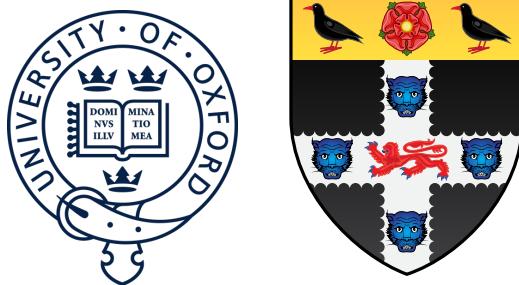


Measurement of Oscillations in Solar Boron-8 Neutrinos and Studies of Optical Scattering in the SNO+ Detector



Daniel Cookman

Christ Church College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2023

Abstract

SNO+ is a large-scale liquid scintillator experiment based in Sudbury, Canada, capable of probing many aspects of neutrinos. One major property of interest is the neutrino's ability to oscillate between different flavours, an indirect demonstration that neutrinos must have mass.

This thesis performs the first ever measurement of oscillations from ${}^8\text{B}$ solar neutrinos in the scintillator phase of SNO+. Assuming the current global fit flux of ${}^8\text{B}$ solar neutrinos, the neutrino oscillation parameter θ_{12} was measured to be $38.9^{\circ+8.0}_{-7.9}$, using an initial 80.6 days of data. This result is consistent with the current global fit result of $33.44^{\circ+0.77}_{-0.74}$. A sensitivity study indicates that the precision of this result can be improved by at least a factor of two within two years of livetime.

On top of this, substantial improvements were made to all aspects of the optical calibration system known as SMELLIE. This is a series of optical-wavelength lasers whose light is emitted from optical fibres attached to the edge of the SNO+ detector. By developing a new analysis, this system was used to measure the scintillator extinction lengths as a function of wavelength and time *in-situ* for the first time. A new analysis was also built and demonstrated to observe changes in scattering and scintillator re-emission properties of the scintillator as a function of time and wavelength. Alongside this, major upgrades were made to both the hardware and simulation of the SMELLIE system, enabling higher-quality data to be taken, and simulations to be made with much greater speed.

Acknowledgements

No man is an island, entire of itself; every man is a piece of the continent, a part of the main¹. This is especially true in Experimental Particle Physics, which in recent decades has required large collaborations of people working together to explore the new frontiers. The making of this thesis is no different; I am deeply indebted to a great number of people who have helped me along the way. Here are some small words of thanks to you all.

To Armin Reichold, thank you so much for being my supervisor: always being there to ask the right questions about my work, and giving me advice about what to look at next. I made it through fighting SMELLIE thanks to your help. I've had so much fun doing research with you, learning so much thanks in part to your words to me: "No black boxes".

To Steve Biller and Jeff Tseng, your advice on matters of analysis, statistics, and computing have been invaluable. My work would be a shambles otherwise. Thank you also for building such a welcoming place to do Physics.

To Kim Proudfoot and Sue Geddes, who are the primary reason why the Particle Physics sub-department runs in any capacity. Also, to the IT Support staff; my work literally could not have been done without you guys keeping the interactive machines and internet connections working nicely. I'm sorry, Vip, for blocking the queue that time.

¹*John Donne, Devotions upon Emergent Occasions*

To the ever-growing family of colleagues and friends who I have had the pleasure of working in the SNO+ office with: Tereza, Iwan, Josh, Rafael, Cal, Gulliver, Abi, Jasmine, Qingyang, and Ellie. Special thanks to Josie for being an awesome thesis-writing buddy, and a brilliant friend. Also, to the elite strike squad of postdocs: Ed, Will, Ana Sofia, and Ben. I would be remiss in forgetting all the folks who have worked on SMELLIE, including Esther Turner and Jeff Lidgard. Jeff, thanks so much for being willing to fight through SMELLIE ‘integration hell’ with me.

To Vic, Sierra, Caroline, Steph, Cindy, Matt, Mark, Aleksandra, Christine, and Ryan, thanks so much for making Sudbury so inviting whilst I was there; be it stuck underground without power, singing at Little Montreal, sliding on inflatables on Ramsey Lake, or getting stuck in Espanola. James and Rafael (again), it was great fun being housemates. Juliette, thanks for being my partner-in-crime as YM representatives: I hope your canoe is still okay.

To my friends outside the SNO+ cult, you are what makes life worth living: Eimear, Soniya, Ynyr, Ciaran, Niamh, Steph, Gwen, Sol, Ben, and so many others. Also, to everyone I’ve played with in Quadball (*née* Quidditch). Never Forget to be Awesome.

And finally, to my family: Mum, Dad, Joel, Ben, and Leo. Thank you all for being so loving and fun, even after so many months in lockdown together. I would not be here without you. ...*All shall be well, and all shall be well, and all manner of thing shall be well.*²

²*Julian of Norwich*

To my parents

and

To Oscar Jacobsson:

The best of us

Table of contents

List of Figures	xiii
List of Tables	xxi
List of Acronyms	xxiii
Introduction	1
1 The Theory of Neutrino Physics	3
1.1 The Standard Model and Neutrinos	3
1.2 Neutrino Oscillations and Neutrino Masses	6
1.2.1 The Evidence for Neutrino Oscillations	8
1.2.2 The Phenomenology of Neutrino Oscillations	14
1.2.3 The Origins of Neutrino Mass	21
2 The SNO+ Detector	27
2.1 Detector Geometry	27
2.2 Experimental Phases	29
2.3 Detecting and Recording an Event in SNO+: A Journey	30
2.3.1 Particle Interactions with Matter	31
2.3.2 Optical Processes	36
2.3.3 Detection by PMTs	42
2.3.4 Data Acquisition and Triggering	44

2.3.5	Operation of the Detector	48
2.4	Detector Calibrations and Modelling	49
2.4.1	Detector Monitoring	49
2.4.2	Electronic and PMT Calibrations	49
2.4.3	Energy and Optical Calibrations	51
2.4.4	Event Reconstruction	53
2.4.5	Event Simulation	56
3	The SMELLIE Calibration System	57
3.1	Lasers	60
3.2	Controlling Laser Intensities	62
3.3	Propagation of Light into the Detector	67
3.4	The Monitoring PMT Unit	71
3.5	Event Triggering and Data Acquisition	73
3.6	Software for SMELLIE Data-taking	76
4	Simulating SMELLIE Events	79
4.1	Improving the SMELLIE Generator Algorithm	80
4.1.1	Previous Attempts at SMELLIE Event Simulation	80
4.1.2	The new generator	83
4.2	Improving the beam profiles	87
4.2.1	Combining beam profile datasets	90
4.3	Comparisons between Data and Simulation	96
4.3.1	Forward Hemisphere Discrepancies	96
4.3.2	Emission Time Discrepancies	100
4.3.3	Backward Hemisphere Discrepancies	102
4.4	Summary and Suggestions for Further Work	106

5 Analysis of SMELLIE Data in the Scintillator Phase	109
5.1 Extinction Length Analysis	109
5.1.1 Mathematical Model	110
5.1.2 Parameter Measurements and Uncertainties	114
5.1.3 Results in Data	117
5.2 Scattering Analysis	128
5.2.1 Historical Approaches and the Problem of Systematics . .	128
5.2.2 New Methodology	130
5.2.3 Results in Data	136
5.2.4 Interpretation	137
5.3 Summary and Suggestions for Further Work	140
6 Solar Oscillation Analysis	145
6.1 Observational Principle	146
6.2 Background Processes	151
6.3 Statistical Approach	160
6.3.1 The Log-likelihood Test Statistic	160
6.3.2 The Bayesian Statistical Approach	161
6.3.3 Markov Chain Monte Carlo	162
6.3.4 Choosing Priors	164
6.3.5 Including Systematics in the Fit	165
6.3.6 Including Oscillations in the Fit	167
6.4 Analysis on Scintillator-Phase data	169
6.4.1 Dataset and Livetime	169
6.4.2 Event Selection	171
6.4.3 Expected Rates and their Constraints	176
6.4.4 Systematics	176
6.4.5 Results	180

6.5 Sensitivity Projections	191
6.6 Summary and Suggestions for Further Work	196
7 Conclusions	201
References	203
Appendix A Expected Rates and Constraints in the Solar Analysis	217
A.1 Boron-8 Signal	217
A.2 Internal Uranium- and Thorium-Chain Backgrounds	219
A.3 Alpha-n Reactions	220
A.4 External Backgrounds	221

List of Figures

1.1	Diagram of the pp chain and CNO cycle within the Sun	9
1.2	Solar neutrino energy spectrum, and associated uncertainties from the SMM	10
1.3	The two tree-level Feynman diagrams associated with neutrino- electron elastic scattering	12
1.4	Comparison of measured solar neutrino fluxes in the SNO CC, NC, and ES modes to the SSM	13
1.5	Measured survival probability versus mean neutrino energy for each solar neutrino type, as observed by Borexino	21
1.6	Masses excesses of the $A = 130$ isobar, with the allowed weak decays being shown	24
1.7	Feynman diagram for $0\nu\beta\beta$ decay, and sketch of the energy spectra for $2\nu\beta\beta$ and $0\nu\beta\beta$	25
2.1	3D model of the SNO+ detector	28
2.2	Comparison of the SNO+ detector media's emission properties, versus optical phase	32
2.3	Comparison between the observed emission time of ^{214}Bi β -decays in data versus MC	34
2.4	Scattering lengths in simulation for UPW, LAB, and 2.2 g/L LAB- PPO	39

2.5	Measured properties of the UPW and acrylic in the water phase	41
2.6	Refractive indices of acrylic, UPW, and LAB-PPO as a function of wavelength; also reflectance as a function of incidence angle	42
2.7	Diagram of the PMT and concentrator ‘bucket’ used within SNO+; and plot of the measured relative angular response of the PMTs in SNO+	43
2.8	Efficiencies of the R1408-type PMTs used as standard within SNO+ .	44
2.9	Example charge spectra for a PMT as a function of the true npe generated	45
2.10	Schematic of the front-end electronics used for data acquisition and triggering in SNO+	46
3.1	Diagram of the SMELLIE calibration system within SNO+	58
3.2	A brief history of SMELLIE hardware work and data-taking campaigns	60
3.3	Diagram of the connections between the various bits of SMELLIE hardware and the rest of the SNO+ detector	61
3.4	Emission wavelength and timing spectra of the lasers used in SMELLIE	63
3.5	Typical impact of driving voltage intensity for PQ446 laser on observed maximum PMT occupancy in the detector	64
3.6	Distribution of the number of PMT hits observed per event for laser PQ446, before the VFA was added	65
3.7	Picture of the VFA during installation into the ELLIE hardware rack	66
3.8	Maximum occupancy in the detector as a function of the attenuation parameter value used	67
3.9	Distribution of the number of PMT hits observed per event for laser PQ446, after the VFA was added	68

3.10 Fibre emission points and nominal pointing directions	70
3.11 Diagram of the SMELLIE coordinate system	71
3.12 Comparison between typical MPU CAEN traces generated by the PQ495 before and after the Summer 2022 hardware upgrades . . .	72
3.13 Comparison of observed hit times in beamspot before and after TUBii firmware update	74
4.1 Comparison of SMELLIE between data and MC using a 1D generator	81
4.2 Number of attempts taken for existing 2D SMELLIE generator to accept direction	84
4.3 Intensity map of existing beam profile, and interpolated sampling map	85
4.4 Histogram of interpolated intensities within the 2D direction-space	86
4.5 Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches	88
4.6 Plot of $\Delta\chi^2$ for both single subruns of a typical PMT, and when all relevant subruns are combined	95
4.7 Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets	97
4.8 Interpolated intensity map for the new updated beam profile of fibre FS055	98
4.9 Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055	100
4.10 Time residual distribution in the beamspot region for data taken in July 2022 with PQ495 through fibre FS007, as compared to a matching simulation	101
4.11 Data–MC comparisons for water phase SMELLIE data in the back-scatter PMT region	103

4.12 Comparison of Laserball data in the water phase to MC	104
5.1 Current RAT models for the extinction lengths of both UPW and scintillator, as a function of wavelength	110
5.2 Diagram showing the two PMT regions used in the SMELLIE extinction length analysis	111
5.3 Linear interpolation being used on Laserball UPW attenuation coefficient data	115
5.4 Time residual distributions of the far and backscatter PMT regions for a simulation using laser PQ407 through fibre FS007, split by the different optical components	118
5.5 Comparison of the backscatter light region time residual distribution with intensity and emission time calculation choice	119
5.6 Unstacked time residual distribution of a simulation of light from the PQ375 laser firing through fibre FS007, split by optical components	121
5.7 Far PMT time window selection applied to a simulation of the PQ407 laser fired through laser FS107	122
5.8 Results of the extinction length analysis using the PQ407 laser . .	123
5.9 Results of the extinction length analysis using the SuperK laser in the 400–410 nm range	125
5.10 Summary of scintillator extinction length measurements made by SMELLIE	127
5.11 Comparison of the signal and total hit distribution in the (t_{res}, α) -space using new and old beam profiles for laser PQ495 through fibre FS055	131
5.12 Hit distribution of SMELLIE data taken in run 302,634, with laser PQ495 through fibre FS055, as a function of both t_{res} and α . . .	132

5.13	Model of light rays propagating in the SNO+ detector during scintillator phase, emanating from an external source such as SMELLIE133	
5.14	Timing distribution of a simulation of the observed hits in the bad light-path PMT selection, split by optical components	134
5.15	Comparison of timing distributions in the bad light-path PMT region for simulations of laser PQ495 through fibre FS055, comparing the old and new beam profile	135
5.16	Plots of R_1 against R_2 for four pairs of scintillator phase datasets, using the SuperK laser	138
5.17	Summary of results from the scattering analysis	139
5.18	Observed time residual distributions for the backscattered light region for FS037, comparing May 2021 to June 2023 data	143
6.1	P_{ee} as a function of true neutrino energy, scanned over a variety of oscillation parameter values	147
6.2	differential cross-section for neutrino-electron elastic scattering of electron neutrinos as a function of T and E_ν	149
6.3	The evolution of energy distributions related to ^8B solar neutrino detection	149
6.4	Comparison of the reconstructed electron energy distribution expected for 9 different oscillation parameter combinations	150
6.5	Solar oscillation parameter contours, from both the NuFit 5.1 global fit, and initial SNO+ sensitivity study	151
6.6	The ^{238}U and ^{232}Th decay chains.	153
6.7	The distributions of the IW $^{212}\text{Bi-Po}$ classifier for Bi-Po and ^8B events in simulation, within the analysis ROI	155
6.8	$^{214}\text{Bi-Po}$, $^{212}\text{Bi-Po}$, and ^8B observed energy spectra before and after OOW and IW cuts	156

6.9	Schematic of (α, n) interactions	156
6.10	(α, n) reconstructed energy spectrum in MC within the analysis ROI, before and after out-of-window cuts have been applied	157
6.11	Radial dependence of external backgrounds relevant in this analysis, as compared to the ${}^8\text{B}$ signal	158
6.12	Distributions of the two external classifiers, for both the ${}^8\text{B}$ signal and ${}^{208}\text{Tl}$ decays in the acrylic	160
6.13	Demonstration of how to handle an energy scaling systematic correctly	166
6.14	Distribution of $\cos \theta_z$, for the events within this dataset, compared to a uniform time distribution	171
6.15	Comparison of tagged ${}^{214}\text{Bi}$ reconstructed energy distributions between data and MC	179
6.16	Examples of the sampled values for parameters within a given MCMC chain as a function of the step number	183
6.17	Marginalised 1D posterior density distributions for all non-oscillation parameters	185
6.18	Correlation matrix between all parameters in the MCMC fit . . .	186
6.19	2D posterior density distribution marginalised onto Δm_{21}^2 and θ_{12}	187
6.20	Comparison of 1D posterior density distribution marginalised onto θ_{12} for different flux constraints	188
6.21	Comparison of data to MC for each radial slice	191
6.22	Posterior density posterior distributions marginalised onto the two solar oscillation parameters, for each livetime scenario	194
6.23	Posterior density posterior distributions marginalised onto each of θ_{12} and Δm_{21}^2 , for each livetime scenario	195
6.24	Posterior densities marginalised onto the two oscillation parameters, for projections over 1 year of livetime, with different background expectations	197

6.25 Posterior densities marginalised onto θ_{12} , for projections over 1 year of livetime, with different background expectations.	198
--	-----

List of Tables

1.1	Global Fit neutrino oscillation parameters	17
1.2	Current best limits on the half-life for $0\nu\beta\beta$ decay for selected isotopes	26
2.1	Current values used to model scintillator emission from electrons in 2.2 g/L LAB-PPO	34
3.1	SMELLIE fibre names, their associated mounting nodes on the PSUP, and their pointing direction.	69
4.1	Water-phase runs used for new beam profiling.	90
5.1	Datasets used in the SMELLIE extinction length analysis	120
6.1	Summary of cuts used for coincidence tagging	154
6.2	Cuts used in this analysis.	172
6.3	Impact of each cut on the quantity of events in data. The efficiency for a given cut is defined here as the fraction of events which have survived all previous cuts, which also survive that cut.	174
6.4	Combined impact of cuts on each MC process	175
6.5	Number of events expected both before and after cuts, along with any constraints	177

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	190
A.1 Measured rates of the external backgrounds during the water phase of SNO+, by Tony Zummo	222

List of Acronyms

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

$2\nu\beta\beta$ Two-neutrino double beta decay

ADC Analogue-to-Digital Converter

AmBe Americium-Beryllium radioactive source

AMELLIE Attenuation Module for ELLIE

AV Acrylic vessel

BD 1,2-Butanediol

BEH Brout-Englert-Higgs (Mechanism)

BHT Butylated hydroxytoluene

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

BSM Beyond the Standard Model

CC Charged Current (weak interaction)

CDF Cumulative Density Function

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

ES Elastic Scattering

EW Electroweak (Theory)

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

IBD Inverse β -decay

IH Inverted Hierarchy (of neutrino masses)

IW In-window events

LAB Linear alkylbenzene

MPU Monitoring PMT Unit (for SMELLIE)

MSW effect Mikheyev-Smirnov-Wolfenstein effect (of neutrinos)

MTC/A+ Analogue Master Trigger Card

MTC/D Digital Master Trigger Card

NC Neutral Current (weak interaction)

NDF Number of Degrees of Freedom (in a statistical fit)

NH Normal Hierarchy (of neutrino masses)

NI Unit National Instruments Bus-Powered M Series Multifunction DAQ Unit
(for SMELLIE)

npe Number of photoelectrons

OOW Out-of-window events

OWLs Outward-looking PMTs

PCA PMT Calibration

PMNS Pontecorvo-Maki-Nakagawa-Sakata (neutrino mixing matrix)

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

SM The Standard Model of Particle Physics

SNO Sudbury Neutrino Observatory

SSB Spontaneous Symmetry Breaking

SSM Standard Solar Model

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)

MC Monte Carlo

nhit Number of PMT hits in an event

RATDB RAT Database

RAT Reactor Analysis Tool

Introduction

It is an exciting time in Neutrino Physics. Over the past decades, the evidence for neutrinos oscillating between different flavours has become overwhelming, to the point that the Nobel Prize in Physics was awarded in 2015 for the discovery [1]. However, as one set of questions gets answered, others get raised: if neutrinos oscillate as they propagate through space, then by Special Relativity they cannot be massless. But what are those masses? What underlying mechanism enables them to be massive, but still requires them to be extraordinarily light? And can we make precision measurements of how this oscillation phenomenon occurs?

These are some of the major questions that current neutrino physics experiments seek to answer. One such experiment is SNO+, a large-scale multipurpose neutrino detector built 2 km underground in Sudbury, Canada. Filled with 800 tonnes of liquid scintillator, neutrinos from a wide variety of sources can be detected. The primary goal is the search for neutrinoless double beta decay ($0\nu\beta\beta$): if discovered, it would be smoking-gun evidence of how neutrinos get mass, and could also provide us a way of measuring the neutrino mass scale.

This thesis describes work done by the author to help progress the SNO+ experiment. These efforts can be split into two categories: optical calibration of the detector, and performing an oscillation analysis of solar neutrinos. In order to provide sufficient context for these results, this document begins firstly with a chapter summarising important results in neutrino physics, followed by a chapter on the details of the SNO+ detector.

Chapters 3–5 cover the work done on one of the optical calibration systems for SNO+, known as SMELLIE. The system is introduced in Chapter 3, with explanations of how the system works, and the hardware upgrades made in the Summer of 2022. Chapter 4 explains how improvements were made to the simulation of SMELLIE events, including a dramatic reduction in the time needed to generate a simulated event. There is also a discussion of the remaining discrepancies that exist between data and simulation, despite improvements made to the simulation of the angular emission distributions. With these systematic effects in mind, Chapter 5 goes over the creation and implementation of two new analyses of SMELLIE data. In one, the extinction lengths of the scintillator were measured *in-situ* as a function both of wavelength and time. In the other, changes in the scattering length and scintillation re-emission properties could be observed. Understanding the optical properties of the detector with precision is critical to being able to perform physics analyses with data from SNO+.

One of the major analyses of interest (other than $0\nu\beta\beta$) for SNO+ is measuring the parameters that govern neutrino oscillations. This can be achieved by looking at the observed energy spectrum of ^8B solar neutrinos: this is the subject of Chapter 6. For the first time, an analysis has been built and performed using scintillator phase data from the detector. The future prospects of this analysis as more data is taken are also investigated.

Chapter 1

The Theory of Neutrino Physics

Light

Light

The visible reminder of Invisible Light

The Rock

T. S. ELIOT

1.1 The Standard Model and Neutrinos

The *Standard Model* (SM) of Particle Physics is the culmination of a century’s work by scientists to understand the fundamental constituent elements of the Universe, and their interactions. Within the SM, fundamental particles are excitations of associated quantum fields within spacetime. One class of particles in the SM are known as the neutrinos, ν : these are spin-1/2 fermions which are neutral in both the strong and electromagnetic force. The only means by which they are known to interact is through the weak nuclear force. There are three ‘flavours’ of neutrino, one associated with each of their charged lepton counterparts: the electron neutrino ν_e , the muon neutrino ν_μ , and the tau neutrino ν_τ .

Crucial to understanding the nature of neutrinos is their interactions with other particles. Within the SM, the weak nuclear force and electromagnetism are unified into the Electroweak (EW) Theory by the work of Glashow, Salam, and Weinberg [2–4]. This is a so-called *chiral gauge field theory*. Gauge field theories are a special type of quantum field theory which demand that the Lagrangian density \mathcal{L} is invariant under certain kinds of transformation, in addition to the usual requirement of Lorentz invariance. For EW, the Lagrangian is invariant under ‘local’ transformations of the fields’ internal degrees of freedom, defined by the ‘gauge’ group $SU(2)_L \times U(1)_Y$, where L and Y are known as the left-handed weak isospin and weak hypercharge, respectively.

A local transformation is one which changes values of the fields in a manner that is dependent on the spacetime coordinates. By demanding invariance under these gauge transformations, as well as Lorentz invariance, the theory naturally predicts the existence of vector (spin-1) boson particles. These are known as the ‘gauge’ bosons of the theory, and they mediate the interactions defined by the gauge group. The massive W^\pm and Z^0 bosons, discovered by the UA1 and UA2 experiments in 1983 [5–7], mediate the weak nuclear force, whilst the massless photon γ mediates the electromagnetic force.

The theory of EW interactions is also *chiral*. Any spinor that defines the wavefunction of a spin-1/2 field can be split into its left- and right-handed ‘chiral’ components, defined through the projection operators $P_{L,R} = \frac{1 \mp \gamma^5}{2}$. The force associated with the $SU(2)_L$ part of the EW gauge group only interacts with the left-handed components of particles, denoted with the subscript L on their wavefunction.

The Lagrangian that defines the weak interactions of neutrinos is:

$$-\mathcal{L} = \frac{g}{2 \cos \theta_W} \sum_{\ell,L} \bar{\nu}_{\ell,L} \gamma^\mu \nu_{\ell,L} Z_\mu^0 + \frac{g}{\sqrt{2}} \sum_\ell \bar{\nu}_{\ell,L} \gamma^\mu \ell_L^- W_\mu^+ + \text{h.c..} \quad (1.1)$$

Here, g is the dimensionless coupling constant associated with $SU(2)_L$, and θ_W is the Weinberg angle. The three lepton flavour fields are denoted by $\ell = e, \mu, \tau$, with their associated neutrino fields being given by ν_ℓ . Similarly, the fields associated with the weak gauge bosons are given by W^\pm and Z^0 . The two components of this Lagrangian are known as the Neutral Current (NC) and Charged Current (CC) weak interactions of neutrinos, respectively. Similar Lagrangians exist that define the NC and CC interactions of quarks, as well as the NC interactions of the charged leptons.

Solidifying this theoretical picture are decades-worth of experimental tests of neutrinos and their place in the SM. The first neutrinos to be detected were electron anti-neutrinos, by Cowan and Reines in 1956 [8, 9]. These neutrinos were generated in the β -decay of radioactive isotopes within the Savannah River nuclear reactor: $n \rightarrow p + e^- + \bar{\nu}_e$. This decay arises from a down quark within the neutron of an atom converting into an up quark via a CC interaction, generating a virtual W^- boson that promptly decays into an electron and $\bar{\nu}_e$. The method by which Cowen and Reines detected these anti-neutrinos was through *inverse β -decay* (IBD): $\bar{\nu}_e + p \rightarrow e^+ + n$. This process also originates from CC interactions. Analogous CC interactions allowed Danby *et al* to discover the muon neutrino in 1962 [10], and the DONUT Collaboration to discover the tau neutrino in 2000 [11].

The existence of NC interactions with neutrinos and anti-neutrinos was first demonstrated by the Gargamelle experiment in 1974 [12–15]. In particular, the observation of anti-muon neutrino electron elastic scattering, $\bar{\nu}_\mu + e^- \rightarrow \bar{\nu}_\mu + e^-$ by the experiment was an unambiguous demonstration of NC interactions.

In 1958, Goldhaber *et al* [16] were able to demonstrate experimentally that the helicity of electron neutrinos, i.e. the component of their spin along the direction of motion, is -1 . Because the weak interactions are chiral, no evidence of neutrinos with positive helicities (or equally, anti-neutrinos with negative helicities) exists. This stands in firm contrast to all other SM particles.

No flavours of neutrino beyond the electron, muon, or tau types have been discovered. A combined analysis of data from the four LEP experiments looking at the decay width of the Z boson was able to indirectly measure the number of neutrino species that could undergo NC interactions and had masses less than one half of the Z boson: $N_\nu = 2.9963 \pm 0.0074$ [17, 18]. This measurement is very strong evidence that no other ‘light’ weakly-interacting neutrinos exist.

1.2 Neutrino Oscillations and Neutrino Masses

So far in this description, no attempt has been made to explain the origin of the masses of the fundamental particles. It is certainly straightforward to naïvely add a mass term such as $m_e \bar{e}_L e_R$ into the SM Lagrangian, where m_e is the mass of the electron. However, one can show that any mass terms added will necessarily violate the $SU(2)_L \times U(1)_Y$ symmetry that defines the EW interactions [19]. The weak vector bosons W^\pm and Z would then need to be massless, in contradiction with observations.

The solution to this problem comes in the form of the *Brout-Englert-Higgs (BEH) Mechanism* [20–22]. In this Mechanism, an additional two-component “Higgs” field H is proposed, which is able to interact with the other fields of the theory in a manner that preserves the SM gauge symmetries. One part of the added Higgs field interactions are the so-called Yukawa terms, which for interactions with leptons are given by:

$$-\mathcal{L}_{\text{Yukawa,lep}} = \sum_\ell y_\ell \bar{L}^\ell H \ell^c + \text{h.c.}, \quad (1.2)$$

where y_ℓ are the “Yukawa” coupling constants for the three lepton flavours, $L^\ell = \begin{pmatrix} \nu_{\ell,L} \\ \ell_L \end{pmatrix}$ are the left-handed lepton doublets of the SM, and ℓ^c are the right-handed charged leptons.

The key to the BEH Mechanism is *Spontaneous Symmetry Breaking* (SSB): the Higgs field is defined in such a way that the ground state takes a non-zero ‘vacuum expectation value’, v . By doing so, the underlying gauge symmetry of the EW interactions is spontaneously broken as $SU(2)_L \times U(1)_Y \rightarrow U(1)_Q$, where $U(1)_Q$ is the residual electromagnetic charge conservation. The above Yukawa Lagrangian term after symmetry breaking generates the mass terms for the charged leptons:

$$-\mathcal{L}_{\text{Yukawa,lep}} \rightarrow \sum_{\ell} m_{\ell} \bar{\ell}_L \ell^c + \text{h.c.}, \quad (1.3)$$

where $m_{\ell} = \frac{v}{\sqrt{2}} y_{\ell}$ are the charged lepton masses. Other terms associated with Higgs interactions in the SM generate mass terms for the quarks and weak vector bosons, as seen in data. A further prediction of this BEH Mechanism is the existence of a massive scalar boson known as the Higgs particle; this was discovered in 2012 by the ATLAS and CMS Collaborations [23, 24].

The one type of fundamental particle not covered by the above argument are neutrinos. If neutrinos were massless, then there is no issue: we observe neutrinos to have only negative helicities, which is equal to left-handed chiralities if they are massless. As the SM contains no right-handed neutrinos, no Yukawa interaction can be built to generate masses for the neutrinos. One can also demonstrate that, in the SM, neutrinos cannot even obtain masses through loop corrections [25].

This assumption of massless neutrinos appears initially consistent with the current observations of direct neutrino mass measurements. The strongest direct limits come from the KATRIN experiment, which looks at the endpoint of the tritium β -decay spectrum. The ‘effective’¹ electron anti-neutrino mass was measured to be $m_{\nu} < 0.8$ eV at a 90% confidence level [26]. Even stronger limits

¹ KATRIN measures the ‘effective’ mass and not the actual mass, because of the phenomenon of neutrino oscillations as described in the following sections.

are available from cosmology, by looking in part at the power spectrum of the Cosmic Microwave Background. Assuming the so-called Standard ‘ Λ CDM’ Model of Cosmology, limits on the sum of all three neutrino flavours $\sum m_\nu < \mathcal{O}(0.1 \text{ eV})$ have been achieved [27].

1.2.1 The Evidence for Neutrino Oscillations

Despite the current lack of any direct measurements, we now know that at least some neutrino flavours must have mass. This is because of the phenomenon of *neutrino oscillations*, which have been observed over a variety of experiments and contexts. The critical pieces of evidence for this process are described here; the underlying mathematical model that is used to explain them quantitatively is described in Section 1.2.2.

Solar Neutrinos

Neutrinos are generated from the Sun as a by-product of the fusion reactions at its core. At the highest level, protons fuse into alpha particles by the overall reaction $4 p \rightarrow {}^4\text{He} + 2 e^+ + 2 \nu_e$, generating also $\sim 25 \text{ MeV}$ of energy that enables the Sun to shine [28]. This process is known as ‘hydrogen burning’. The Standard Solar Model (SSM) is the current best quantitative description of stellar evolution for main sequence stars, and our Sun in particular. It covers the nuclear reactions that generate both the energy that powers the star and the changes in the relative isotopic abundances, how the energy is transported out through the star via radiation of photons and convection, and how the outward pressures caused by this radiation is balanced by gravity to maintain hydrostatic equilibrium. An introduction to the SSM can be found in [28].

In the Sun, two sets of nuclear reactions enable hydrogen burning to occur: the *proton-proton (pp) chain* and *CNO cycle*. Diagrams of these reaction chains

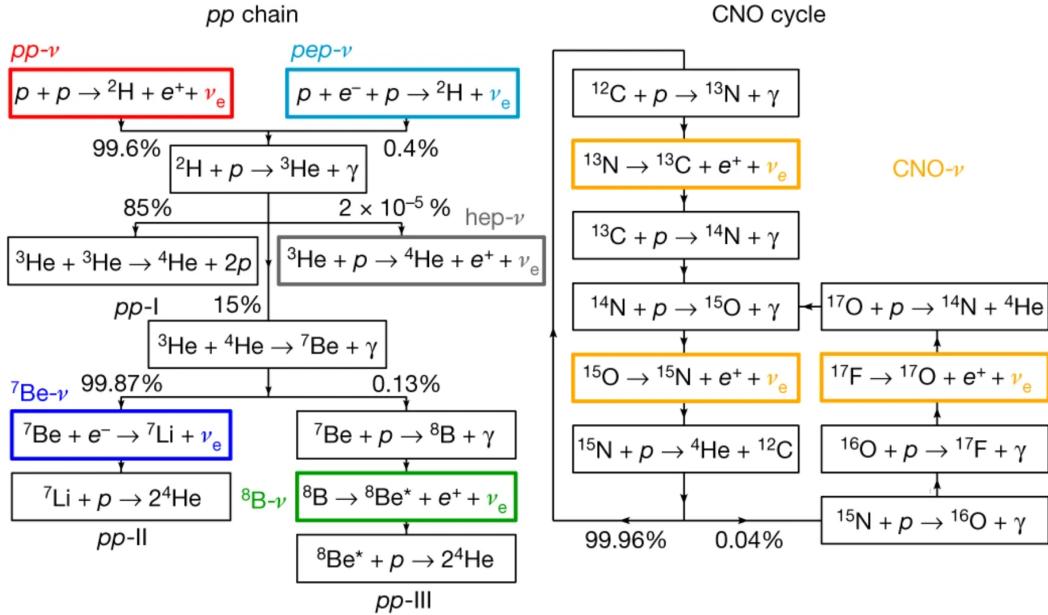


Fig. 1.1: Diagram of the pp chain and CNO cycle within the Sun, with the reactions that generate neutrinos highlighted. Taken from [29].

are shown in Fig. 1.1. In the pp chain, protons are first fused together to form a ^2H nucleus through the ‘pp’ and ‘pep’ reactions, the latter also using an electron. Once a deuterium nucleus has been generated, it strongly interacts with a proton to create a ^3He nucleus. The dominant method for hydrogen burning to then terminate is for two ^3He nuclei strongly interact to generate ^4He and two protons.

Two other nuclear reactions with ^3He are possible. In one, ^3He fuses with ^4He to generate a ^7Be nucleus, which can then generate a neutrino either from the creation of ^7Li via electron capture, or from the additional fusing into ^8B which promptly β^+ -decays. These are known as the ^7Be and ^8B solar neutrino generation reactions, respectively. The final and rarest reaction within the pp chain that generates a neutrino is the so-called ‘hep’ reaction, in which ^3He directly fuses with a proton.

The CNO cycle is a secondary means by which the Sun can burn hydrogen. This is achieved through the aid of a ^{12}C nucleus as a catalyst. Part of the cycle

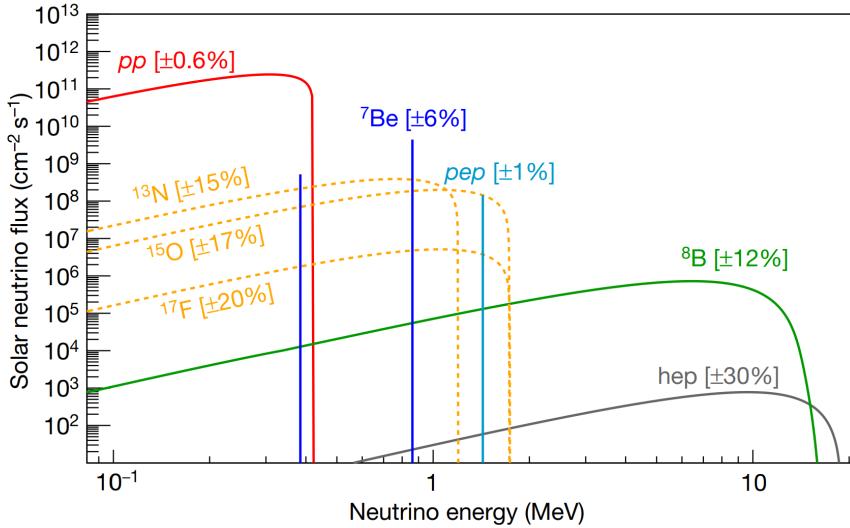


Fig. 1.2: Solar neutrino energy spectra, and associated uncertainties from the SSM. Figure is taken from [29, 31]; note that the flux is given in units of $\text{cm}^{-2} \text{s}^{-1} \text{ MeV}^{-1}$ for the continuum sources, and $\text{cm}^{-2} \text{s}^{-1}$ for mono-energetic sources.

involves the generation of unstable isotopes ^{13}N and ^{15}O , both of which β^+ -decay, creating electron neutrinos. In a rare side-chain of the CNO cycle, it is possible to also generate ^{17}F which also weakly decays to generate a neutrino. This is the final method of generating neutrinos in our Sun.

The SSM quantitatively predicts the flux and energy spectra of neutrinos generated through each of the above processes, as incident on the Earth. This is shown in Fig. 1.2. The shapes of the energy spectra are determined by the nuclear reactions that define the process: for example, the broad shape of the $^8\text{B} \nu_e$ energy spectrum comes from the β^+ -decay of ^8B isotopes, and has been measured in nuclear beam experiments to high precision [30].

The rate of the ^8B decays in the Sun, and hence the generated neutrino flux, depend strongly on the radial temperature distribution of the Sun, the cross-sections of the pp chain nuclear reactions, and the Sun's chemical composition. The latter point remains a topic of some controversy: measurements of the relative abundances in 1998 through spectroscopy of the Sun's photosphere as well as

meteorites give a ‘metal-to-hydrogen’² ratio of $Z/X = 0.023$ [32], whereas a more recent study in 2009 has a substantially lower value of $Z/X = 0.018$ [33]. These two models are called the ‘high-metallicity’ **GS98** model and ‘low-metallicity’ **AGSS09met** model, respectively. The current best SSM associated with these two abundance models, denoted **B16_GS98** and **B16_AGSS09met**, have ${}^8\text{B}$ flux predictions of $\Phi_{\text{sB}} = (5.46 \pm 12\%) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ and $\Phi_{\text{sB}} = (4.50 \pm 12\%) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$, respectively [31].

Since the late 1960s, a number of experiments have been able to detect solar neutrinos, using different techniques. The earliest of these was the Homestake Chlorine detector, which used the capture of electron neutrinos on ${}^{37}\text{Cl}$ nuclei to generate ${}^{37}\text{Ar}$ atoms that could be chemically extracted and counted [34]. In a similar vein, the SAGE and GALLEX/GNO experiments detected the capture of electron neutrinos on ${}^{71}\text{Ga}$ [35, 36].

The Kamiokande experiment, and its successor Super-Kamiokande, are large water Cherenkov detectors, sensitive to solar neutrinos through neutrino-electron elastic scattering (ES): $\nu_x + e^- \rightarrow \nu_x + e^-$. All flavours of neutrino are capable of scattering through a NC process, but there is an additional CC mode for electron neutrinos, as shown in Fig. 1.3. The differential cross-section for this interaction as a function of the scattered electron’s kinetic energy T is given by [37]:

$$\frac{d\sigma_{\nu_i}}{dT} = \frac{2G_F^2 m_e}{\pi} \left\{ g_{L,i}^2(T) \left[1 + \frac{\alpha}{\pi} f_-(z) \right] + g_{R,i}^2(T) (1-z)^2 \left[1 + \frac{\alpha}{\pi} f_+(z) \right] - g_{R,i}(T) g_{L,i}(T) \frac{m_e}{E_\nu} \left[1 + \frac{\alpha}{\pi} f_{+-}(z) \right] \right\}, \quad (1.4)$$

where $i = e, \mu$ ³, G_F is the Fermi coupling constant, m_e is the electron mass, α is the fine-structure constant, E_ν is the incident neutrino energy, and $z = T/E_\nu$. $g_{L,R}(T)$ are the left- and right-handed running chiral couplings, which have a

²‘Metal’ here is used in the astrophysical sense: elements heavier than hydrogen or helium.

³ $\frac{d\sigma_{\nu_\mu}}{dT} = \frac{d\sigma_{\nu_\tau}}{dT}$ because both flavours of neutrino only undergo the NC interaction.

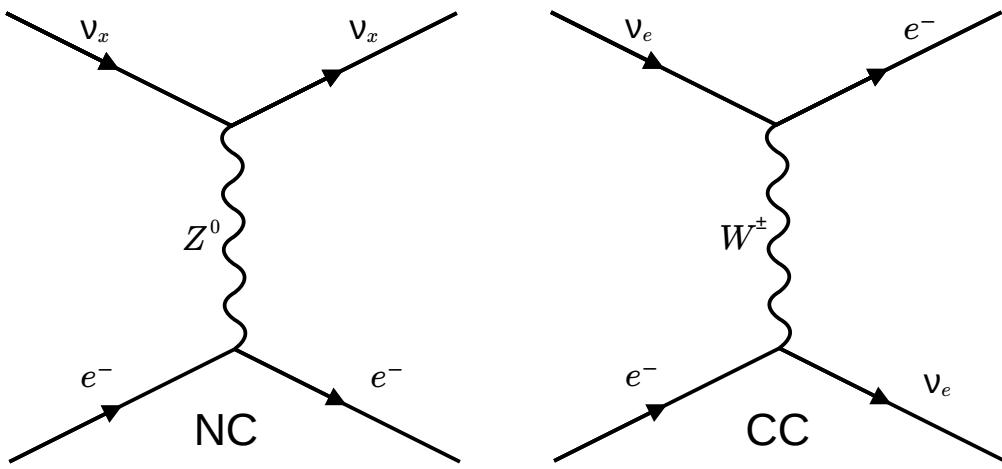


Fig. 1.3: The tree-level Feynman diagrams associated with the NC and CC modes of neutrino-electron elastic scattering.

T -dependence because of radiative corrections. Similarly, $f_-(z)$, $f_+(z)$, and $f_{+-}(z)$ are all QED radiative correction terms.

A consistent finding in all of these solar neutrino experiments was that the measured rate of neutrino interactions was substantially less than the expectation given by the SSM [35, 36, 38, 39]. This was known as the *Solar Neutrino Problem*. To solve it, the Sudbury Neutrino Observatory, SNO, was built. A large spherical acrylic vessel 2.2 km underground was filled with 1000 tonnes of heavy water, ${}^2\text{H}_2\text{O}$, from which Cherenkov light due to particle interactions could be detected [40]. Neutrinos were able to interact with the heavy water via three complementary modes: the CC process $\nu_e + d \rightarrow e^- + 2 p$, the NC process $\nu_x + d \rightarrow \nu_x + p + n$, and the ES process described above.

Results of the measured fluxes of ν_e and $\nu_{\mu,\tau}$ solar neutrinos for the three detection modes in SNO, ES results from Super-Kamiokande, and comparison to the SSM are shown in Fig. 1.4. Without a flavour-changing process like neutrino oscillations, no $\nu_{\mu,\tau}$ solar neutrinos should be observed. Because the NC interaction is insensitive to neutrino flavour, it is able to directly measure the total flux of ${}^8\text{B}$ solar neutrinos, regardless of flavour. One can see that, from the results, the

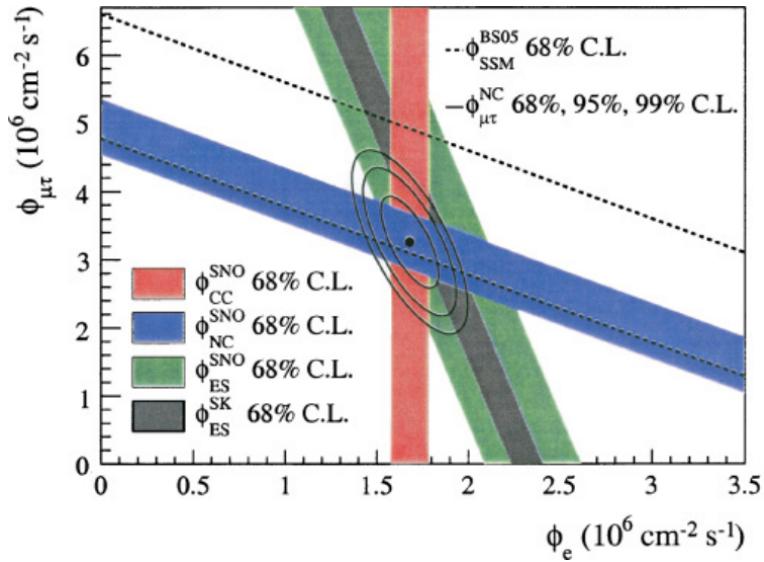


Fig. 1.4: Measured solar neutrino fluxes from electron neutrinos versus muon and tau neutrinos, in the SNO CC, NC, and ES modes (coloured bands). Also shown is the expectation from the SSM (dotted lines), and the ES rate measured from Super-Kamiokande (black band). The combined probability contours are shown in black. Taken from [41].

NC flux measurement is consistent with the total flux expected from the SSM. In contrast, the CC mode is only sensitive to the ν_e flux, whilst the ES mode is sensitive to an admixture of the different neutrino flavours. The results of all measurements of solar neutrinos from both SNO and Super-Kamiokande lead to consistent values of the flux of ${}^8\text{B}$ neutrinos for each flavour, consistent also with the total flux expected by the SSM. Because the measured ν_e flux was low, but the total flux neutrino flux was correct, this was strong evidence for some form of neutrino oscillations.

Other Evidence for Neutrino Oscillations

There are three other major areas in which neutrino oscillations have been observed. The first are atmospheric neutrinos, which come from the decays of cosmic rays in the Earth's atmosphere. In the IBM [42], Kamiokande [43], and Super-Kamiokande [44] experiments, these neutrinos were detected via CC interactions in water, with the ability to distinguish between electron and muon tracks coming

from electron and muon neutrinos, respectively. It was consistently observed that the observed ratio of ν_μ to ν_e events was below expectations, a phenomenon known at the time as the *Atmospheric Neutrino Anomaly*. Super-Kamiokande was able to show [45] that the anomaly could be explained by a disappearance of muon neutrinos that depends on the ratio L/E , where L is the estimated distance from production of the neutrino, and E is the neutrino energy.

The next class of evidence comes from the detection of electron anti-neutrinos from nuclear reactors. These have consistently been detected through the use of IBD, all the way back to the first observation of neutrinos by Cowan and Reines [8, 9]. Contemporary experiments such as KamLAND [46], Daya Bay [47], Double Chooz [48], and RENO [49], have all measured disappearances in the rate of IBD interactions that depended in an oscillatory manner on the ratio L/E .

Finally, a variety of experiments have observed neutrinos that were generated from particle accelerators. In particular, the T2K [50, 51] and NOvA [52, 53] experiments have both been able to show the disappearance of both ν_μ and $\bar{\nu}_\mu$ from their accelerator beams, along with the appearance of ν_e and $\bar{\nu}_e$. Recently, the two experiments have seen evidence for differences in the appearance and disappearance rates of neutrinos and antineutrinos [54, 55]. In addition, the OPERA experiment was able to demonstrate the appearance of ν_τ in their detector [56].

1.2.2 The Phenomenology of Neutrino Oscillations

Oscillations in Vacuum

Taken together, the observations described in the previous section naturally lead to the notion of neutrino oscillations: the idea that as neutrinos propagate through space they are capable of changing flavour. Special Relativity precludes massless particles from experiencing time evolution, so any theory of neutrinos changing flavour between one another must require non-zero mass states.

The initial theories describing neutrino oscillations were made by Pontecorvo, Maki, Nakagawa, and Sakata in the 1960s [57, 58]. These theories initially assumed a two-neutrino model of oscillations, but now that ν_τ particles have been observed a three-neutrino model has been adopted. The theory starts by assuming that the flavour eigenstates of neutrinos, $\nu_{e,\mu,\tau}$ are different from the neutrino mass eigenstates, $\nu_{1,2,3}$, and are instead related to one another through the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix, U :

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U \cdot \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \cdot \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (1.5)$$

Because U must be unitary in order to preserve total probability, its components can be parameterised as follows:

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{CP}} & c_{13}c_{23} \end{pmatrix} \cdot \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (1.6)$$

This matrix uses three “mixing angles” $0 \leq \theta_{12}, \theta_{13}, \theta_{23} \leq \pi/2$, and one further parameter called the “CP-violating phase”, $0 \leq \delta_{CP} \leq 2\pi$, with the abbreviations $s_{ij} = \sin \theta_{ij}$ and $c_{ij} = \cos \theta_{ij}$ used in the above expression.

Given that a neutrino flavour eigenstate $|\nu_\alpha(0)\rangle$ ($\alpha = e, \mu, \tau$) is produced in some CC process at time $t = 0$, because mass eigenstates $|\nu_i(0)\rangle$ ($i = 1, 2, 3$) are simultaneously the energy eigenstates when propagating in free space, the time evolution of the neutrino state is given by:

$$|\nu_\alpha(t)\rangle = \sum_{i=1}^3 U_{\alpha i} e^{-E_i t} |\nu_i(0)\rangle. \quad (1.7)$$

E_i is the energy eigenvalues corresponding to the associated mass eigenstates. The oscillation probability of going from one flavour α to another β is given by $P(\nu_\alpha \rightarrow \nu_\beta) = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2$. Assuming that the neutrino is ultra-relativistic so that $E_i \gg m_i$, where m_i is the mass of the i^{th} mass eigenstate, and that all mass eigenstates have the same definite momentum, one can show that the oscillation probability becomes [59]:

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \delta_{\alpha\beta} - 4 \sum_{i < j} \operatorname{Re} \{W_{\alpha\beta,ij}\} \sin^2 \left(\frac{\Delta m_{ij}^2 L}{4E} \right) \\ &\quad - 2 \sum_{i < j} \operatorname{Im} \{W_{\alpha\beta,ij}\} \sin \left(\frac{\Delta m_{ij}^2 L}{2E} \right), \end{aligned} \quad (1.8)$$

where $\delta_{\alpha\beta}$ is the usual Kronecker delta, $W_{\alpha\beta,ij} = U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}$, $\Delta m_{ij}^2 = m_i^2 - m_j^2$, L is the distance between the creation and detection of the neutrinos, and $E_i \approx E$ is the average energy of the neutrino. As can be seen, the probability will oscillate as a function of L/E , in accordance with what was seen in Section 1.2.1.

If anti-neutrinos are produced, then oscillations are governed by U^* , which is equivalent to U but with the CP phase changing sign: $\delta_{CP} \rightarrow -\delta_{CP}$. Therefore, the difference between $P(\nu_\alpha \rightarrow \nu_\beta)$ and $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$ is given by twice the third term of Eq. 1.8.

In a neutrino flavour disappearance measurement, $P_{\alpha\alpha} = P(\nu_\alpha \rightarrow \nu_\alpha)$ is the *survival probability* of the neutrinos. In this case, $W_{\alpha\beta,ij} = |U_{\alpha i} U_{\alpha j}^*|^2$ is real, and the survival probability formula simplifies to:

$$P_{\alpha\alpha} = 1 - 4 \sum_{i < j} |U_{\alpha i} U_{\alpha j}^*|^2 \sin^2 \left(\frac{\Delta m_{ij}^2 L}{4E} \right). \quad (1.9)$$

An experiment is only sensitive to neutrino oscillations from mass splitting Δm_{ij}^2 if X_{ij} is $\mathcal{O}(1)$, where X_{ij} is the phase within the relevant oscillation proba-

Parameter	Normal hierarchy	Inverted hierarchy
$\theta_{12}[\circ]$	$33.44^{+0.77}_{-0.74}$	$33.45^{+0.77}_{-0.74}$
$\theta_{23}[\circ]$	$49.2^{+1.0}_{-1.3}$	$49.5^{+1.0}_{-1.2}$
$\theta_{13}[\circ]$	$8.57^{+0.13}_{-0.12}$	$8.60^{+0.12}_{-0.12}$
$\delta_{CP}[\circ]$	194^{+52}_{-25}	287^{+27}_{-32}
$\Delta m_{21}^2[10^{-5} \text{ eV}^2]$	$7.42^{+0.21}_{-0.20}$	$7.42^{+0.21}_{-0.20}$
$\Delta m_{3\ell}^2[10^{-3} \text{ eV}^2]$	$+2.515^{+0.028}_{-0.028}$	$-2.498^{+0.028}_{-0.028}$

Table 1.1: Global fit results for the neutrino oscillation parameters and mass splittings, as performed in NuFit 5.1 [60]. The results for both the Normal and Inverted Hierarchy are shown: $\Delta m_{3\ell}^2 = \Delta m_{31}^2$ in the former, Δm_{32}^2 in the latter. The results used do not include atmospheric neutrino data from Super-Kamiokande.

bility term:

$$X_{ij} = \frac{\Delta m_{ij}^2 L}{4E} = 1.27 \frac{\Delta m_{ij}^2 [10^{-3} \text{ eV}^2] L [\text{km}]}{E [\text{MeV}]} . \quad (1.10)$$

If $X_{ij} \ll 1$, then $\sin^2(X_{ij}) \rightarrow 0$, and no oscillations are seen due to that mass splitting. If instead $X_{ij} \gg 1$, then as detectors have limited distance and energy resolutions, and neutrino sources generate them over a region of space, what can only be observed is the average effect over many oscillations, $\langle \sin^2(X_{ij}) \rangle = 1/2$.

From the results of a variety of neutrino oscillation experiments, the oscillation parameters and magnitudes of the mass splittings have now been measured, with varying degrees of precision. A global fit of the experimental data by the NuFit group in October 2021 [60] gave the values shown in Table 1.1. All the parameters appear to have non-zero values, implying that mixing between all neutrino flavours is possible. Furthermore, $|\Delta m_{21}^2| \ll |\Delta m_{31}^2| \sim |\Delta m_{32}^2|$, meaning there are two distinct ‘scales’ in L/E that are sensitive to neutrino oscillations: $L/E \sim 10 \text{ km MeV}^{-1}$ and $L/E \sim 0.3 \text{ km MeV}^{-1}$.

In addition to this information, we also know from solar data that the sign of Δm_{21}^2 must be positive, due to the MSW Effect: this will be explained in the next section. However, the same cannot be yet said for Δm_{31}^2 . This leads to two possible scenarios of the ordering of the neutrino mass states. If the sign of

Δm_{31}^2 is positive, then $m_{\nu_1} < m_{\nu_2} < m_{\nu_3}$, known as the Normal Hierarchy (NH). Alternatively, $m_{\nu_3} < m_{\nu_1} < m_{\nu_2}$, known as the Inverted Hierarchy (IH).

Neutrino Oscillations in Matter

Considering only neutrino oscillations in vacuum is insufficient to understand the results of certain neutrino experiments, especially those detecting high-energy solar neutrinos, such as those from the ${}^8\text{B}$ chain. Using Eq. 1.10 for the case of solar neutrinos, one finds that $X_{ij} \gg 1$ for all i, j , so all effects of neutrino oscillations in the vacuum should be washed out. This leads to an expected electron neutrino survival probability for solar neutrinos of:

$$P_{ee} = 1 - 2 \sum_{i < j} |U_{ei} U_{ej}^*|^2 \quad (1.11)$$

$$= 1 - \frac{1}{2} \sin^2(2\theta_{13}) - \frac{1}{2} \sin^2(2\theta_{12}) \cos^4(\theta_{13}) \quad (1.12)$$

$$= 0.55, \quad (1.13)$$

using the parameters in Table 1.1. This survival probability value appears somewhat consistent with those measured in low-energy solar neutrino experiments such as SAGE and GALLEX/GNO, but not with experiments with higher energy thresholds such as Homestake, Super-Kamiokande, or SNO.

The resolution of this apparent problem is that the effect of matter on neutrino oscillations have not been considered. When neutrinos travel through a medium, the electrons, protons, and neutrons within that medium interact weakly with those neutrinos, leading to coherent forward elastic scattering. The resulting phenomenon is known as the *MSW effect*, after its discovery by Mikheyev, Smirnov, and Wolfenstein [61, 62].

One can show [63] that there is an effective potential due to the CC weak interactions from the electrons within the medium of the form $V_{CC}(x) = \sqrt{2}G_F n_e(x)$,

where $n_e(x)$ is the electron number density at a position x . This effective potential is only felt by electron neutrinos, because of the additional CC interaction possible, as seen in Fig. 1.3. All neutrino flavours experience NC interactions identically, and so the effective potential generated by NC interactions from electrons, protons, and neutrons can be ignored in what follows. The effective potential modifies the Hamiltonian, and therefore by the Schrödinger Equation the propagation of the neutrino wavefunctions are modified.

The resulting dynamics of the MSW effect for 3- ν oscillations in a general medium can become quite complex. However, for the case of solar neutrinos two simplifying assumptions can be made that make many of the equations far more tractable: a full discussion can be read in e.g. [64]. Firstly, for solar neutrinos it can be shown that $A_{CC} = 2EV_{CC} \ll \Delta m_{31}^2$, leading to the evolution of the ν_3 state decoupling from the ν_1 and ν_2 states. There exists a new basis in which the Hamiltonian is diagonalised, known as the matter eigenstate basis, leading to new effective oscillation parameters:

$$\begin{aligned}\tan 2\theta_{12}^M &= \frac{\tan 2\theta_{12}}{1 - \frac{A_{CC}}{A_{res}}}, \\ \Delta m_{M,21}^2 &= \Delta m_{21}^2 \sqrt{\sin^2 2\theta_{12} + \cos^2 2\theta_{12} \left(1 - \frac{A_{CC}}{A_{res}}\right)^2}, \\ A_{res} &= \frac{\cos 2\theta_{12} \Delta m_{21}^2}{\cos^2 \theta_{13}}.\end{aligned}\quad (1.14)$$

A_{res} is the value of A_{CC} at which a resonance occurs, leading to the effective mass splitting $\Delta m_{M,21}^2$ being minimised. In the core of the Sun, this resonance occurs at an energy of $E \sim 2$ MeV. If $A_{CC} \ll A_{res}$, then the effective oscillation parameters reduce back to the vacuum oscillation parameters, as expected. If instead $A_{res} \ll A_{CC} \ll \Delta m_{31}^2$, then $\theta_{12}^M \rightarrow \pi/2$. In this case, ν_e that are created in the medium are completely driven into the $|\nu_2^M\rangle$ effective mass eigenstate.

Because the values of the effective oscillation parameters are a function of the electron density of the medium, calculating the full impact of the MSW effect of a medium with strongly-changing density can be challenging. However, if the rate of change of electron density is slow enough as a function of distance, a second approximation can be made: this is known as the *adiabatic approximation*. For solar neutrinos, this approximation has been shown to be valid assuming the SSM and the measured values of the oscillation parameters. Under this approximation, there are no transitions between effective mass eigenstates as the neutrinos propagate. Therefore, an electron neutrino with energies $\gg 2 \text{ MeV}$ will have been produced in the $|\nu_2^M\rangle$ state in the Sun's core, and then smoothly transformed into the equivalent $|\nu_2\rangle$ mass eigenstate once it has reached the Sun's surface. This state then travels through space without any further oscillations occurring, because the neutrino has been transformed into a mass eigenstate. For such a neutrino, the detection probability becomes simply:

$$P(\nu_2 \rightarrow \nu_e) = |\langle \nu_e | \nu_2 \rangle|^2 = |U_{e2}|^2 \quad (1.15)$$

$$= \sin^2 \theta_{12} \cos^2 \theta_{13} \quad (1.16)$$

$$= 0.30, \quad (1.17)$$

using the oscillation parameters from Table 1.1. This survival probability is almost half of the value seen in Eq. 1.11.

At intermediate energies, $E_\nu \sim 2 \text{ MeV}$, solar neutrinos are only partly driven into the $|\nu_2^M\rangle$ eigenstate, leading to a survival probability that is somewhere in-between the two extremes: $0.30 \leq P_{ee} \leq 0.55$. The value of P_{ee} for a given neutrino will be dependent on both the neutrino's energy and its location of production in the Sun: a neutrino that has to travel through regions of greater electron density will drive the effective mixing angle θ_{12}^M larger.

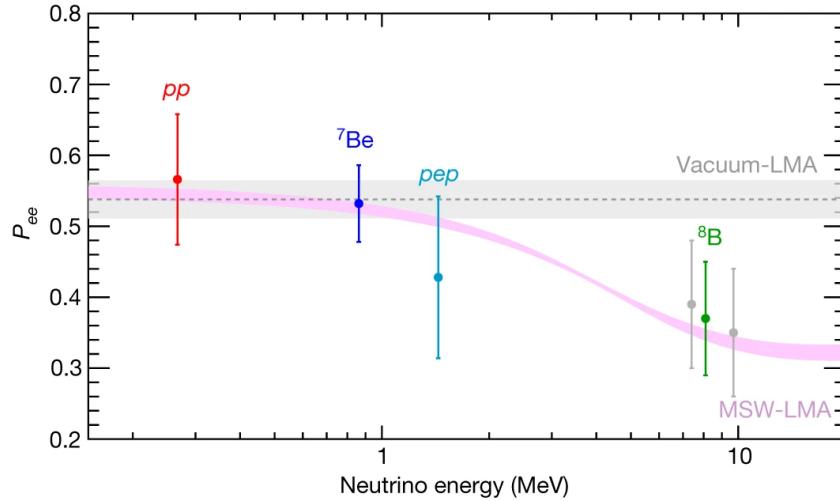


Fig. 1.5: Measured survival probability versus mean neutrino energy for each solar neutrino type, as observed by Borexino. Taken from [29]. For comparison, the expected survival probability due to vacuum oscillations and the MSW effect are shown as grey and pink bands, respectively.

Looking at data, Fig. 1.5 shows the measured values of P_{ee} for a number of different types of solar neutrino, each with their own characteristic energy spectrum. Also shown is the expectation after considering the MSW effect: as can be seen, the solar neutrino data appears consistent with this model.

Important to note is that the resonance phenomenon of the MSW effect requires a positive sign for Δm_{21}^2 to occur, given that $0 \leq \theta_{12} \leq \pi/4$. If $\Delta m_{21}^2 < 0$, then in the case where $|A_{CC}| \gg |A_{res}|$, θ_{12}^M will be driven to 0 instead of $\pi/2$. This would lead to the survival probability of solar neutrinos increasing at higher energies, entirely counter to what is seen in data. Because of this, the sign of Δm_{21}^2 is known to be positive, as mentioned in the previous section.

1.2.3 The Origins of Neutrino Mass

Having seen a wide variety of experiments over many decades observe neutrino oscillations, and the phenomenology that describes them requiring at least two neutrino mass states to be non-zero, a critical question is how neutrino masses can be included into the SM. There are two different approaches to doing so. If

neutrinos are given a *Dirac mass term*, then a term similar to the one in Eq. 1.2 is added, now using right-handed neutrino terms ν_j^c :

$$-\mathcal{L}_{\text{Dirac}} = \sum_{i,j} y_{ij}^\nu \bar{L}^i C H \nu_j^c + \text{h.c.} \quad (1.18)$$

$$\rightarrow \sum_{i,j} m_{ij}^\nu \bar{\nu}_{i,L} \nu_j^c + \text{h.c..} \quad (1.19)$$

The 3×3 matrix y_{ij}^ν describes the Yukawa coupling strengths between the neutrinos and the charge conjugate of the Higgs doublet, CH . After SSB, a neutrino mass matrix is obtained $m_{ij}^\nu = \frac{v}{\sqrt{2}} y_{ij}^\nu$, which can be related to the PMNS mixing matrix [19]. The theoretical downsides of this approach are two-fold: firstly, because $v = 246.22 \text{ GeV} \gg m_{\nu_i} \sim \mathcal{O}(10^{-2} \text{ eV}^2)$, this requires fine-tuning of the Yukawa coupling parameters down to values $\sim \mathcal{O}(10^{-14})$.

A second assumption needed for neutrinos to be Dirac particles is the existence of right-handed, ‘sterile’ neutrinos. These are so-called because their right-handed nature precludes them from interacting via any of the three main fundamental forces of particle physics. Sterile neutrinos would only be able to interact with the rest of the SM either via perturbative loops (creating a non-zero neutrino magnetic moment, for example), or the above mass term of the Lagrangian; this implies that neutrinos could oscillate into a sterile neutrino state. The LSND and MiniBooNE short-baseline neutrino experiments have seen excesses at low energies that could be explained by oscillations of a sterile neutrino with a mass splitting of $\Delta m_{41}^2 \sim \mathcal{O}(1 \text{ eV}^2)$ [65, 66]; however, these results seem at odds with recent results by the MicroBooNE Collaboration [67].

An alternative approach to generating neutrino masses without needing to posit the existence of right-handed neutrino states is through a *Majorana mass term*:

$$\mathcal{L}_M = \frac{1}{2} \sum_{i,j} m_{ij}^\nu \nu_{i,L}^T C \nu_{j,L} + \text{h.c.}, \quad (1.20)$$

where there is now no longer a coupling to the Higgs field, and instead the charge conjugate of the left-handed neutrino states is used. This term breaks SM gauge symmetry, so in theories Beyond the Standard Model (BSM) that want to include such a term typically introduce a higher-order term that reduces after SSB down to the Majorana term [68]. Many of these BSM theories, for example ones that include a so-called ‘Seesaw Mechanism’, also have a means of explaining why neutrinos have such light masses, without having to resort to ‘unnatural’ Yukawa coupling strengths [69]. For this term to exist neutrinos must be a ‘Majorana particle’, in which they are their own antiparticle, expressed mathematically as $\nu^c = \nu$. This is named after Ettore Majorana, who realised that a mass term such as the above could exist when neutrinos are their own antiparticles [70].

The existence of this Majorana mass term would have major consequences, beyond just allowing for neutrino masses. Crucially, the term violates lepton number, as it allows for neutrinos to annihilate one another. Fig. 1.7a shows an example of a process which violates lepton number by two, as two electrons have been created without any anti-leptons. Because all other terms in the SM conserve the ‘accidental’ symmetry of lepton number conservation, any evidence of lepton number violation with neutrinos can provide strong evidence that neutrinos are Majorana particles.

One prominent search mode for determining whether neutrinos have a Majorana mass term is by looking for *neutrinoless double beta decay*, $0\nu\beta\beta$. This is a variant of the radioactive decay known as *two-neutrino double beta decay*, $2\nu\beta\beta$, and was first hypothesised by Wendell H Furry [71]. $2\nu\beta\beta$ is a nuclear process theorised by Maria Goeppert-Mayer [72], in which two β -decays occur simultaneously in one nucleus, generating two electrons and two electron anti-neutrinos. This is only significant in the subset of isotopes for which $2\nu\beta\beta$ is energetically allowed, but the usual single β -decay is not (or indeed any other form of nuclear decay).

One example of an isotope capable of $2\nu\beta\beta$ is ^{130}Te . Fig. 1.6 shows the mass excesses of the nuclear ground state energy levels for isotopes in the isobar $A = 130$. As can be seen, ^{130}Te is not capable of β^- -decay to ^{130}I , but decay via $2\nu\beta\beta$ down to the stable isotope ^{130}Xe is possible. This process has been observed by the CUORE experiment [73], and $2\nu\beta\beta$ has been similarly observed in a number of other isotopes such as ^{76}Ge [74], ^{136}Xe [75], and ^{150}Nd [76].

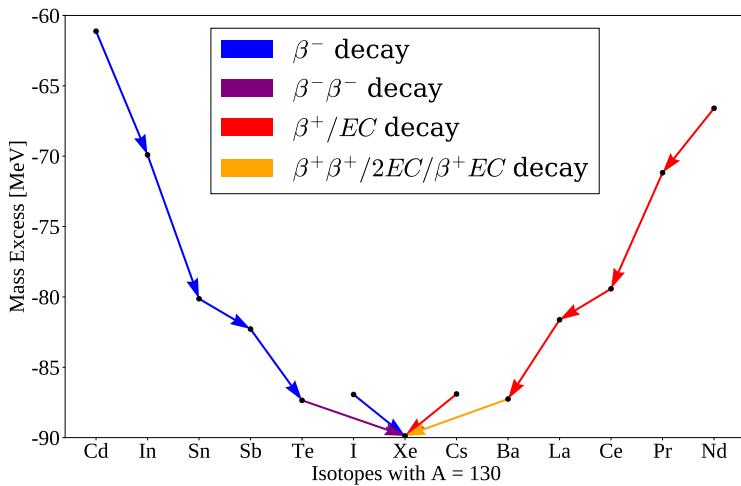


Fig. 1.6: Masses excesses of the $A = 130$ isobar, from [77]. The allowed weak decays between isotopes are shown by coloured arrows; note how both ^{130}Te and ^{130}Ba can only decay via second-order weak processes.

Unlike $2\nu\beta\beta$, $0\nu\beta\beta$ would emit no neutrinos during the decay, and instead a virtual anti-neutrino emitted by one nucleon would be captured on another as a neutrino. Fig. 1.7a shows a Feynman diagram for this process. A theorem by J. Schechter and J.W.F. Valle [78] says that, as long as the weak interaction is governed by some form of local gauge theory, any observation of $0\nu\beta\beta$ guarantees the existence of a Majorana mass term for neutrinos.

The lack of neutrinos generated in $0\nu\beta\beta$ compared to $2\nu\beta\beta$ enables a method for distinguishing between the two processes. In $2\nu\beta\beta$, the energy of the decay is shared between the two electrons and two anti-neutrinos generated. Given that the anti-neutrinos rarely ever interact, the observed energy of the event will come only

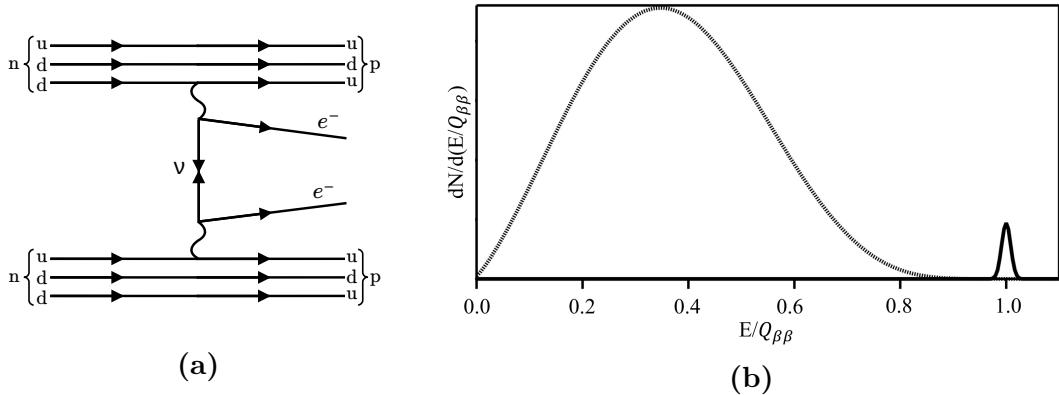


Fig. 1.7: (a): Feynman diagram for $0\nu\beta\beta$ decay. (b): Sketch of the energy spectra for $2\nu\beta\beta$ (dashed line) and $0\nu\beta\beta$ decays (solid line). The Q -value of the decay is written here as $Q_{\beta\beta}$. Taken from [79].

from the electrons (ignoring the negligible kinetic energy of the daughter nucleus). This leads to a broad observed energy spectrum. In comparison, with $0\nu\beta\beta$ all the decay energy is passed onto the electrons, and so the observed energy spectrum will be a thin peak at the Q -value of the decay. This is shown schematically in Fig. 1.7b. As can be seen, $0\nu\beta\beta$ can be searched for by looking for an excess of radioactive decay events generated by a $2\nu\beta\beta$ -decaying isotope at the Q -value.

No evidence of $0\nu\beta\beta$ has been seen at the time of writing. However, numerous experiments have been searching for the decays in a variety of isotopes, and more are under construction or being planned. A summary of the current state of searches with the most prominent isotopes that could theoretically allow $0\nu\beta\beta$ is shown in Table 1.2. In absence of an observation, experiments report a limit on the minimum possible half-life of $0\nu\beta\beta$ that is consistent with the non-observation, $T_{1/2}^{0\nu\beta\beta}$. If $0\nu\beta\beta$ is observed, the measured half-life can be used to help determine the neutrino masses, through the formula:

$$\frac{1}{T_{1/2}^{0\nu\beta\beta}} = \frac{|m_{\beta\beta}|^2}{m_e^2} G_\nu |\mathcal{M}_{0\nu\beta\beta}|^2, \quad (1.21)$$

where G_ν and $\mathcal{M}_{0\nu\beta\beta}$ are the phase space factor and matrix elements of the decay, and $m_{\beta\beta}$ is known as the ‘effective $0\nu\beta\beta$ mass’, defined as:

$$m_{\beta\beta} = \sum_{i=1}^3 U_{ei}^2 m_{\nu_i}. \quad (1.22)$$

Isotope	$T_{1/2}^{0\nu\beta\beta}$ [years]	Experiment	
^{76}Ge	$> 1.8 \times 10^{26}$	GERDA	[80]
^{100}Mo	$> 1.5 \times 10^{24}$	CUPID-Mo	[81]
^{130}Te	$> 2.2 \times 10^{25}$	CUORE	[82]
^{136}Xe	$> 2.3 \times 10^{26}$	KamLAND-Zen	[83]

Table 1.2: Current best limits on the half-life for $0\nu\beta\beta$ decay, for a selection of isotopes. All limits given are for a 90% CL.

Chapter 2

The SNO+ Detector

The light-soaked days are coming.

JOHN GREEN

2.1 Detector Geometry

The SNO+ detector is a large, multi-purpose neutrino detector built in the SNOLAB underground laboratory near Sudbury, Canada. The main detector structure is taken from the Sudbury Neutrino Observatory (SNO) [40], which can be seen in Fig. 2.1. SNO+ is an optical detector: light is generated within a central spherical detector medium, with those photons being detected by an array of 9362 inward-facing Photomultiplier Tubes (PMTs). The properties of physics events are then estimated by looking at the number of photons that were detected, as well as their timing and spatial distributions. Because of this, the detector must be optimised to achieve a high, stable, and well-understood detection efficiency of photons, as well as a timing resolution of detected photons $\mathcal{O}(1\text{ ns})$. A more detailed description of how physics events get detected in SNO+ is given in Section 2.3.



Fig. 2.1: 3D model of the SNO+ detector [84].

The main detector medium of SNO+ changes depending on the phase of the experiment; specifics are given 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 the PMTs. 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 also affixed to the outside of the PSUP, allowing for the effective vetoing of cosmic ray muons.

Directly above the detector is the Deck, upon which all the detector electronics are kept. Access within the AV for calibration tools and filling is possible through the acrylic ‘neck’ on top of the AV. Full details of the design of the current detector can be found in [84].

2.2 Experimental Phases

SNO+ was designed to fulfil a number of physics goals over multiple ‘phases’ of the detector’s lifetime. The phases are distinguished by the medium that fills the AV. The first main phase (after a brief **Air Fill Phase** used only for detector commissioning) was that of the **Water Fill Phase**, with data taken between May 2017 and July 2019. This was used to perform fundamental optical calibrations of the detector [85], measurements of the solar neutrino flux [86], observation of neutrino oscillations in reactor anti-neutrinos [87], and searches for nucleon decay [88, 89].

After this, the detector was filled with 780 tonnes of a type of liquid scintillator known as linear alkylbenzene (LAB), mixed with the fluor 2,5-diphenyloxazole (PPO). More information on the physics of scintillators can be found in Section 2.3.1. Filling of the LAB-PPO cocktail had to be paused in March 2020 due to the COVID-19 pandemic, leading to the detector having its bottom half still filled with UPW, and the top half filled with LAB and PPO at 0.5 g/L. This impromptu phase became known as the **Partial Fill**, and allowed for some creative analyses to be performed: an initial neutrino oscillation analysis from reactor anti-neutrinos [90], as well as the first ever observation of directionality in a high light yield scintillator [91, 92]. Eventually, filling of the detector with liquid scintillator completed in May 2021. At that point, the concentration of PPO in the detector was at 0.6 g/L, markedly below the target level of 2.0 g/L. A further ‘PPO top-up’ campaign then proceeded, finishing in April 2022 with a final concentration of 2.2 g/L PPO. Thus began the **Scintillator Phase** of the experiment, which continues on during the time of writing. The main goals for this phase include a number of solar neutrino analyses (including the one described in Chapter 6), a precision measurement of the neutrino oscillation parameter

Δm_{21}^2 using reactor anti-neutrinos [90], further calibrations of the detector, and measurements of the various backgrounds.

Two further chemicals are being added to the scintillator cocktail at the time of writing. The antioxidant butylated hydroxytoluene (BHT) has been added in July 2023 to capture any free-radicals within the liquid scintillator, hopefully preventing any oxidation reactions that could lead to the ‘yellowing’ of the scintillator, a degradation of its optical properties. The addition of BHT is not expected to directly impact the detector’s optics in any substantial way. However, the other substance to also be added, 1,4-Bis(2-methylstyryl)benzene (bis-MSB), will impact the optics. Bis-MSB acts as a ‘wavelength-shifter’ which enables the scintillator cocktail to transmit light with a greater overall detection efficiency — more on the details of this in Section 2.3.1.

Finally, in the near future the detector will be loaded with Tellurium for the **Tellurium Phase**, allowing for the flagship analysis of the experiment to begin: neutrinoless double beta decay. In order to load Te within the liquid scintillator in a stable manner, a chemical loading process has been developed, as described in [93]. The Te starts within Te(OH)_6 (telluric acid, otherwise known as TeA), which after purification will be reacted with 1,2-butanediol (BD) via heating and addition of N,N-Dimethyldodecylamine (DDA), which acts as a stabiliser. What results is tellurium-loaded scintillator, TeLS.

2.3 Detecting and Recording an Event in SNO+: A Journey

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

2.3.1 Particle Interactions with Matter

All observable physics events within the detector begin by the generation of some form of ionising radiation: α , β^\pm , p , μ or π . These can be created via numerous processes, both exciting (e.g. $0\nu\beta\beta$ or interactions of neutrinos) and annoying (e.g. decay of background radioisotopes): see Section 6.2 for some of them. Regardless of their origin, these particles begin propagating through the detector, and interacting with the detector medium. A number of mechanisms then allow 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 [94]:

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

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.

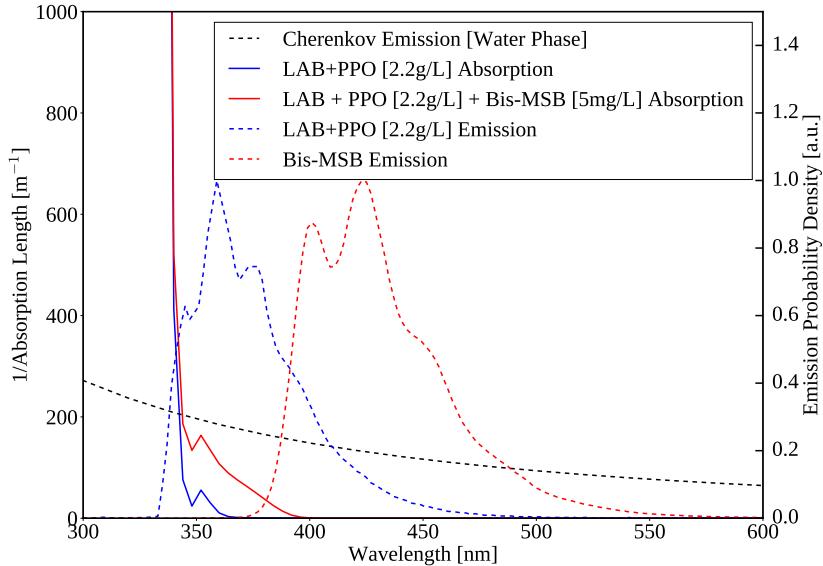


Fig. 2.2: Comparison of the SNO+ detector media’s emission and absorption properties, versus optical phase [95, 96].

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 atoms nearby a moving charged particle can lead to the generation of optical-wavelength light, in a process known as **scintillation**. 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’ [97]. A major example of these π-bonds are found in benzene rings, which are present in LAB, PPO, and bis-MSB.

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 [97] 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.

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-PPO 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 LAB-PPO [98]. A plot from R. Hunt-Stokes showing this fit between data and simulation is shown in Fig. 2.3.

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 LAB-PPO [98, 99].

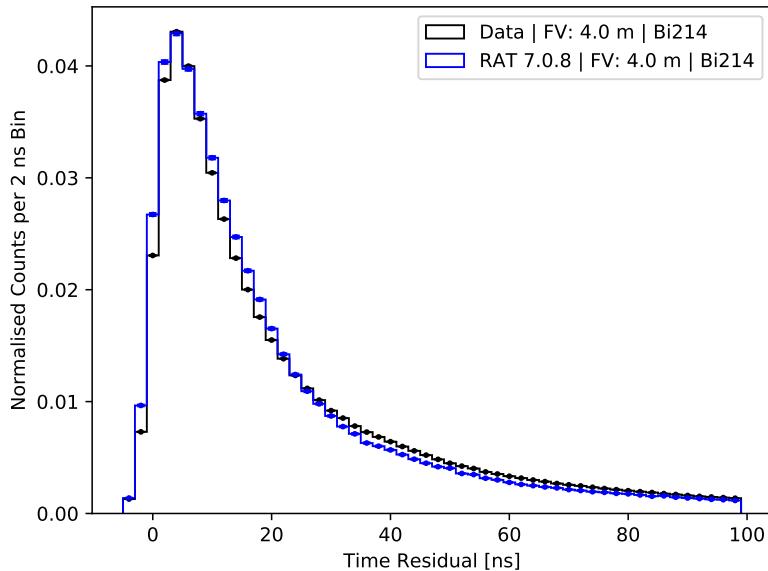


Fig. 2.3: Comparison between the observed emission time distribution of electrons from tagged ^{214}Bi β -decays and a production of matching simulated events, after fitting the timing constants. Adapted from [98]. “FV” corresponds to the Fiducial Volume used in this plot; “RAT 7.0.8” is the version of the simulation software used to compare to data.

In SNO+, a separate scintillating component, PPO, has been added to the LAB. When an LAB molecule is excited, that energy can be transferred to a PPO molecule through what is known as a ‘non-radiative transfer’. In short, this transfer of energy occurs not through the emission and absorption of optical photons, but through the coupling of the molecules’ electric dipoles during a collision. The now-excited PPO molecule can then de-excite to emit scintillation

light. The additional pathway that PPO provides substantially increases the light yield of the scintillator.

The compound bis-MSB is also being added to the scintillator cocktail at the time of writing. This is a ‘wavelength-shifter’: scintillation light at short wavelengths is absorbed, and then re-emitted at longer wavelengths, where the detection efficiency of the PMTs is greatest (~ 420 nm). More on the properties of the PMTs in SNO+ can be found in Section 2.3.3. This shift in wavelength further boosts the measured light yield of the scintillator within the detector. The net effect of the three scintillating components 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 [100]:

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

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”. In the 2.2 g/L LAB-PPO scintillator currently within SNO+, S and k_{Birks} are measured to be approximately $14,000 \gamma/\text{MeV}$ and $0.077 \text{ mm MeV}^{-1}$, respectively [101]. For minimum-ionising particles such as a 6 MeV electron $\frac{dE}{dx} \approx 2 \text{ MeV cm}^{-1}$ [102],

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. For example, α -particles are generated at 5.304 MeV from the decays of ^{210}Po nuclei [77]. However, these events generate light equivalent to a 0.45 MeV event in the detector.

2.3.2 Optical Processes

Once optical-wavelength photons have been created within the detector, various processes can then occur that can hinder their 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.

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 generates 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 different wavelength than the incident radiation: in which case, the scattering was *inelastic*. However, this latter type of scattering, also known as Raman scattering, occurs negligibly in SNO+.

The simplest form of elastic optical scattering is known as *Rayleigh scattering*, after the initial formulation by Lord Rayleigh [103], and relies on the following assumptions of a system of particles:

- the particles are an ideal gas, i.e. there are negligible inter-molecular forces;
- the particles are spherical;
- the particle radii are much less than the wavelength of the incident light;
- the induced dipole moments of the particles can be established on a timescale much less than the period of the electromagnetic wave.

If the latter two assumptions are lifted, one ends up with the more general *Mie Theory*, first described by Gustav Mie [104] and Ludvig Lorenz [105].

The optical media within SNO+ are all liquids or solids, and so the first assumption above of negligible inter-molecular forces does not at all hold. A different theory was developed by Einstein [106], Smoluchowski [107], and Cabannes [108], in which light scatters off of the local charge density fluctuations that naturally are present in a medium because of the thermal motion of molecules. This theory predicts that the *Rayleigh ratio* R , the fraction of the incident light that gets scattered at a 90° angle, per unit volume per unit solid angle, is given by [109]:

$$R = \frac{\pi^2}{2\lambda^4} \left[\rho \left(\frac{\partial \varepsilon}{\partial \rho} \right)_T \right]^2 k_B T \kappa_T \frac{6 + 6\delta}{6 - 7\delta}. \quad (2.4)$$

Here, ρ is the density of the medium, $\left(\frac{\partial \varepsilon}{\partial \rho} \right)_T$ is the partial derivative of the dielectric constant ε with respect to a changing density assuming a constant temperature T , k_B is the Boltzmann Constant, κ_T is the medium's isothermal compressibility, and δ is the *depolarisation ratio* of the medium. This latter variable describes how anisotropic the medium's electric polarisability is — a medium with no anisotropy in the polarisability has $\delta = 0$. The $1/\lambda^4$ dependence indicates that short wavelengths of light will get scattered to a far greater extent than longer wavelengths.

Various alternative versions of this formula exist, converting $\rho \left(\frac{\partial \varepsilon}{\partial \rho} \right)_T$ into something more straightforward to measure via use of thermodynamical or empirical

equations — see [109, 110] for discussions. Also, in liquid mixtures, the total observed scattering can substantially exceed what would be expected from the sum of the individual components. This is because fluctuations in the dielectric constant can also be caused by fluctuations in the relative composition of the medium in a given volume element [111, 112].

Although R is the scattering quantity most easy to measure in bench-top laboratory experiments, for SNO+ two different properties are more relevant. The main observable 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 the Rayleigh scattering length is given by [113]:

$$l_{\text{Ray}} = \left[\frac{8\pi}{3} \frac{2 + \delta}{1 + \delta} R \right]^{-1}. \quad (2.5)$$

The other important feature of Rayleigh scattering is its angular dependence. The scattered intensity as a function of the scattered angle, $I(\theta)$, has an equation of the form:

$$I(\theta) \propto \left(1 + \frac{1 - \delta}{1 + \delta} \cos^2 \theta \right). \quad (2.6)$$

In-situ measurements of the scattering of the UPW were first made indirectly during SNO [114]. Subsequently, the scattering lengths were measured to be scaled down by a factor of $(1.28 \pm 0.05(\text{stat.}) \pm 0.14(\text{sys.}))$ by Esther Turner during the SNO+ water phase [115]. Ex-situ measurements of the Rayleigh scattering within LAB and LAB-PPO have also been made by groups in both the SNO+ and JUNO Collaborations [116–120], 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 LAB-PPO 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. The depolarisation ratios of both

UPW and LAB have been measured to be non-zero [119, 121]; currently, non-zero values of δ are not considered in simulations of SNO+.

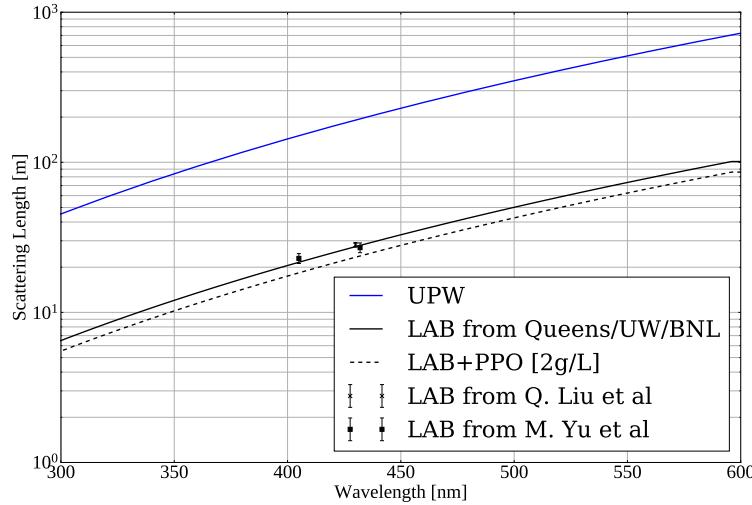


Fig. 2.4: Scattering lengths used in simulation for UPW, LAB, and 2 g/L LAB-PPO. The UPW shape is taken from indirect in-situ measurements in [114] with an additional divisive scaling factor from [115]. The LAB shape is taken from ex-situ measurements by SNO+ members at Queen’s University, University of Washington, and Brookhaven National Laboratory [116, 117]. An additional divisive scaling factor of 1.176 due to PPO was made by [118]. For comparison, measurements by members of the JUNO Collaboration for LAB are also shown [119, 120].

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

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

$$\frac{1}{l_{\text{ext}}} = \frac{1}{l_{\text{abs}}} + \frac{1}{l_{\text{Ray}}}. \quad (2.7)$$

In the water phase, the ‘Laserball’ calibration system was used by Ana Sofia Inácio to measure various optical properties of the detector, including the extinction lengths of the UPW and acrylic as a function of wavelength [79, 85]. Using the water phase scattering measurements made by E. Turner, Eq. 2.7 allowed for the estimation of the absorption lengths of these two materials, shown in Figure 2.5. Measurements of the extinction length in the scintillator phase is discussed in detail in Chapter 5.

Surface reflection and refraction

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

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

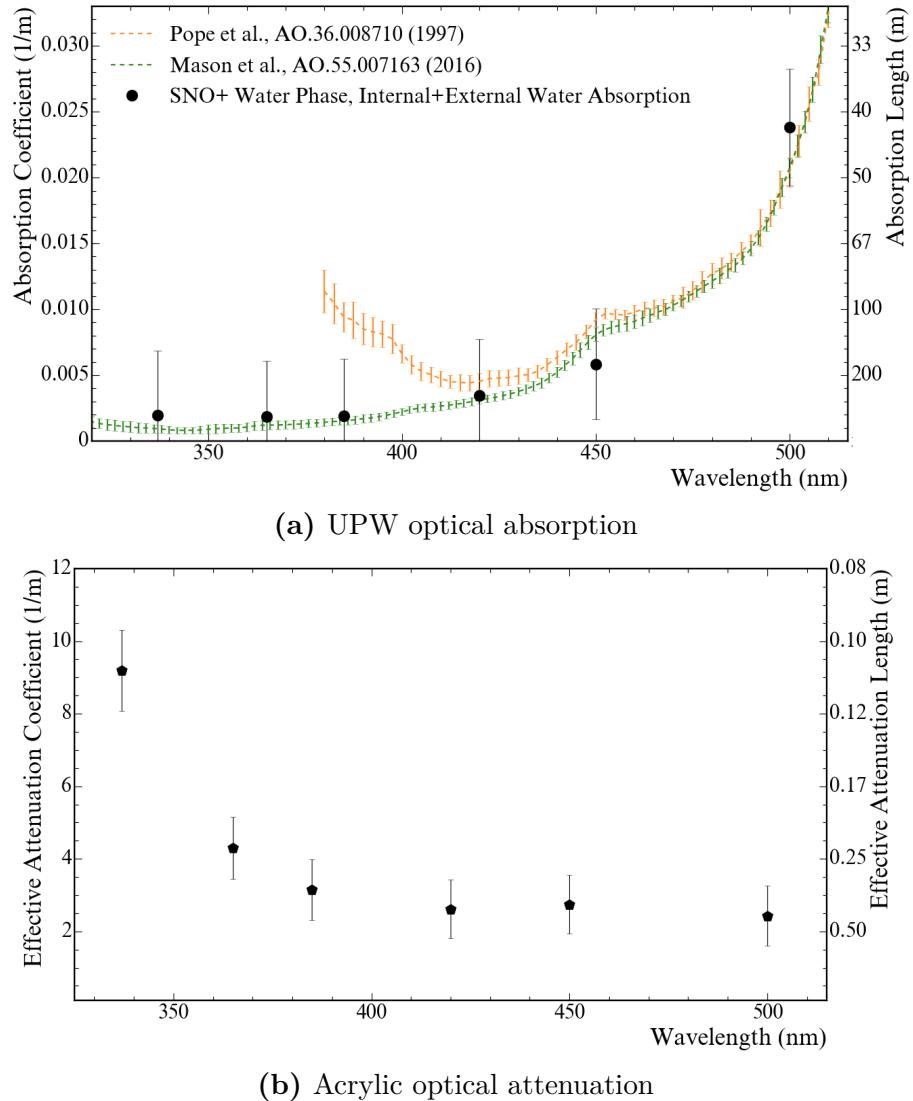


Fig. 2.5: Properties of the UPW and acrylic in the water phase, measured by A. S. Inácio in [79, 85].

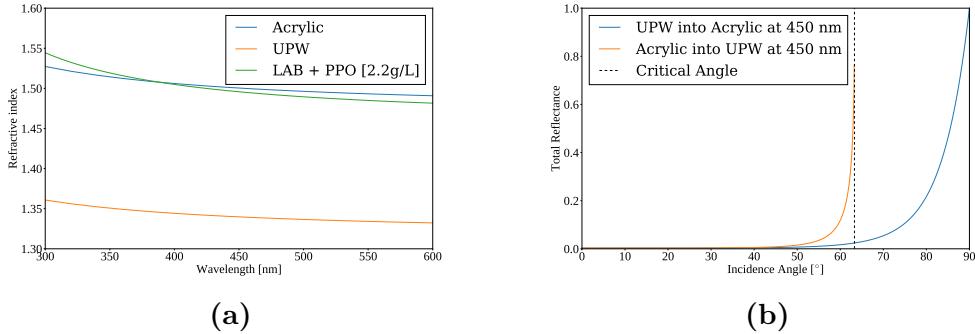


Fig. 2.6: (a): Refractive indices of acrylic, UPW, and LAB-PPO as a function of wavelength. The values for acrylic and UPW come from model fits to data made in SNO [114, 122], whereas those for LAB-PPO come from data taken in [123]. (b): Reflectance of an unpolarised beam of light at 450 nm going between UPW and acrylic.

The *Fresnel Equations* determine the reflectance of an interface [124]:

$$R_s = \left| \frac{n_1 \cos \theta_i - n_2 \cos \theta_t}{n_1 \cos \theta_i + n_2 \cos \theta_t} \right|^2, R_p = \left| \frac{n_1 \cos \theta_t - n_2 \cos \theta_i}{n_1 \cos \theta_t + n_2 \cos \theta_i} \right|^2, \quad (2.8)$$

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. In SNO+, there is no sensitivity of the PMTs to different polarisations, so what matters is 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 the detector is detection by a PMT. Almost all PMTs in SNO+ are of the Hamamatsu R1408 design [40]. The 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

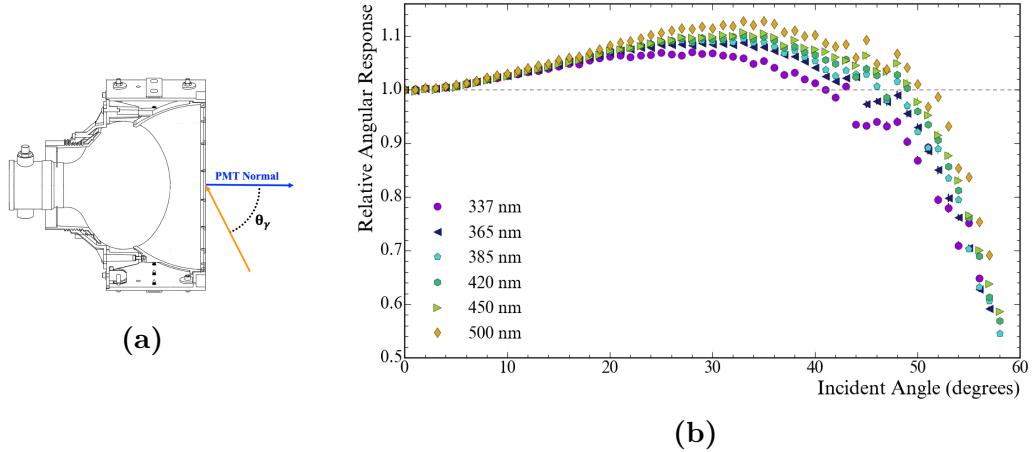


Fig. 2.7: (a): Diagram of the PMT and concentrator ‘bucket’ used within SNO+, showing also the definition of the incidence angle. (b): Plot of the measured relative angular response of the PMTs in SNO+, as a function of both incidence angle and wavelength. Both figures taken from [85].

maximise the collection efficiency of light emanating from within the AV, whilst minimising the collection efficiency of light outside the AV [125]. 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. In addition, the collection efficiency defines the probability that a generated photoelectron actually created a recorded signal. The combined measured efficiency of PMTs tested *ex-situ* for SNO can be seen in Fig 2.8. Once this photoelectron has been created, it is accelerated by an electric field within the PMT, onto the PMT’s first dynode. The natural spread in drift times is known as the ‘Transit Time Spread’ (TTS) of the PMTs: for SNO+, the RMS of the TTS for the R1408-type PMTs is ~ 1.7 ns [40]. The collision of the photoelectron with

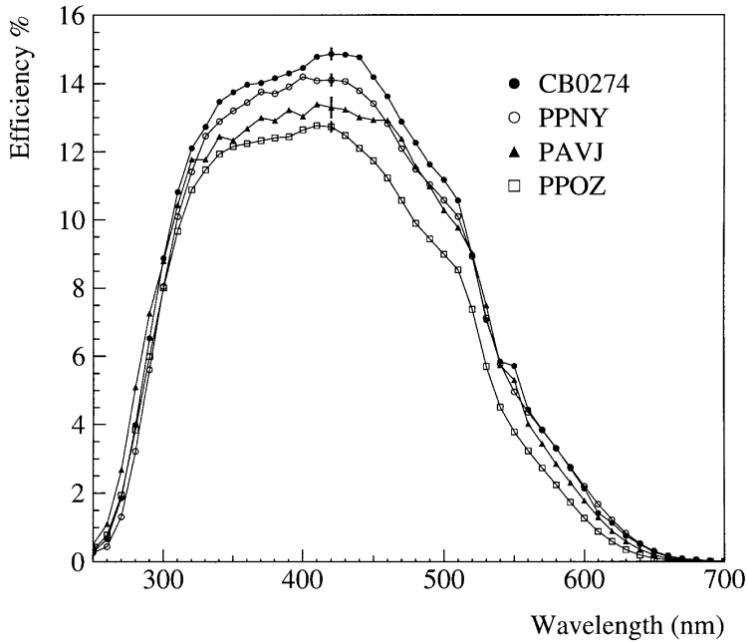


Fig. 2.8: Efficiencies of four R1408-type PMTs tested for calibration by [126].

the dynode generates further electrons, which collide with subsequent dynodes to generate a cascade that eventually produces an observable voltage signal.

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

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,

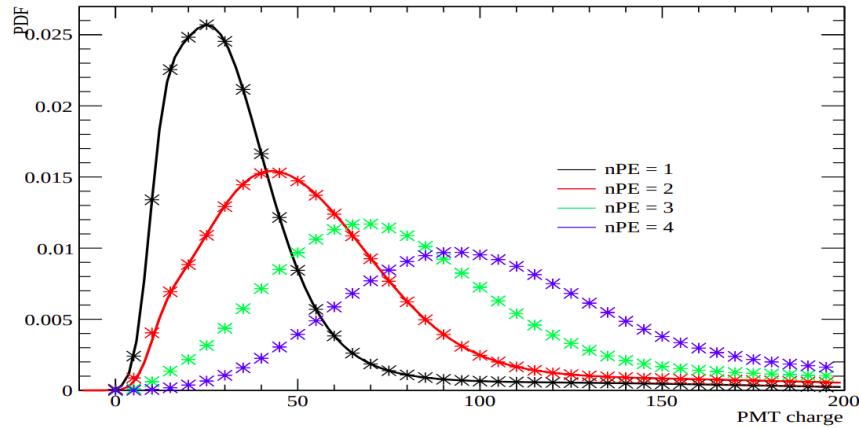


Fig. 2.9: Example charge spectra for a PMT as a function of the true npe generated. Figure taken from [127].

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 [84].

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 a ‘hit’ is said to have been detected 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 Converter (TAC). TAC also corresponds to the resulting analogue voltage on the output of the TAC being measured.
2. Begins integrating the total charge signal for that channel in three ways, known as QHS, QHL, and QLX. These correspond to using different integration times and gain settings.

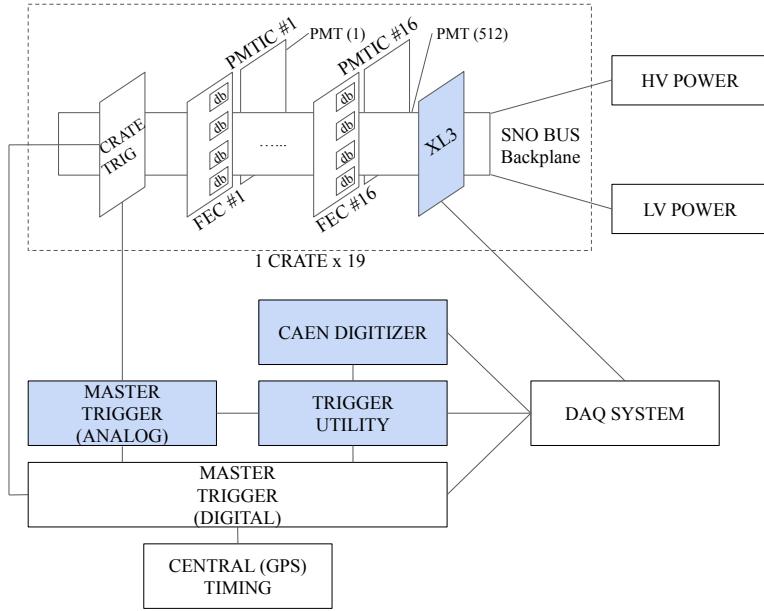


Fig. 2.10: Schematic of the front-end electronics used for data acquisition and triggering in SNO+, taken from and discussed in [84].

3. Generates trigger pulses for that channel for each available trigger type.

Three main trigger signals are the ‘N20’ (a square pulse for 20 ns), ‘N100’ (a square pulse for 100 ns), and ‘ESUMHI’ (a pulse copying the shape of the voltage signal for that channel).

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

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

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

2.3.5 Operation of the Detector

Control of the detector’s DAQ system is handled through a custom-built GUI known as ORCA [128]. 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 LAB-PPO 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 Detector Calibrations and 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.

2.4.1 Detector Monitoring

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

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

2.4.2 Electronic and PMT Calibrations

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

During an ECA, two main quantities are measured. Firstly, because of noise the integrated charge measured on each channel is offset by some amount. This offset, known as the ‘pedestal’, is recorded for each channel. The other quantity is the ‘time slope’ for each channel, which allows one to convert from the ADC TAC counts into an uncalibrated hit time of that channel’s PMT. Both 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 allows any channels with unusual behaviour to be spotted, 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 Laserball is an 11 cm quartz sphere, filled with small glass beads suspended in silicone gel, which diffuses laser light that is sent into it [85, 114]. The result is a near-isotropic light source that can be deployed within the detector in different positions, via use of a series of ropes. The TELLIE system 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 contain all the information of an event, but now the timing and charge information have been calibrated. It is this file type used in the optical calibration work of Chapters 3–5.

2.4.3 Energy and Optical Calibrations

The next stage of calibrating the detector is modelling its optical properties. These properties include all the processes covered in Section 2.3.2, such as scintillator emission, optical absorption, re-emission, and Rayleigh scattering. This is crucial, as it allows us to reconstruct information about events within the detector.

In addition to deployments of the Laserball (discussed in Section 2.3.2), two further calibration sources are used in SNO+ to measure properties of light propagation: AMELLIE and SMELLIE. These are the ‘Attenuation Module’ and ‘Scattering Module’ of the ELLIE calibration system. Like TELLIE, AMELLIE and SMELLIE consist of optical light sources that shine through optical fibres into the detector. The former uses LEDs from TELLIE, whilst the latter uses optical wavelength lasers. Despite the names both subsystems are similar enough

that they are both capable of measuring attenuation and scattering within the detector. More details about the SMELLIE hardware can be read in Chapter 3.

Another critical component of the detector to calibrate well is the energy response: given a specific amount of energy deposited in the water/scintillator, how many hits are observed? For this, a number of radioactive sources are used at a variety of energies. In the scintillator phase, there are three main sources. The first is an americium-beryllium (AmBe) source inherited from SNO [129], which contains ^{241}Am that α -decays. These α particles can be captured by the ^9Be within the source, leading to the emission of a neutron as well as production of a ^{12}C 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 (typically) in the detector, leading to a characteristic 2.2 MeV delayed γ being generated [84]. 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 [129], which is important for the analysis of antineutrino IBD events.

The second radioactive source is the ^{16}N deployable source, also originally used for SNO [130]. 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 [85]. Although external deployments during the scintillator phase with the Laserball, ^{16}N , and AmBe sources have been made, there have been no such internal source deployments. This is a result of the substantial work currently underway [131]

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 two deployed sources, a third 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 energy spectra that can be observed in SNO+, and can be used to calibrate the scintillator’s energy response. $^{214}\text{BiPo}$ events in particular have been used for energy scale calibration with the solar oscillation analysis, as discussed in Section 6.4.3.

2.4.4 Event Reconstruction

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

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

for a given PMT hit is defined as:

$$t_{res} = t_{\text{hit}} - t_{\text{TOF}} - t_{\text{emm}}, \quad (2.9)$$

where t_{hit} is the calibrated hit time of the PMT, t_{TOF} is the time one expects for light to travel directly from the reconstructed position to that PMT (the time-of-flight), and t_{emm} is the reconstructed emission time. The t_{res} distribution of hits in a given event can be very useful in understanding the physics of that event.

Whilst a number of algorithms have been developed for reconstructing position and time in SNO+, they all work on the same basic principle. Because of the 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.

A likelihood-based approach is currently used to reconstruct position and time in SNO+. 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 [132–134]. 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

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, this is accounted for when coordinating the position fitters.

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 [135].

The other critical piece of reconstructed event information is the event’s energy. In energy reconstruction, the particle is assumed to be an electron, and the energy corresponds to the kinetic energy of the electron. 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).

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 high energies, many PMTs will have had multiple photoelectrons generated, so merely using the `nhit` will give an underestimate of the true energy. Also, 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 [136, 137].

After position, time, and energy reconstruction, `ScintFitter` calculates a number of additional quantities, 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 6.2.

All Physics data runs, as well as certain calibration data such as AmBe and ^{16}N , have 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** [84]. 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 7.0.14, inclusive.

Chapter 3

The SMELLIE Calibration System

There's a certain Slant of light,

Winter Afternoons —

That oppresses, like the Heft

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 which can be fired through 15 optical fibres, whose endpoints are attached to the PSUP. A collimator is attached at the end of each fibre, narrowing the beam across the detector. A diagram of SMELLIE in the detector is shown in Fig. 3.1.

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

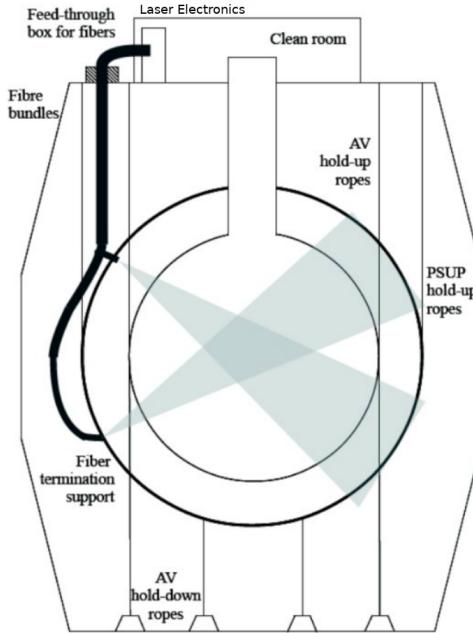


Fig. 3.1: Diagram of the SMELLIE calibration system within SNO+. Modified from [138].

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, one 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 observing the fraction of light emitted by the fibre that gets observed on the far side of 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. 2.7.

Both the scattering and absorption lengths of the detector medium impact the propagation of light from physics events, and hence which PMTs get hit along with the times of those hits. If incorrect values for these lengths are assumed in event reconstruction, 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 a larger fraction of photons is lost because not all absorbed light is re-emitted. Scintillator absorption naturally leads to a non-uniformity in the measured light yield as a function of radius; an incorrect absorption length would then lead to an incorrect model of this non-uniformity. 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 affect the position reconstruction and many classifiers.

A full description of the initial hardware setup of SMELLIE that was used during the air fill and early parts of the water fill phase can be read in [139, 140]. Since then, a series of hardware upgrades have been made, with [115, 141] covering the hardware status used in data taken throughout the water phase. Fig. 3.2 shows a timeline of the hardware upgrades as well as some of the calibration data taking campaigns performed using SMELLIE. The current layout of the SMELLIE

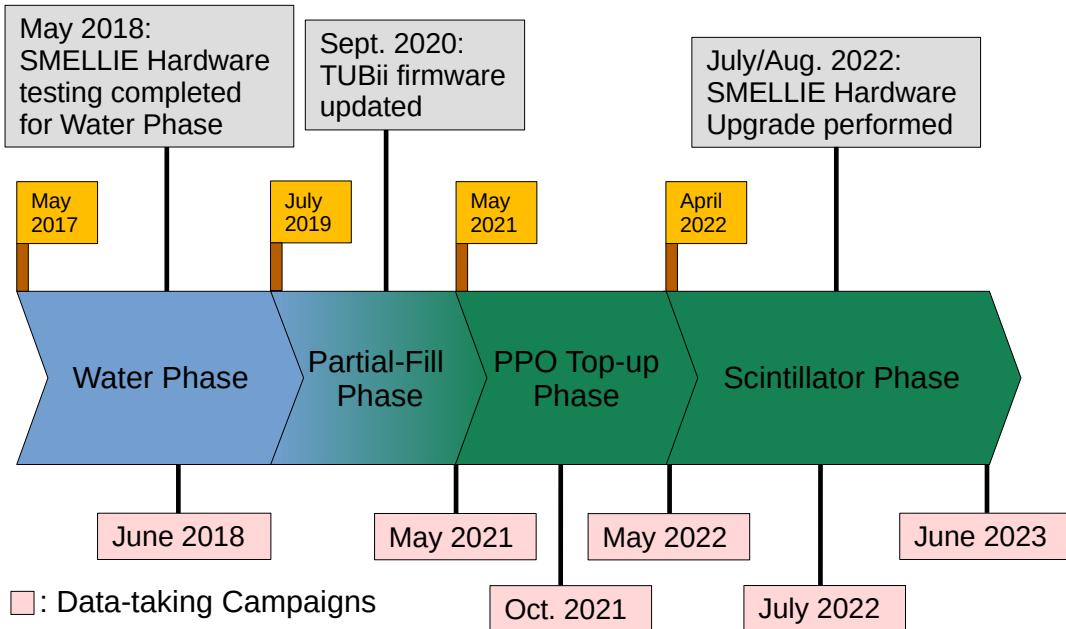


Fig. 3.2: A brief history of the SMELLIE hardware work and calibration data-taking campaigns of interest for this thesis.

hardware, showing the connections between each of the devices within the system, can be seen in Fig. 3.3. This chapter briefly summarises the current contents of the calibration system, along with descriptions of the major hardware changes made since the water phase.

3.1 Lasers

Fundamental to the SMELLIE calibration system are 5 optical-wavelength lasers. Four of these are fixed-wavelength pulsed-diode lasers from PicoQuant. These ‘PQ’ lasers each emit with a different narrow wavelength spectrum, peaking at 375 nm, 407 nm, 446 nm, and 495 nm. These are referred to as the PQ375, PQ407, PQ446, and PQ495 lasers, respectively. In addition to these lasers, a SuperK Compact laser made by NKT Photonics (hereafter referred to as the SuperK laser) is also used¹. Unlike the PQ laser heads, the SuperK is a super-continuum laser

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

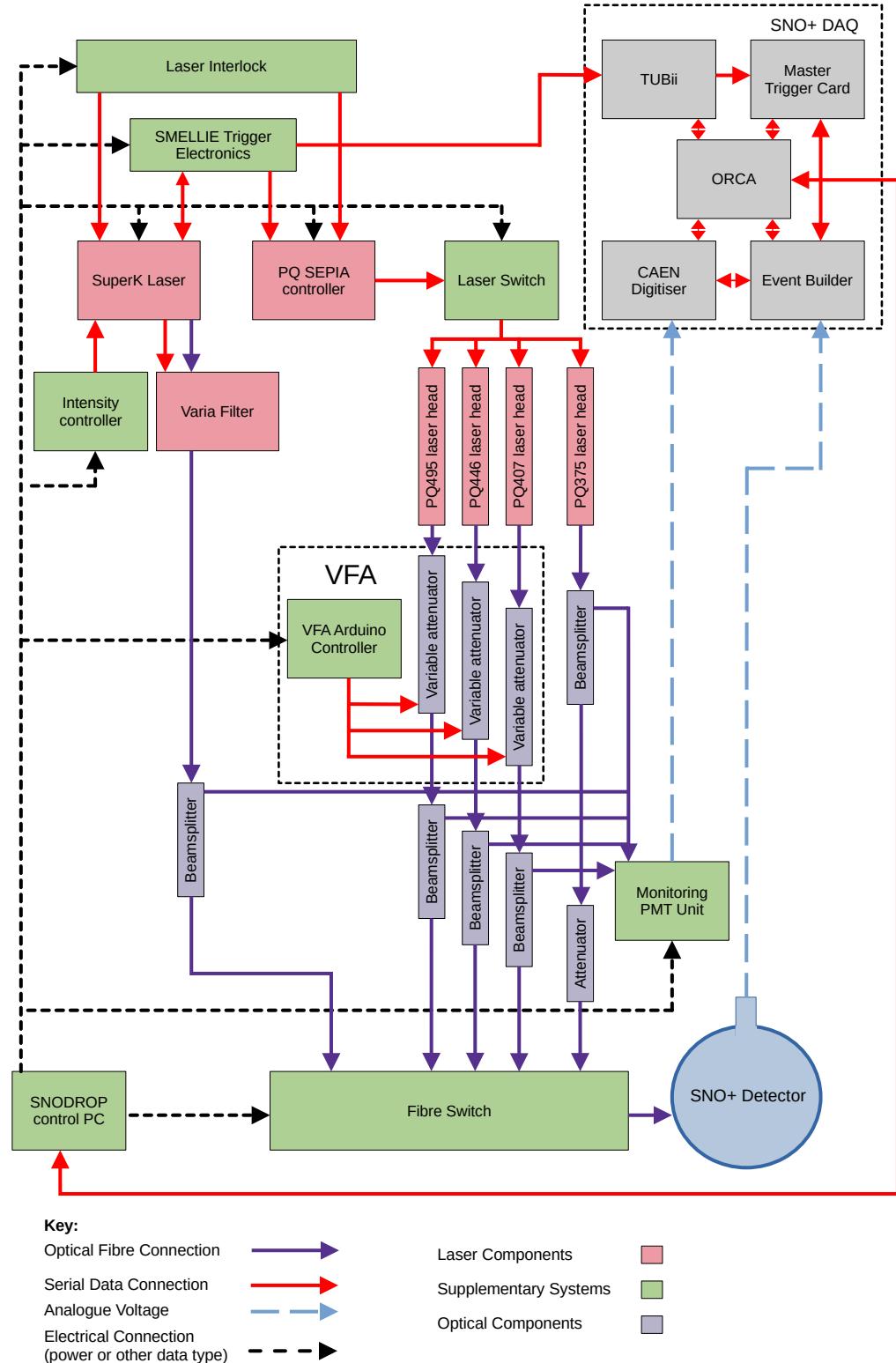


Fig. 3.3: Diagram of the connections between the various bits of SMELLIE hardware and the rest of the SNO+ detector, after the changes made in the Summer of 2022. Adapted from [115].

able to produce laser light over a very broad wavelength spectrum. Because the interest is almost always in determining optical properties at specific wavelengths, a variable bandpass filter also built by NKT Photonics known as the SuperK Varia has been included. This allows 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 of the laser pulses are essential for scattering analyses as they allow for precise predictions in knowing when photons should arrive at PMTs in the detector. It is much harder to produce such short pulses with LEDs, which is why SMELLIE uses lasers for its light generation, unlike the LEDs used for AMELLIE and TELLIE.

3.2 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. The emitted light needs to be intense enough to generate sufficient PMT hit statistics for an analysis, but too much light can overwhelm the detector’s DAQ. This intensity control 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. All light pulses get generated at the full power, but are 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. For both types of laser, the driving voltage is given as a number between 0 and 1000, where 1000 indicates the maximum allowed driving voltage for that laser. Problematically, the dependence of the raw output intensity of the PQ laser on

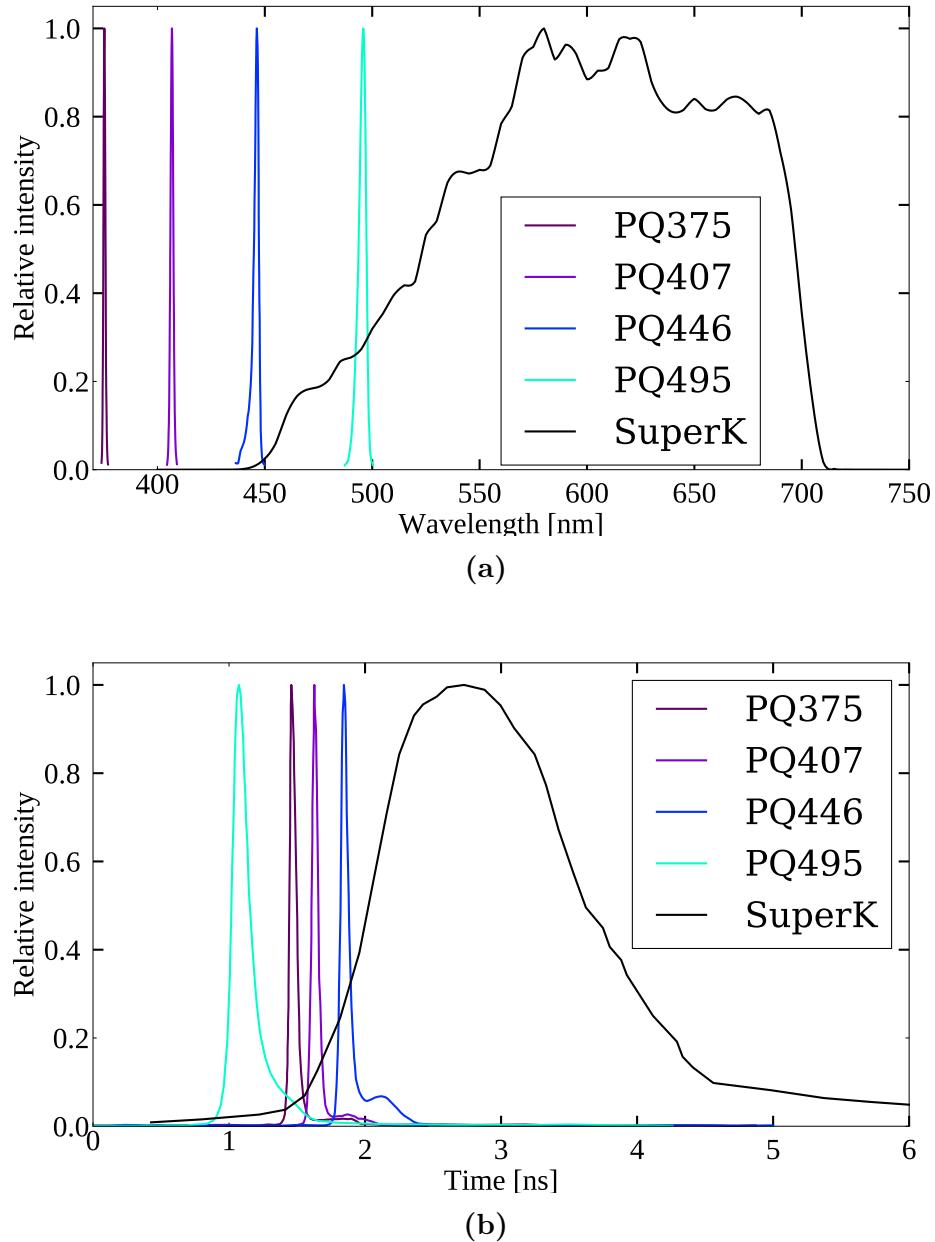


Fig. 3.4: Emission wavelength (a) and timing (b) spectra of the lasers used in SMELLIE. PQ spectral information taken from the manufacturer; SuperK information measured by J. Lidgard [142]. The SuperK wavelength spectrum shown is the full spectrum before the bandpass filter has been applied.

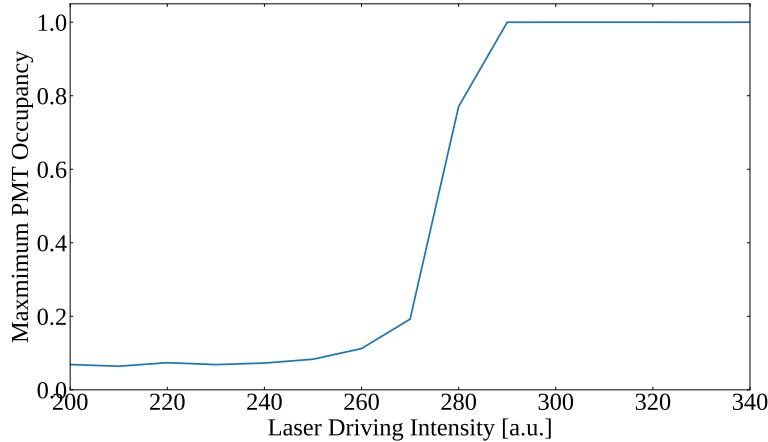


Fig. 3.5: Typical impact of driving voltage intensity for PQ446 laser on observed maximum PMT occupancy in the detector. Data taken on February 22nd, 2021.

this voltage is highly nonlinear. To demonstrate, consider the observed fraction of events in which a given PMT was hit, known as the PMT occupancy. The largest PMT occupancy in the detector (known as the ‘maximum PMT occupancy’) is shown as a function of driving voltage for SMELLIE events using the PQ lasers in Fig. 3.5.

For low driving voltages, the resulting maximum PMT occupancy is very small, and rises slowly. However, near the ‘lasing threshold’ the occupancy observed rapidly climbs. Above this threshold, some PMTs in the beamspot ‘saturate’, having an occupancy of 1. As will be seen in Chapter 4, information about the light intensity incident on a given PMT can be derived in a straightforward manner only if the light level is stable for a given set of data, and the occupancy on the PMT is below 1. When driving a laser near its lasing threshold the shot-to-shot variation in observed light intensity in the detector can also become substantial: see Fig. 3.6 for an example of this occurring.

During the water and scintillator phases up until Summer 2022, some data taken, especially using the PQ407 and PQ446 lasers, suffered from large shot-to-shot intensity variations. Throughout this period, after the light was generated by

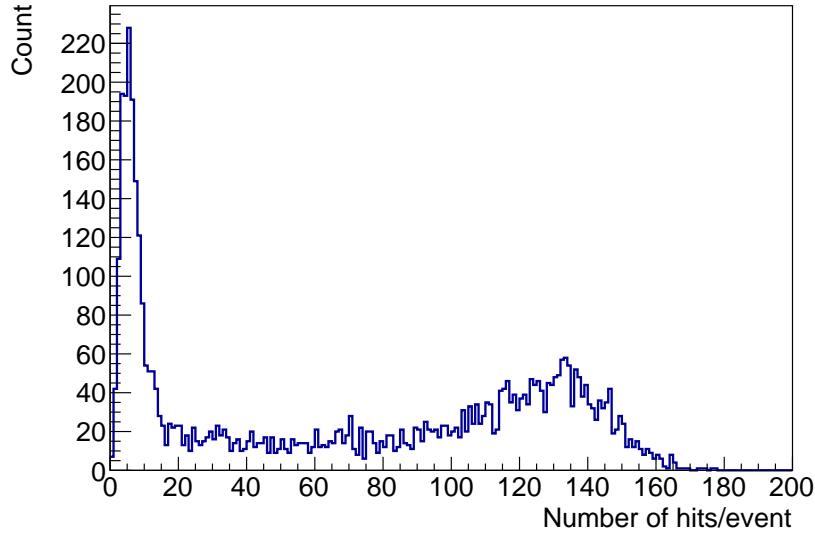


Fig. 3.6: Distribution of the number of PMT hits observed per event in the detector for laser PQ446 when the driving voltage value was 280. Data taken on February 22nd, 2021.

a PQ laser head it would then be passed through a manually-variable attenuator, fixed to some nominal attenuation setting for each laser. However, the emitted intensity from a given fibre emission point varies substantially. In theory, one could solve the intensity variation problem by deliberately setting the intensity well beyond the lasing threshold. This would then mean that one would have to change the attenuation of the attenuator in-person every time a different set of SMELLIE run conditions were used, in order to obtain the occupancies within the detector one is interested in.

Instead, Jeff Lidgard built a piece of hardware called the remotely-controllable Variable Fibre Attenuator (VFA), shown in Fig. 3.7. Contained within a metal housing were a ‘precision variable attenuator’ from DiCon Fibreoptics [143] for each PQ laser, along with an Arduino running firmware written by J. Lidgard to enable communication with each of the attenuators. Commands could be sent to a given attenuator asking for a specific attenuation expressed as a number between 0 and 3000, where the number divided by 100 gives the theoretical attenuation in

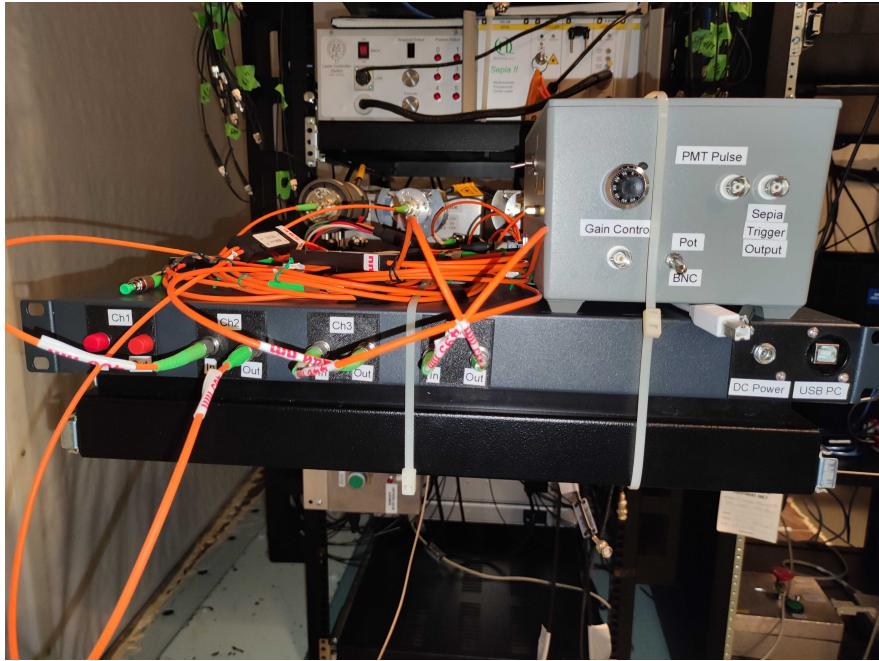


Fig. 3.7: Picture of the VFA during installation into the ELLIE hardware rack. The VFA box rests beneath the fibres connecting to the beamsplitters, as well as the newly-installed Monitoring PMT Unit (see Section 3.4 for details).

dB. Following ex-situ testing by J. Lidgard and Jasmine Simms, the VFA was installed underground by myself and Armin Reichold in July 2022, with some assistance from J. Lidgard in integration of the hardware and SMELLIE server software.

During testing of the VFA in-situ, it was discovered that the actual attenuation of the variable attenuator at the minimum setting of 0 dB for the PQ375 laser was so strong that negligible light was ever observed in the detector. This is likely because the precision variable attenuator components were primarily designed by the manufacturer with infrared fibre-optic transmission in mind: its nominal attenuation properties were calibrated by the manufacturer at 850 nm. Because of this, the PQ375 was not hooked up to the VFA, and kept its original attenuator setup. After fixing the driving voltage settings for PQ407, PQ446, and PQ495 to be 1000, 750, and 750 respectively, the observed maximum PMT occupancy in the detector was once again compared to the input laser intensity setting. The

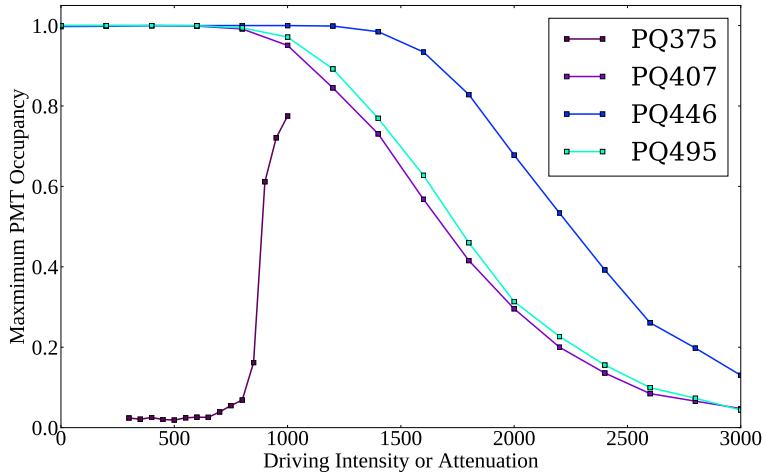


Fig. 3.8: Maximum occupancy in the detector as a function of the attenuation parameter value used, for each of PQ407, PQ446, and PQ495. Because PQ375 is not plugged into the VFA, its dependence on the driving intensity is shown instead. Data taken on the 12th of January, 2023.

PQ375 laser had its intensity controlled by the driving voltage still; the others were now controlled by only the attenuation setting. The results can be seen in Fig. 3.8. As hoped for, the three PQ lasers hooked up to the VFA can now have continuously-variable intensities of light observed in the detector over multiple orders of magnitude of observed intensity. Fig. 3.9 shows an example of the `nhit` distribution for the same laser and fibre as was used in Fig. 3.6; the event-by-event intensity stability can clearly be seen.

3.3 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 [144]. These fibres were chosen in part for their low cost and intrinsic radioactivity [145], 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

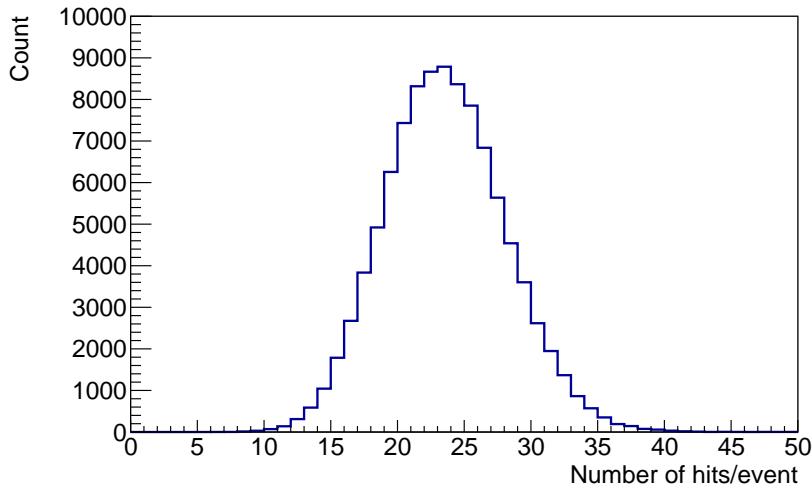


Fig. 3.9: Distribution of the number of PMT hits observed per event in the detector for laser PQ446, for a subrun of data taken in June 2023 after the VFA has been installed.

that the initial sharpness 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 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 3.4), 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 also contains collimators, designed to reduce the range of angles with which the light is emitted. This is particularly important for SMELLIE, because unlike the other

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 [115].

ELLIE systems, a thin beam of light across the detector is ideal for measuring scattering [139]. The mounting also points each of the three fibres in different directions: 0°, 10°, and 20° from the direction radially towards the centre of the detector.

Each fibre is given a name that nominally refers to both its mounting position and its pointing direction. For example, the label ‘FS107’ corresponds to the SMELLIE fibre mounted at node 07 with a pointing direction of 10°. Unfortunately, during installation some fibres were mislabelled, leading to the node mounting points and pointing directions of some fibres being inconsistent with the labelling convention. The actual mounting nodes and pointing directions for each fibre can be seen in Table 3.1. The 3D positions and pointing directions of all the fibres are determined in [141], and shown in Fig. 3.10.

When performing analysis with SMELLIE, a fibre-centric spherical-polar coordinate system is used. This coordinate system was first fully-developed by E. Turner; full details can be read in [115]. Fig. 3.11 shows a diagram of the

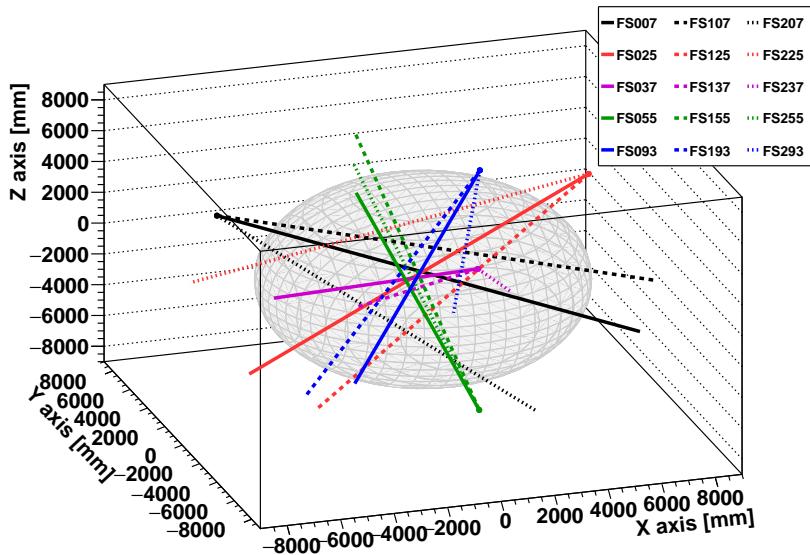


Fig. 3.10: Fibre emission points and nominal pointing directions within the detector.

coordinate system. 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 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: this is shown on the right-hand side of the diagram. 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:

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

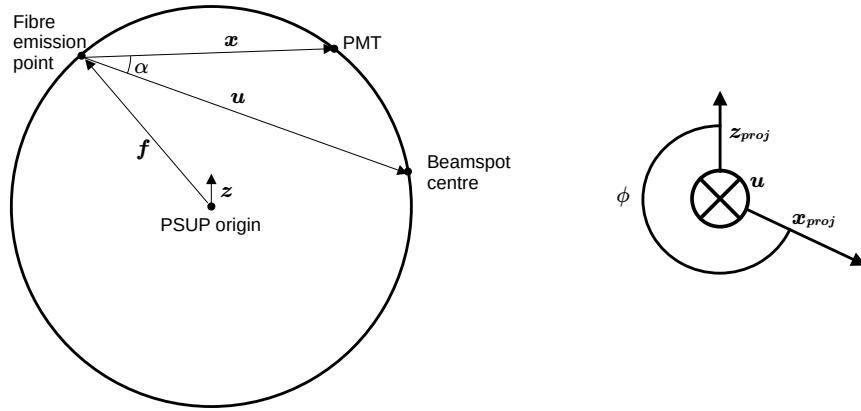


Fig. 3.11: Diagram of the SMELLIE coordinate system, adapted from [115]. The right-hand drawing is being shown in the plane orthogonal to $\hat{\mathbf{u}}$.

3.4 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 is used for monitoring purposes. This is achieved with a box known as the Monitoring PMT Unit (MPU). As the name suggests, the MPU contains a small PMT that generates an electronic signal pulse from the laser light. This signal is then shaped by electronics, and passed to the detector's central CAEN digitiser to have that pulse digitised.

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

Because of this, a new MPU was commissioned. Built by Adam Baird and Johan Fopma from the Oxford Physics Central Electronics Group, this MPU had

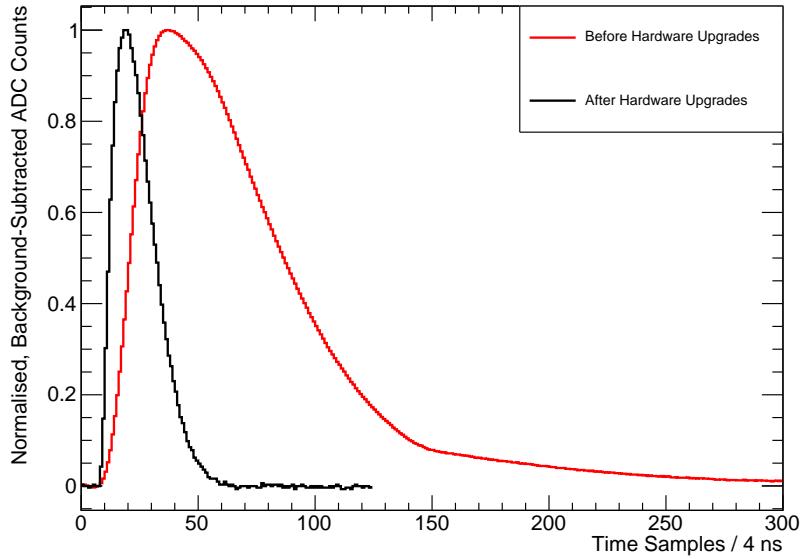


Fig. 3.12: Comparison between typical MPU CAEN traces generated by the PQ495 before and after the Summer 2022 hardware upgrades. The data before was taken on July 24th 2022, whilst the data after was taken on June 17th 2023. The traces have had their baselines subtracted off, and their peak values normalised to 1.

updated electronics such that the pulse was shaped shorter. In addition, the rise time of the pulse was made faster, in the hopes that the emission time of the light pulse for a given event could be captured more accurately. The new MPU was installed by myself and A. Reichold at the same time as the VFA, in Summer 2022.

Alongside the installation of the new hardware, the settings in ORCA for the CAEN digitisation of the MPU signal were updated. In particular, the number of samples made by the CAEN was shortened from 300 down to 124. The timing of the CAEN sampling and delay on the TUBii trigger (more on the trigger shortly) was also modified. As a result of these changes, a much shorter trace was now being generated, without missing any part of the MPU pulse. Fig. 3.12 shows a comparison between typical MPU pulses taken before and after the upgrades.

3.5 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, an EXTA signal can be sent from the firing laser to trigger the detector precisely when a light pulse is within the detector.

Trigger signal pulses are created by a National Instruments Bus-Powered M Series Multifunction DAQ Unit (the ‘NI Unit’), in place alongside the rest of the SMELLIE electronics. These trigger pulses are sent to either the SEPIA controller or the SuperK laser to induce the firing of the relevant laser. If a PQ laser is triggered in this way, a trigger pulse is simultaneously sent to TUBii, which after a tunable 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 pulse. Instead of sending the trigger signal to TUBii in parallel with the SuperK trigger, a photodiode contained within the SuperK laser system is able to detect the generated pulse, and from that detection a new trigger signal is derived and sent to TUBii.

A major problem with the handling of the trigger signal delay by TUBii was present throughout the collection of water phase SMELLIE data. This delay was ‘latched’ to the 100 MHz clock present within the electronics of TUBii. As a result, the observed hit times of PMTs in the detector (which are relative to the GT time) was a convolution of the true hit times and a top hat function of width 10 ns arising from the latching to the clock.

An example of this effect can be seen in Fig. 3.13. Light from the PQ495 laser was fired through fibre FS007 by E. Turner in June 2018, during the water phase. Looking only at PMTs that were hit in the beamspot, defined here as PMTs with

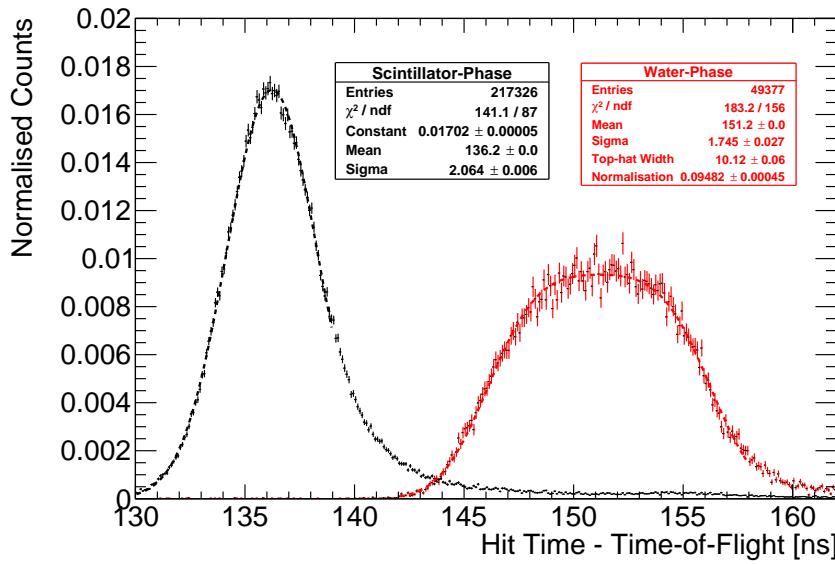


Fig. 3.13: Comparison of observed hit times in beamspot for laser PQ495 through fibre FS007, before and after T. Zummo's firmware update to TUBii. The initial data was taken by E. Turner June 20th 2018 during the water phase, whilst the data post-fix was taken by myself on June 24th 2023 during the scintillator phase. The peak of the former is fit to the convolution of a Gaussian with a top-hat function, whilst the latter is just fit with a Gaussian function.

$\alpha < 3^\circ$, the trigger times can be fit to the convolution of a Gaussian of width 1.7 ns with a 10 ns top hat.

During the partial fill phase, Tony Zummo updated the firmware of TUBii so that this latching would no longer occur, and the arrival times to the MTC/D were truly asynchronous to any clocks. The results are also shown in Fig. 3.13: notice how the width of the peak is now much sharper. This width is determined by the TTS of the hit PMTs, with the width of the PQ495 timing pulse being a subdominant effect.

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

In this thesis, two similar approaches are used to measure the event-by-event emission time t_{emm} , the time at which laser light first emanates from the fibre in 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 [115] 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. It will be seen that in Chapter 5, 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. 3.14 one can see the hit times of beamspot PMTs for SuperK light of wavelengths in the interval $[490, 500]$ nm from fibre FS125, taken on June 17th 2023. 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 it would be expected 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

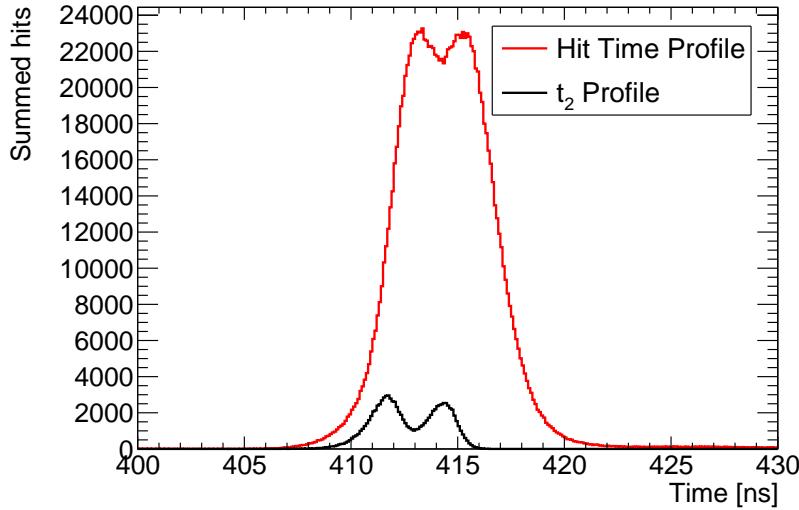


Fig. 3.14: Hit time and t_2 distributions for a subrun of SMELLIE data taken with the SuperK laser in the [490, 500] nm range through fibre FS125 on June 17th 2023. The double-peaked structure is clear in both distributions.

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

3.6 Software for SMELLIE Data-taking

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

To begin data taking, ORCA is used to put the DAQ into a ‘SMELLIE’ run type, which disables almost all the usual ‘Physics’ triggers such as N100 and N20. The only triggers still masked into the MTC/D are:

- EXTA: This is needed to capture the trigger signals sent from SMELLIE.
- ESUMHI: This residual physics trigger is kept in with a high threshold, to capture any `high-nhit` events not coming from SMELLIE, e.g. if a supernova were to go off whilst SMELLIE data taking was occurring.
- PULSEGT: This is a special trigger that fires at a rate of 50 Hz independent of all other systems. With this trigger, the noise rate can be calculated on a run-by-run basis for each channel.

The noise rates derived from the PULSEGT triggers are automatically calculated 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 PMT when generating simulations to match data. When performing analysis on SMELLIE data, only events from an EXTA trigger are considered.

Using ORCA, one can load any run plan already uploaded to the CouchDB server, and execute it. To execute the run, ORCA sends commands enacting the instructions laid out in the run plan to a SMELLIE-specific server-program running on a computer called SNODROP. This server converts these high-level commands into the low level instructions understood by the specific pieces of SMELLIE hardware needed to actually send the correct number of pulses of laser light of the right wavelength and intensity through the correct fibre. Details of the integration of SMELLIE with ORCA can be found in [146]

Whilst SMELLIE data is being taken, a “run description” `JSON` file is generated and then uploaded to the SMELLIE CouchDB server, describing the settings of the SMELLIE run actually executed. This differs from the run plan slightly in two

main ways: firstly, the same run plan can be run on multiple occasions. However, a run description file is associated to exactly one run of data that was actually taken. Secondly, if a SMELLIE run was terminated early either by the detector operator, or from a hardware problem, the run description file will correctly show only the subruns that were actually performed. The code repository used in all SMELLIE analysis has been updated by myself to use a given run description file in concert with the associated SMELLIE data file, so that the subrun-level metadata can be known during analysis.

Chapter 4

Simulating SMELLIE Events

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 MC simulation of the SMELLIE system. By modelling the laser light emission into, and propagation within the detector correctly, how SMELLIE light will be impacted by changing scattering lengths in the detector can be simulated. 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 requires a number of parameters to be correctly described. In particular:

- **Fibre emission positions** were recorded during the installation of the fibres.

- **Wavelength and emission timing distributions** of light pulses were taken from measurements of the laser heads by their manufacturers, or by colleague Jeff Lidgard in the case of the SuperK wavelength distribution [142].
- **The “emission intensity,”** defined as the mean number of photons simulated per event, is determined on a subrun-by-subrun basis. The number of photons for a given event is assumed to follow a Poisson distribution.
- **The beam profiles**, which describe the angular emission distributions of each fibre, is the focus of this chapter. These are necessary because unlike scintillation light, the light emitted from SMELLIE fibres is not isotropic.
- **Nominal fibre emission directions** attempt to define the centre of the beam for a given fibre.

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

4.1 Improving the SMELLIE Generator Algorithm

4.1.1 Previous Attempts at SMELLIE Event Simulation

Before the beam profiles can be determined, it must first be decided how to specify them. Previous observations show that different fibres can have notably different beam profiles [139], so each fibre’s beam profiles are allowed to be unique. It is assumed for now that a given fibre’s beam profile is stable over time, and independent of the wavelength of light fired. A straightforward, naïve approach

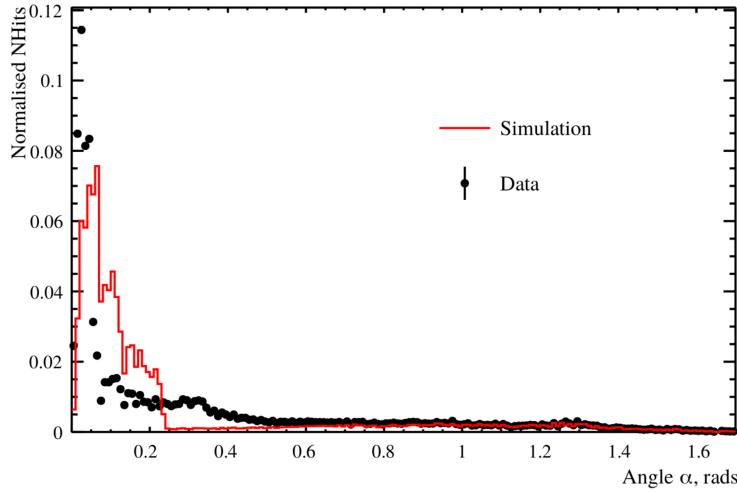


Fig. 4.1: Comparison between a simulation of one of the fibres, made from the 1D beam profile generator (red), with the associated data subrun that was used to create that beam profile (in black). For both MC and data, what is plotted is the PDF of observed PMT hits, as a function of the α angle. Poissonian 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 [115].

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 effect likely occurs because of the interference of different fibre propagation modes. This led to E. Turner building a SMELLIE generator that could handle 2D beam profiles:

dependent on both α and ϕ [115]. The distribution was stored as a map from each inward-pointing PMT in the detector to a relative intensity value. This was chosen because the beam profile shapes were calibrated from existing SMELLIE data — more on this in section 4.2.

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

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

This generator 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) in ~ 1 ms. However, the 2D generator specified here could take upwards of ~ 50 s *per event* to generate. Because SMELLIE analyses have historically required many millions of events to be simulated, 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. In addition, as many calculations were tried to be put before run-time as possible. Starting with the existing PMT relative intensity maps, these were plotted in the 2D direction-space $(1 - \cos \alpha, \phi)$: see Figure 4.3a. In a toy-MC simulation, 500,000 directions were 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

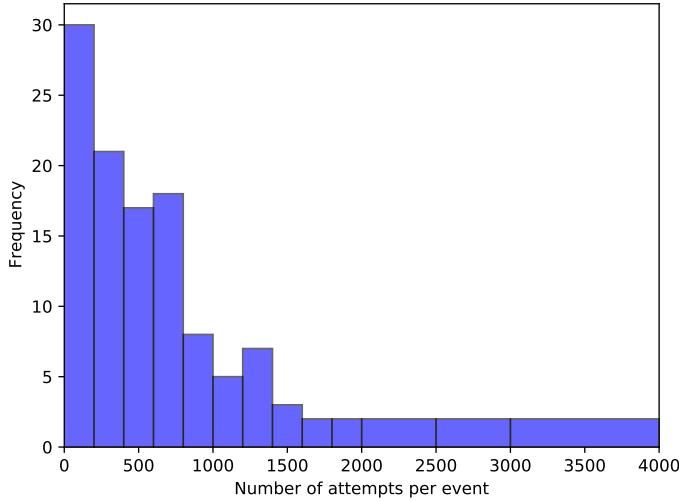


Fig. 4.2: Typical distribution of the number of attempts it takes for the existing 2D generator before the test direction gets accepted, per event.

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

Following this, the sampled intensities were then binned into a 2D histogram, where the bin value corresponds to the sum of all intensities for all directions found within this bin. Choosing a sensible binning procedure is important: too few bins, and necessary information about the shape of the beam is lost, whilst too many bins can oversample the data and capture statistical artefacts in the sampling process instead of just the beam profile. As a balance, 15 bins were chosen along the ϕ direction, and 60 in $r = 1 - \cos \alpha$. This was chosen to ensure that a reasonable number of PMTs were located within each bin, lessening the impact of any statistical fluctuations.

Although the bins in ϕ were chosen to have uniform width, this was decided to be not the case for the other axis, as there is far more important information near $r = 0$ (the beamspot). Instead, the width of the bins in r were calculated so that roughly the same total probability was contained in each r -strip. By consequence, bins near the beamspot typically are of significantly smaller size than ones much

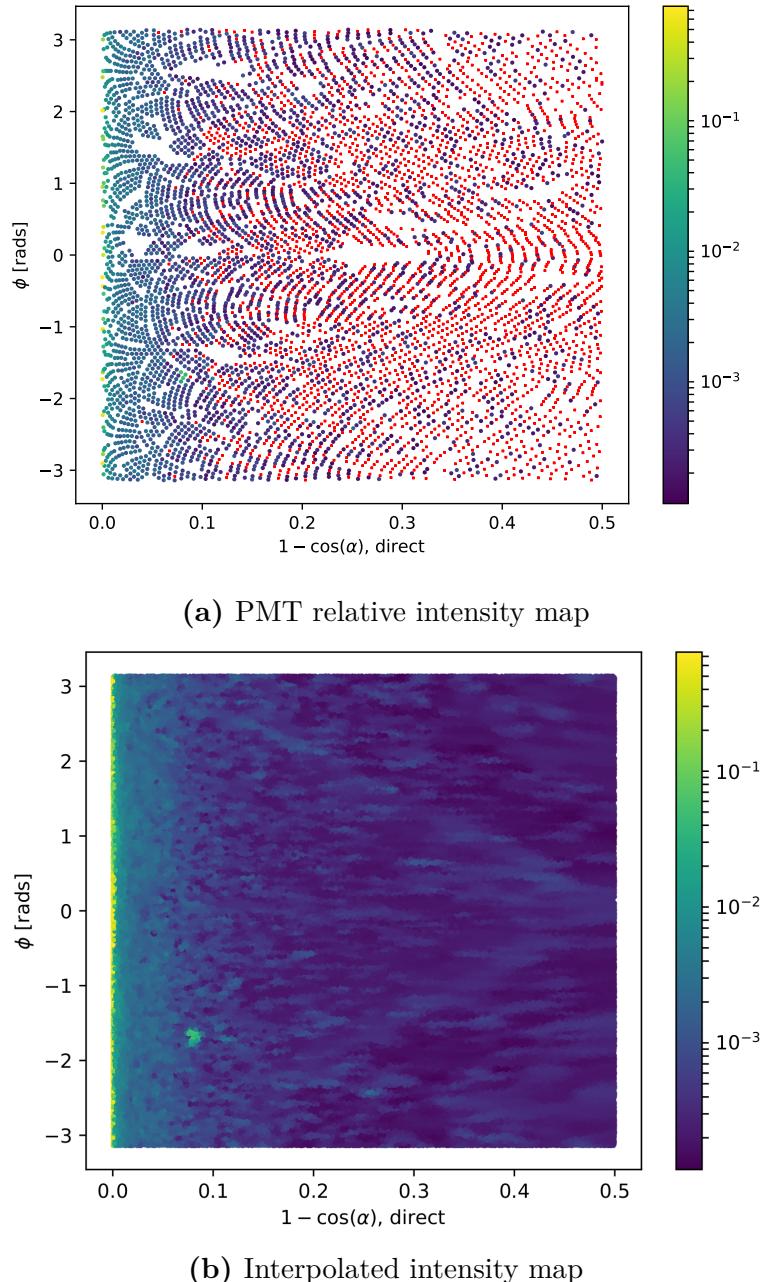


Fig. 4.3: The first step in the new method for preparing the new generator. In (a), the relative intensities used for the existing beam profile of fibre labelled FS055 are shown for each PMT, the position on the plot indicating the location of that PMT in the fibre coordinates. The colour indicates the 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.

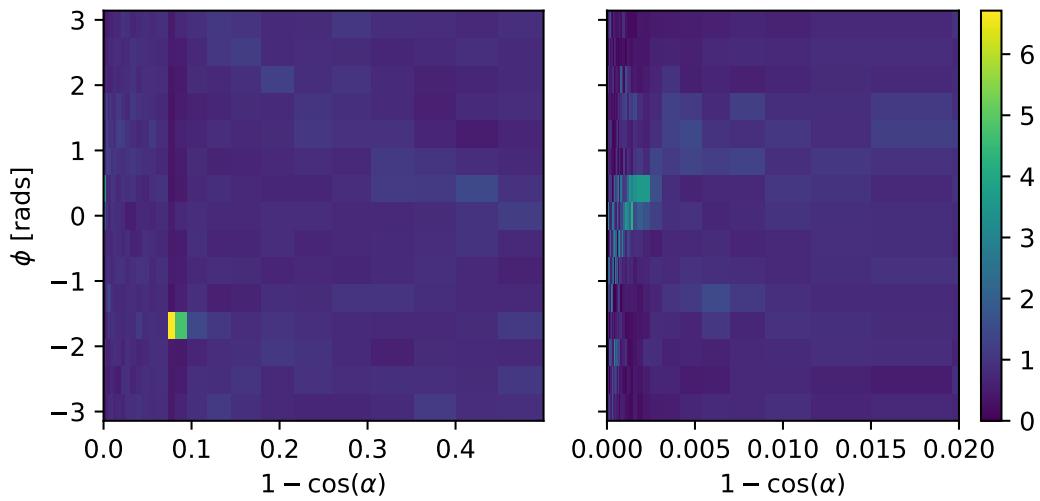


Fig. 4.4: Histogram of interpolated intensities within the 2D direction-space. The left view shows the full histogram; the right is a zoomed-in version near the beamspot. Unlike the binning in ϕ , the bin widths in r are not at all uniform. Instead, they have been determined such that the area summed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

further out. This captures 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 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:

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

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

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

4.2 Improving the beam profiles

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

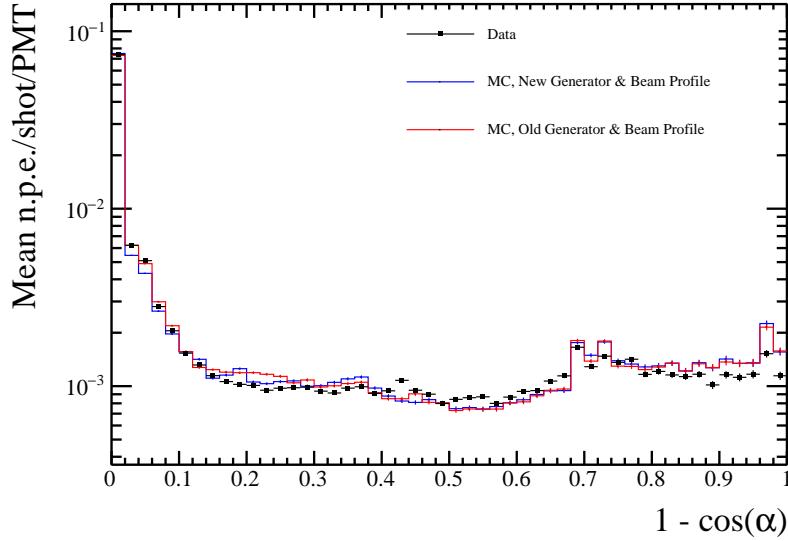


Fig. 4.5: Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches. The new generator also uses the updated beam profiles, as discussed in Section 4.2. The MC distributions have been scaled to match the total npe of the data.

of zero, providing little useful information about the beam shape. It was shown in [115] 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 try and 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 negligible in the data.

As mentioned in Section 2.3.3, 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 it is important to avoid PMTs with occupancies of 100%: they preclude one’s ability to convert into a value for μ by looking at occupancy alone. This conversion from occupancy into npe is known as the “multi-hit correction” [127]. 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, it is still expected for certain PMTs to be illuminated more than others because of e.g. reflections off the AV, ropes blocking the path, or the solid angle subtended by the PMT

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.

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 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 the only considered subruns were ones which used low or medium intensity set-points.

For each subrun j of data per fibre, only PMT hits for each PMT i that has been identified as “good” for that subrun are looked at¹, $i \in G_j$. G_j here

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.

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.

The hits arriving directly from the fibre without reflecting, scattering, or being noise were isolated through use of a time residual cut. 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, as estimated with the Collaboration’s “Light Path Calculator” utility. Then, the emission time was also subtracted, t_{emm} , as estimated by using the t_2 approach described in Section 3.5. 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’s emission 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 beamspot **nhits** below 9. 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, had no data subruns that satisfy this condition, and as such was not processed.

Extracting the underlying beam profiles from these datasets 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 emission intensity of the subrun. 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 of interest in this chapter.

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:

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

$$= \binom{N_j}{m_{ij}} \left(1 - e^{-\mu_{ij}}\right)^{m_{ij}} e^{-\mu_{ij}(N_j - m_{ij})}. \quad (4.6)$$

Here, equation 4.2 has been used. Note 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:

$$\langle \mu_{ij} \rangle = -\ln \left(1 - \frac{m_{ij}}{N_j}\right) = \ln (1 - o_{ij}) \quad (m_{ij} \neq N_j), \quad (4.7)$$

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, as expected.

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,

$$L(\{I_j\}, k_i | \{m_{ij}\}) = \prod_j L(I_j, k_i | m_{ij}) = \prod_j \binom{N_j}{m_{ij}} (1 - e^{-I_j k_i})^{m_{ij}} e^{-I_j k_i (N_j - m_{ij})}. \quad (4.8)$$

This leads to a log-likelihood distribution of

$$\mathcal{L}(\{I_j\}, k_i | \{m_{ij}\}) = \sum_j [\ln \left(\binom{N_j}{m_{ij}} \right) + m_{ij} \ln (1 - e^{-I_j k_i}) - I_j k_i (N_j - m_{ij})]. \quad (4.9)$$

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

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

$$i \in \mathcal{I} = \bigcap_j G_j. \quad (4.10)$$

The summed npe for a given subrun can then be defined as $S_j = \sum_{i \in \mathcal{I}} \text{npe}_{ij}$, with the assertion that $I_j = cS_j$ for some constant of proportionality c . By finding a value proportional to I_j , there is now enough information to maximise the log-likelihood $\mathcal{L}(k_i | \{m_{ij}\}, \{I_j\})$ with respect to k_i for each PMT independently, and hence obtain estimates for these k_i parameters.

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

For each isotropic MC run, both I_j^{MC} and k_i^{MC} were calculated via the method described above. Because the simulations were isotropic, the underlying value for b_i was constant across all the PMTs, and so $ak_i^{MC} = f_i$. By doing some rearranging of equation 4.4, one finds that:

$$\mu_{ij} = I_j b_i f_i = c S_j b_i a k_i^{MC} = (acb_i) S_j k_i^{MC}. \quad (4.11)$$

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.

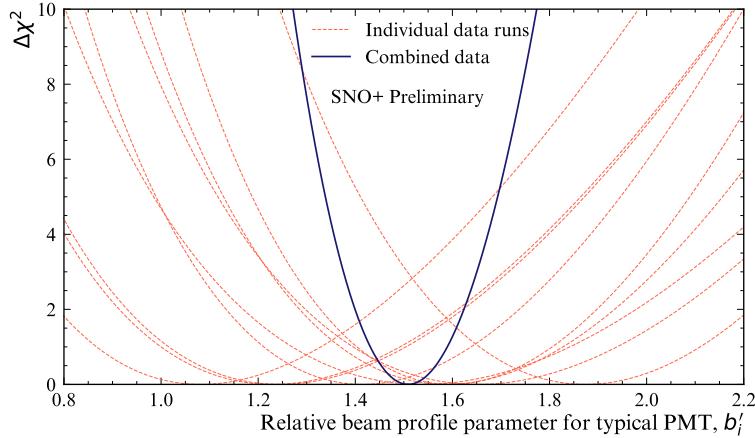


Fig. 4.6: Plot of $\Delta\chi^2 \simeq X_i$, twice the negative log-likelihood ratio, for both single subruns of a typical PMT, and when all relevant subruns are combined.

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 [147] allows one to produce 1σ confidence intervals about the maximum likelihood estimate for b'_i , $\langle b'_i \rangle$, because

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

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

Only a couple of exceptions to this approach of parameter estimation are possible. In the case where $m_{ij} = N_j$, i.e. a PMT has 100% occupancy, no maximum likelihood estimate exists: this only occurs in the high-intensity data not included in this analysis. On the other end of the intensity spectrum, there are some PMTs for certain fibres where even after all subruns of data have been included, there are no hits. In this scenario, one can show that the log-likelihood

becomes linear in the beam profile parameter:

$$\mathcal{L}(b'_i | \{m_{ij} = 0\}) = b'_i k_i^{MC} \cdot \sum_j [I_j N_j]. \quad (4.12)$$

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

$$b'_{i,ulim} = -\frac{k_i^{MC} \sum_j [I_j N_j]}{\ln \left[1 - \text{erf} \left(1/\sqrt{2} \right) \right]}, \quad (4.13)$$

where $\text{erf}(x)$ is the error function.

4.3 Comparisons between Data and Simulation

In this section, the major systematic differences between simulations of SMELLIE and data that remain after the above improvements have been made are discussed. These fall into three broad categories. The two of these are the differences seen in the forward and back hemispheres of the detector relative to a given fibre; the third are differences in the timing distributions of the light.

4.3.1 Forward Hemisphere Discrepancies

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.

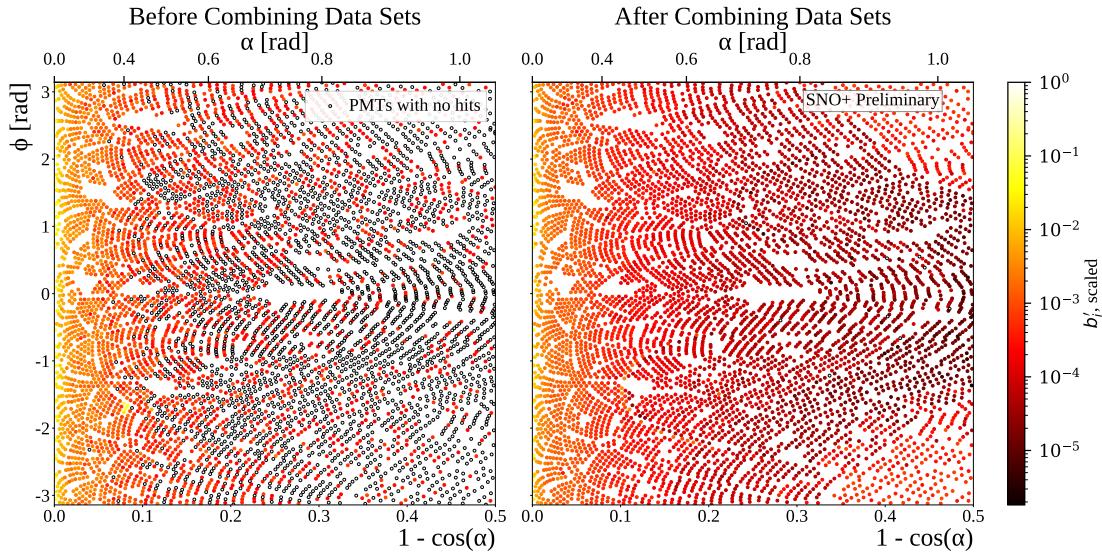


Fig. 4.7: Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets. Once again, the relative intensities (b'_i) for each PMT are given by the colour of each point, the position of each plotted in the 2D (r, ϕ) -space. The relative intensities have been both 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.

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 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 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

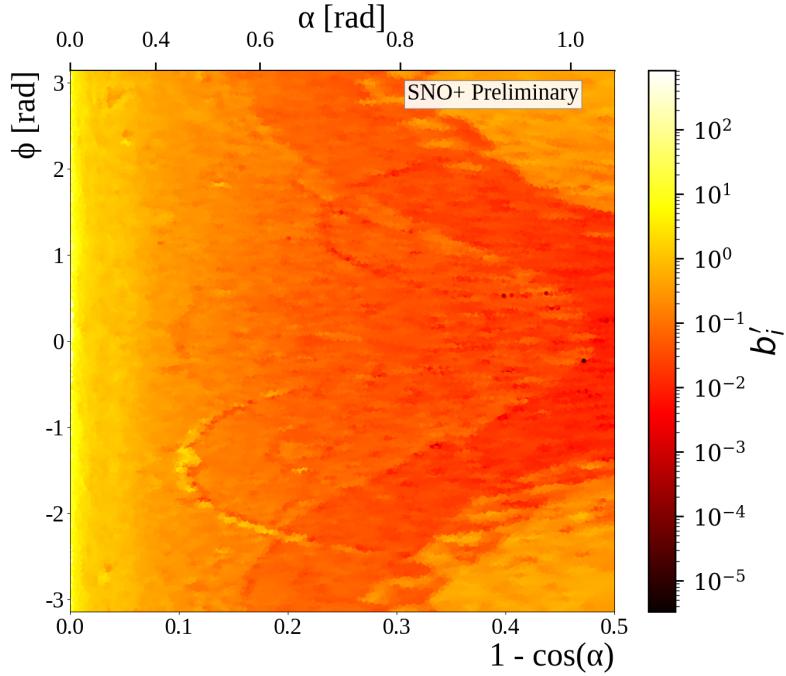


Fig. 4.8: Interpolated intensity map for the new updated beam profile of fibre FS055. The misalignment of rope shadows and AV effects, can both be seen.

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 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.

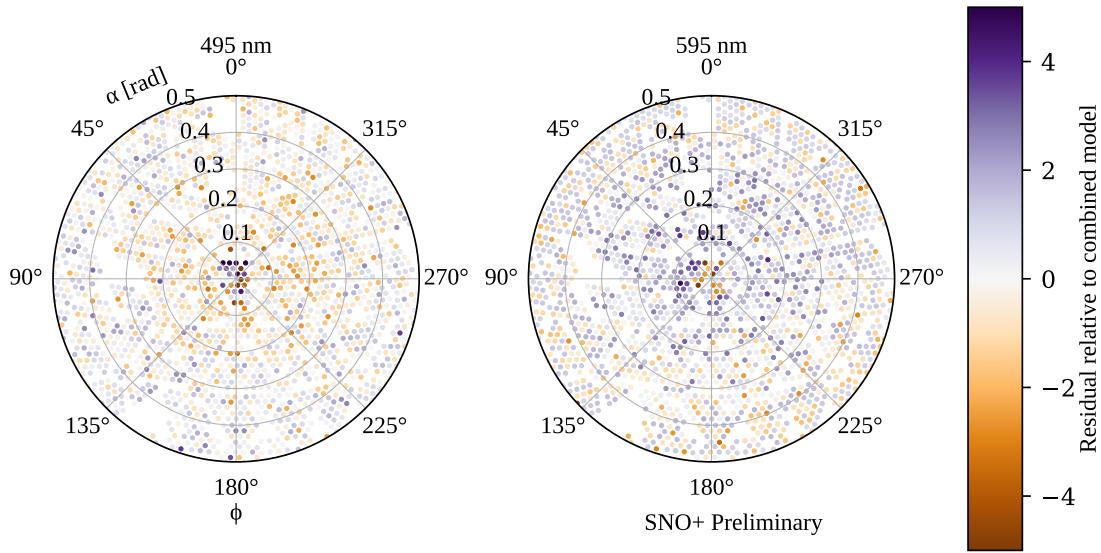


Fig. 4.9: Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055. A positive sign, and hence bluer colours, indicate that the combined model underestimates the observed intensity for that particular subrun. Values with a magnitude beyond 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).

4.3.2 Emission Time Discrepancies

In addition to systematics in the spatial distribution of light observed in the forward distribution, for two sets of lasers there are notable emission timing discrepancies. The first is the double-peaked timing structure of the SuperK laser, which was discussed in Section 3.5. The other main difference is associated with the PQ495 laser.

Fig. 4.10 shows the time residual distribution of hits in the 10° beamspot PMT region of fibre FS007, due to PQ495 laser light fired in July 2022. The “ t_{med} ” approach to determining the emission time was used, as described in Section 3.5. Also shown in this plot is the time distribution from a simulation of this same setup, and the PMT transit time distribution, as taken from [148]. The spike at $t_{\text{res}} = 0$ for data and MC distributions corresponds to events with only a single hit in the beamspot. Almost all hits shown here are known to come from direct light. The shape of the distribution in MC is largely determined by the transit

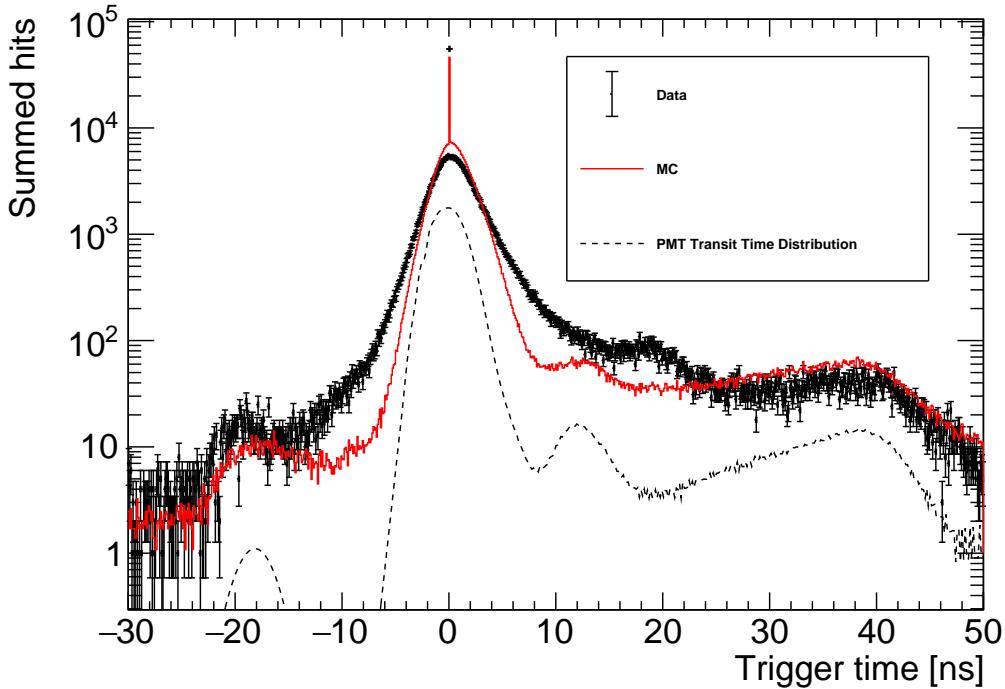


Fig. 4.10: Time residual distribution in the beamspot region for data taken in July 2022 with PQ495 through fibre FS007, as compared to a matching simulation. Also shown is the PMT transit time distribution, taken from [148]. The width of the data’s prompt peak is much broader than expected, and there is a notable bump at 19 ns not present in the MC or the PMT transit time distribution.

time distribution of the PMTs, which give rise to the four peaks seen in the MC. Aside from the prompt peak around $t_{res} = 0$, the earlier peak is associated with PMT pre-pulsing, and the latter two peaks with PMT late-pulsing [149].

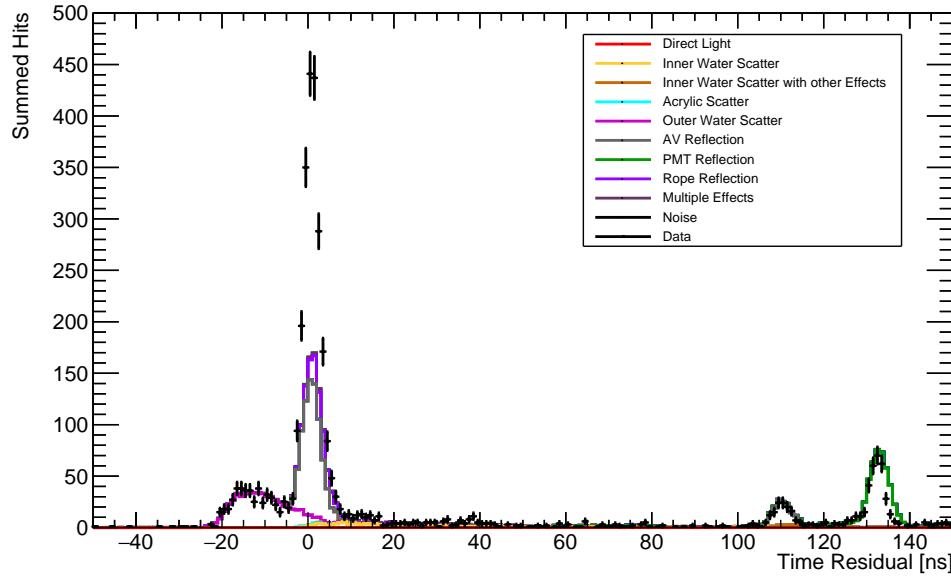
The two major discrepancies between data and MC seen for this laser is the much broader prompt peak seen in data than expected, and the bump at ~ 19 ns. Neither of these effects are seen in the other PQ lasers. Because these hits come almost exclusively from direct light, the most likely explanation for this difference is a change in the emission time distribution of the PQ495 laser, relative to the timing distributions used in RAT which were provided by the laser’s manufacturer. One method for resolving this problem could be to measure the laser’s emission timing distribution with an oscilloscope, in the same manner used by J. Lidgard to determine the timing distribution for the SuperK laser [142].

4.3.3 Backward Hemisphere Discrepancies

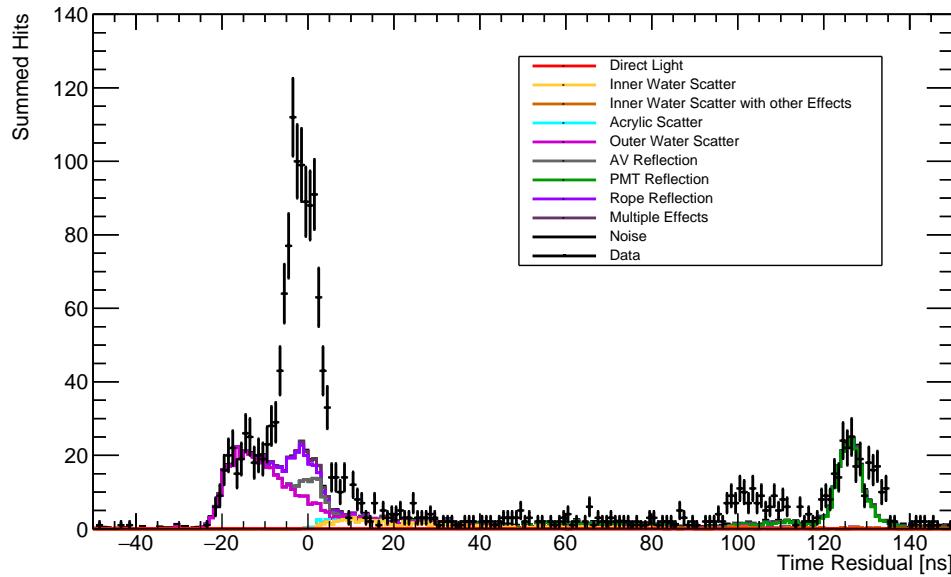
A third class of systematic differences between data and simulation of SMELLIE events are those discrepancies observed in the distributions of hits near a given fibre emission point. Consider the region of 50 PMTs closest to a given fibre. This defines what shall be called the “back-scatter” region of PMTs, because these will be the PMTs which will see the greatest intensity of light that has Rayleigh scattered in the UPW outside the AV. For a given subrun, only the subset of these PMTs which were ‘good’ in the sense described in Section 4.2.1 were used. This PMT region will become useful to the analyses described in Chapter 5.

Fig. 4.11 shows the time residual distributions for data taken in water phase run 114,018 in the back-scatter PMT region, using the SuperK laser in the wavelength range 490–500 nm. The method of emission time calculation used here is that of t_{med} . Because it is not possible for light to travel straight from the fibre emission point to a PMT in the back-scatter region, the light-paths used for calculating the time-of-flight in the time residuals here correspond to those which reflect off of the AV surface. Also shown in these plots are matching simulations of these subruns. The distributions from simulation are displayed as stacked histograms, where each individual histogram corresponds to the subset of hits associated with photons that underwent specific types of path. The overall intensity of the simulation has been scaled so that the number of hits in the $[-30, -10]$ ns time residual window is equal in data and MC, corresponding to a window in which only light that back-scattered in the outer water was observed.

Looking at Fig. 4.11a in particular, corresponding to the 0° fibre FS007, both data and simulation observe five distinct regions of interest. After the earliest peak from light which back-scattered in the outer water, there is a prominent peak near 0 ns arising from reflections off of the AV surface nearest to the fibre, as well as reflections from ropes. After this, in the 10–50 ns range, there is a



(a) Fibre FS007



(b) Fibre FS155

Fig. 4.11: Plots showing comparisons between the time residual distributions in the ‘back-scatter’ PMT region for water phase SMELLIE data taken in run 114,018, with the SuperK laser in the wavelength range 490–500 nm for two different fibres, and corresponding MC.

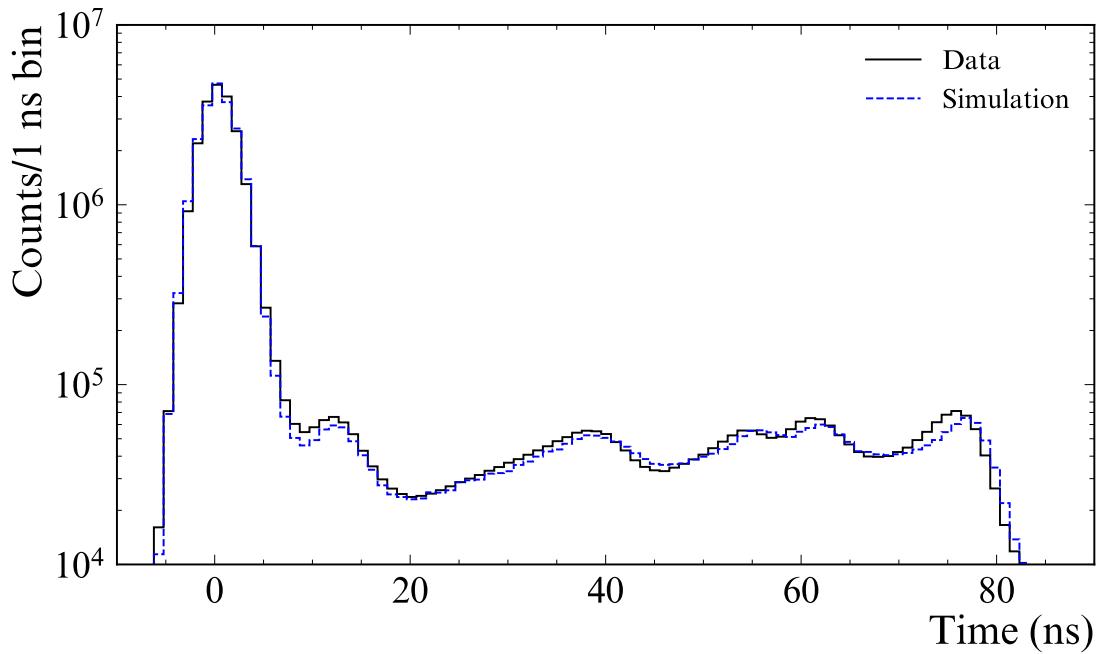


Fig. 4.12: Comparison of Laserball data in the water phase to MC, with a wavelength of 420 nm. The latest peak comes from PMT reflections. From [85].

long tail of hits due to a number of effects, including back-scattering in the inner water. Much later, there are two further peaks: one at 110 ns due to reflections off of the far AV surface, and one at 132 ns due to reflections off of the PMTs and concentrators on the far side of the detector.

Although all the features described above are seen qualitatively in both data and MC, there are a couple of notable discrepancies. The first to note are the differences in shape seen in the PMT reflection peaks of both fibres shown. The FS007 data has a tighter peak than expected from simulation, whilst the FS155 data sees the opposite. These differences are likely due to the current limitations of the PMT reflection model used in RAT. Fig. 4.12 shows the best-fit of Laserball data to simulations with varying PMT reflection models, taken from [85]. The last peak in this plot comes from PMT reflections, which shows a small disagreement between data and MC. It is presumably this residual systematic in the modelling of the PMT reflections that leads to the shape differences in the peaks seen in the SMELLIE data.

The other prominent difference is the magnitude of the peak at 0 ns, with data appearing to have a peak over twice the size of simulation. This can clearly be seen for the results of both fibres in Fig. 4.11. What could cause this discrepancy? One simple explanation is that the intensity calibration of the MC is wrong. Because the intensity of the simulation shown here has been forced to match that of the outer water back-scattered light, if the scattering length of this water is systematically off in simulation then so too would the intensity calibration. If this were the case, then the true Rayleigh scattering length of the outer water would have to be substantially longer than expected in order to explain the discrepancy in the peak at 0 ns. However, this would cause a substantial change in the magnitudes of the two later peaks. Given that these two late peaks have magnitudes which appear to agree with data under the current intensity calibration, a large systematic in the UPW scattering length does not seem like the primary cause of this systematic.

A second hypothesis for explaining the difference in peak heights at 0 ns is that there is far more light being reflected off of the AV surface than expected. As discussed in Section 2.3.2, standard specular reflection is entirely determined by the refractive indices of the two media, and the angle of incidence of the light. The refractive indices of the UPW and acrylic are well-known, and because FS007 is a 0° fibre, the angles of incidence of light rays will be close to 0° also. The fact that there appears to be good agreement between data and MC in the peak from reflections off of the far AV surface casts strong doubt on a substantial miscalculation of the specular reflection. An alternative theory is the existence of a diffusive reflection component on the AV surface, due to a surface roughness. This could possibly explain the existence of a peak seen in data in Fig. 4.11b, but not in the MC.

There are two further hypotheses which could possibly help to explain the 0 ns peak discrepancy. One is that the extra light in this peak comes from scattering in the acrylic, with the scattering length being much shorter than currently

modelled. The current **RAT** model indicates that only a negligible fraction of light gets scattered in the acrylic, so there would need to be a dramatic change in the scattering length. The final hypothesis is that the rope reflections are currently being mismodelled. Currently, the model for rope reflections in **RAT** assumes 60% reflectivity with the reflections being purely diffusive. However, this is based on preliminary estimates only, with no formal calibration to back this reflection model up. It could be the case that a different reflection model could help to alleviate the tension between data and model in the peak, although it is challenging to see how the all the difference could be attributed to purely this.

4.4 Summary and Suggestions for Further Work

This chapter described two updates to the simulation of SMELLIE events. Firstly, a new algorithm for the generation of SMELLIE events was built, leading to a dramatic speed-up in simulation times of multiple orders of magnitude. This was achieved by converting a rejection sampling approach to sampling a fibre's beam profile into an ‘inverse-CDF’-style one, as well as pre-computing many calculations that were historically performed at run-time. The beam profiles were then updated by combining multiple datasets together, using a carefully-designed statistical model. By using substantially more data to build the beam profiles, a much greater dynamic range in the profiles was possible.

The updating of the beam profiles also uncovered a number of systematics still present in the simulations. By using interpolated forms of the new beam profiles, there is clear mismodelling in acrylic attenuation and the position of rope shadows. There is also a notable wavelength-dependence to the shape of the beam profiles. In addition, other data-MC discrepancies exist: the PQ495 laser has an observed emission time spectrum that does not match expectation, and there appears to be far more light coming from AV and/or rope reflections than expected, if one

assumes that the modelling of the outer UPW scattering is correct. There are also differences in the shape of the time distributions of light reflected off of PMTs and their concentrators.

There are a number of things that can be done to further the work performed in this chapter. These come in two main classes: improvements to the SMELLIE generator, and improvements to the optical model of the detector in **RAT** more generally. For the former, the emission time distribution for the PQ495 laser should be measured external to the detector's DAQ, possibly using an oscilloscope. Alternatively, the PQ495 emission time distribution could be recovered by deconvolving the observed time distribution with the known PMT transit time distribution.

For the beam profiles, even more water phase data could be used if one were to include the ‘high’ intensity data that was taken back in 2018, but never used. This would require some thought about how to appropriately handle PMTs which were fully-saturated: the npe would then have to be estimated by looking at the charge information. If one can work that out, then the benefit would be far greater statistics. In doing so, this would alleviate the reliance on combining beam profiles at multiple different wavelengths. Instead, separate beam profiles could be developed for different wavelengths, with the beam profile for a given simulated wavelength being the interpolation/extrapolation of these different beam profiles. It would also be beneficial to have a physical understanding of why this wavelength-dependence occurs.

However, the quality of the beam profile is fundamentally limited by the accuracy of the associated isotropic MC. This is linked to the second class of possible improvements: those of the optical model of the detector itself. In multiple beam profiles there are clear indications of rope shadows being incorrectly modelled. This could either come from the ropes being simulated in the wrong place, or the fibre emission points being incorrect, or a combination of both. By

using the information from these beam profiles, it might be possible to determine where the ropes and fibres are. Also, the water phase SMELLIE data could be used to try and calibrate the optical models of the rope and PMT reflections.

There are indications of problems with the modelling of the acrylic optics from both the interpolated beam profile plots, and the time residual distributions of the back-scattered light region of PMTs. These could be related to the long-standing problem of attempting to model physics events near the AV — for example, Iwan Morton-Blake saw in the partial fill phase that the `nhits` distribution of $^{214}\text{Bi}-\text{Po}$ radioactive background events had up to a 20% difference in median value between data and MC near the AV [90]. More information about this background can be found in Section 6.2. A similar effect was seen in the main scintillator phase by T. Zummo [150]. Investigating different models of the acrylic optics to fix these systematic problems is worthwhile not just for SMELLIE, but for the experiment more broadly. It would be important to make sure that any optical model changes would maintain consistency with the data taken by the Laserball.

Chapter 5

Analysis of SMELLIE Data in the Scintillator Phase

*Due to Local Education Authority cuts,
The light at the end of the tunnel has been
switched off.*

DR. FRANK CHEW

In this chapter, two innovative analyses using data from SMELLIE are built and implemented. The first is a method for measuring the extinction lengths of the SNO+ scintillator as a function of both wavelength and time; the second is a way to monitor changes in Rayleigh scattering. A key theme of both is the careful design of their methodologies to avoid the known systematics discussed in Section 4.3.

5.1 Extinction Length Analysis

The first analysis discussed in this chapter is the measurement of the extinction lengths of the liquid scintillator deployed in the SNO+ detector, as a function of both wavelength and time. As seen in Fig. 3.2, SMELLIE data was taken during

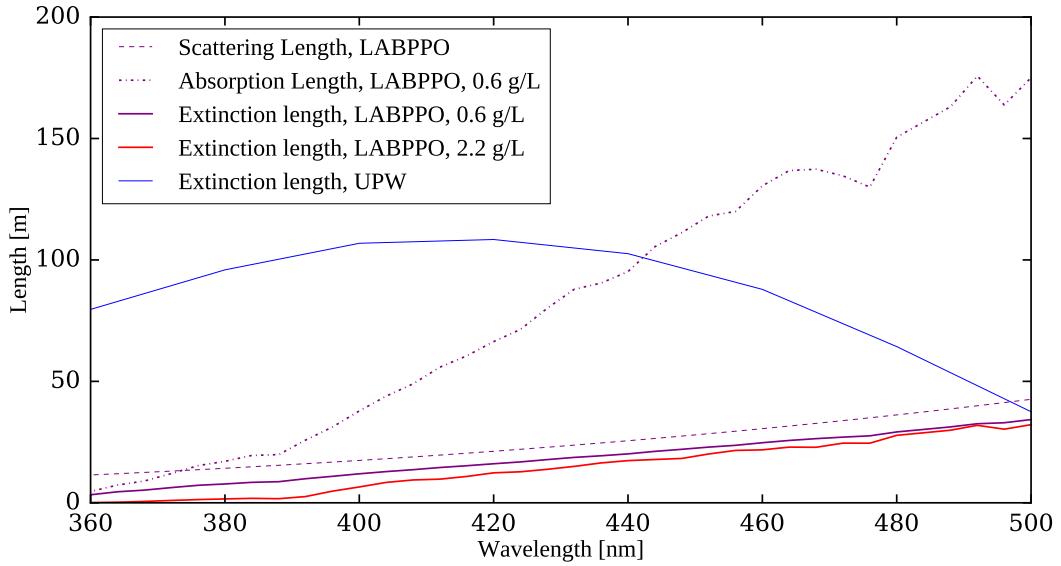


Fig. 5.1: Current RAT models for the extinction lengths of both UPW and scintillator, at different PPO concentrations as a function of wavelength. For the scintillator, the absorption and scattering length distributions are also shown as a dotted line, showing that it is expected for scattering to dominate over absorption in the 400–500 nm range.

the water phase, and in the scintillator phase when the PPO concentration of the LAB-PPO was at a variety of levels. Fig. 5.1 shows the changes in extinction length with detector phase, as a function of wavelength, according to the optical model described in Section 2.3.2, and implemented in RAT. This will be discussed further in Sections 5.1.3 & 5.2.4.

5.1.1 Mathematical Model

To begin, consider light emission from a particular SMELLIE fibre of fixed wavelength λ_j , with a mean emission intensity of I_j photons emitted per event in subrun j during detector phase p . Modifying Eq. 4.4, a general PMT i will have a mean npe per event of $\mu_{ij,p}(\lambda_j) = I_{j,p} b_{i,p}(\lambda_j) f_{i,p}(\lambda_j)$. A dependence on the detector target medium has been added to all terms, because:

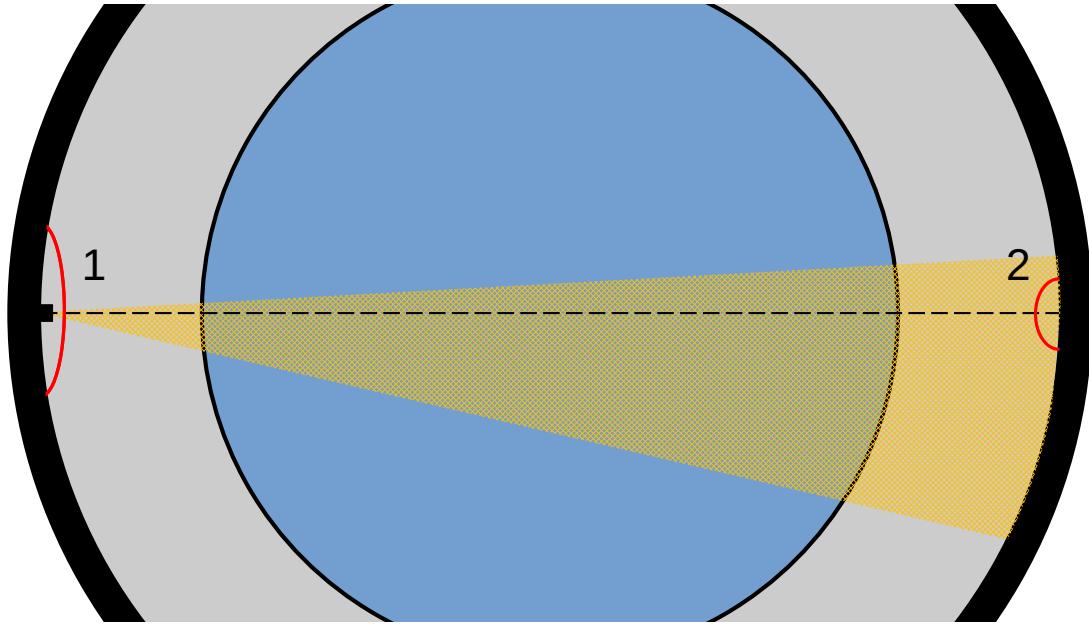


Fig. 5.2: Diagram showing the two PMT regions used in this analysis. Region 1 is the set of back-scatter PMTs, whilst region 2 is the set of far PMTs.

- The emission intensities chosen, $I_{j,p}$, can vary between different data-taking campaigns. These changes are due to changes in hardware (as detailed in Chapter 3), as well as changes in the settings used to run SMELLIE.
- The beam profile of a given fibre at a given wavelength is not expected to change with detector phase. However, if the refractive index of the inner detector medium changes (e.g. from the water phase to scintillator phase), then the fraction of light emitted that is pointed in the correct direction to be detected by PMT i , $b_{i,p}(\lambda_j)$, can change. Because the light-path needed to get from the fibre to the PMT changes, so too will the probability that a photon travelling such a path will actually make it and generate a photoelectron, $f_{i,p}(\lambda_j)$.

In this analysis, two regions of PMTs will be used for a given fibre; Fig. 5.2 shows a schematic of the PMT selections. The first region corresponds to PMTs diametrically-opposite to the fibre emission point in the detector, i.e. the light-paths travel orthogonally through all surface boundaries. Considering the various

contributions to the value of $f_{i,p}(\lambda_j)$, a PMT in this ‘far’ selection will have a mean npe per event, $\mu_{ij,p}^{\text{far}}(\lambda_j)$, of:

$$\begin{aligned}\mu_{ij,p}^{\text{beam}}(\lambda_j) = I_{j,p} b_{i,p}(\lambda_j) \exp\left(-\frac{L_{ij,p}^{\text{extern}}(\lambda_j)}{l_{\text{extern}}(\lambda_j)}\right) \exp\left(-\frac{L_{ij,p}^{\text{acr}}(\lambda_j)}{l_{\text{acr}}(\lambda_j)}\right) \exp\left(-\frac{L_{ij,p}^{\text{inner}}(\lambda_j)}{l_p^{\text{inner}}(\lambda_j)}\right) \\ \cdot T_{ij,p}(\lambda_j) \epsilon_{ij,p}(\lambda_j).\end{aligned}\quad (5.1)$$

Here, $L_{ij,p}^{\text{extern,acr,inner}}(\lambda_j)$ is the path length traversed in a given detector medium through which light travels, for the external water, acrylic, and inner detector medium, respectively. $l_p^{\text{extern,acr,inner}}(\lambda_j)$ is the extinction length of each detector medium for a given wavelength — it is assumed that only the inner detector medium has changing optics in different phases. Some fraction of light is lost when passing between two media with different refractive indices; this is captured by $T_{ij,p}(\lambda_j)$, the product of the Fresnel transmission components for all optical boundaries along a photon’s path. Finally, $\epsilon_{ij,p}(\lambda_j)$ is the probability that a photon along a given path, incident on a given PMT, will generate a photoelectron that is detected.

The second selection of PMTs are those near the fibre emission point. The first light observed by these PMTs will be from photons which have ‘back-scattered’ in the UPW outside the AV. This is followed by light reflected off of the AV surface. The details of isolating this backscattered light is discussed in Section 5.1.2. Assuming that the Rayleigh scattering properties of the UPW outside the AV have been unchanged throughout the lifetime of the detector, then the expected number of photoelectrons observed in a selection of these PMTs during a time period in which only back-scattering can occur will be simply:

$$\mu_{j,p}^{\text{back}} = k I_{j,p}. \quad (5.2)$$

k here is just some general constant of proportionality. Therefore, observing this back-scattered light can be used as a measure of the relative intensity of the subrun.

Because $\mu_{j,p}^{\text{back}}$ is proportional to the intensity, the ratio $R_{ij,p} = \mu_{ij,p}^{\text{beam}}(\lambda_j)/\mu_{j,p}^{\text{back}}$ will be independent of $I_{j,p}$. A similar method can be used to remove the dependence of $R_{ij,p}$ on k , by taking the ratio of $R_{ij,p}$ with $R_{ij,\text{H}_2\text{O}}$, where $R_{ij,\text{H}_2\text{O}}$ is the measured values of $R_{ij,p}$ in the water phase. This ratio becomes:

$$\begin{aligned} \frac{R_{ij,p}}{R_{ij,\text{H}_2\text{O}}}(\lambda_j) &= \frac{k}{k} \frac{b_{i,p}(\lambda_j)}{b_{i,\text{H}_2\text{O}}(\lambda_j)} \frac{T_{ij,p}(\lambda_j)}{T_{ij,\text{H}_2\text{O}}(\lambda_j)} \frac{\epsilon_{ij,p}(\lambda_j)}{\epsilon_{ij,\text{H}_2\text{O}}(\lambda_j)} \exp\left(-\frac{L_{ij,p}^{\text{extern}}(\lambda_j) - L_{ij,\text{H}_2\text{O}}^{\text{extern}}(\lambda_j)}{l^{\text{extern}}(\lambda_j)}\right) \\ &\quad \cdot \exp\left(-\frac{L_{ij,p}^{\text{acr}}(\lambda_j) - L_{ij,\text{H}_2\text{O}}^{\text{acr}}(\lambda_j)}{l^{\text{acr}}(\lambda_j)}\right) \exp\left(-\frac{L_{ij,p}^{\text{inner}}(\lambda_j)}{l_p^{\text{inner}}(\lambda_j)} + \frac{L_{ij,\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)}{l_{\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)}\right) \\ &= \frac{b_{i,p}(\lambda_j)\epsilon_{ij,p}(\lambda_j)}{b_{i,\text{H}_2\text{O}}(\lambda_j)\epsilon_{ij,\text{H}_2\text{O}}(\lambda_j)} \frac{T_{ij,p}(\lambda_j)}{T_{ij,\text{H}_2\text{O}}(\lambda_j)} \exp\left(-\frac{L_{ij,p}^{\text{inner}}(\lambda_j)}{l_p^{\text{inner}}(\lambda_j)} + \frac{L_{ij,\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)}{l_{\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)}\right), \end{aligned} \quad (5.3)$$

where it has been assumed that any change in path length through the external UPW or acrylic relative to their extinction lengths is negligible.

Importantly, when considering the first PMT selection, a further simplification can be made to the above formula. Because the light travels orthogonally through the AV boundaries, its path is unaffected by changes in the refractive index of the inner detector medium. Therefore, the impact of the beam profile and PMT efficiency will be unchanged, and so the formula simplifies to:

$$R_{ij,\text{H}_2\text{O}}(\lambda_j) = \frac{T_{ij,\text{H}_2\text{O}}(\lambda_j)}{T_{ij,p}(\lambda_j)} \exp\left(\frac{L_{ij,p}^{\text{inner}}(\lambda_j)}{l_p^{\text{inner}}(\lambda_j)} - \frac{L_{ij,\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)}{l_{\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)}\right) \cdot R_{ij,p}(\lambda_j). \quad (5.4)$$

The measurable quantities $R_{ij,\text{H}_2\text{O}}(\lambda_j)$ and $R_{ij,p}(\lambda_j)$ are then proportional, with the constant of proportionality being a function of the variable of interest $l_p^{\text{inner}}(\lambda_j)$.

This methodology for measuring $l_p^{\text{inner}}(\lambda_j)$ deliberately avoids the systematic effects described in Section 4.3. By taking the ratio of two different datasets

instead of using MC, any systematics of the simulation are removed. The selection of the ‘far’ PMT region ensures that the npe observed in this region is independent of the refractive index of the inner detector medium.

5.1.2 Parameter Measurements and Uncertainties

As a result of Eq. 5.4, measuring the extinction length of the scintillator first requires measuring a number of other quantities with knowledge of their uncertainties.

UPW Extinction Lengths

As discussed in Section 2.3.2, the attenuation lengths of the UPW were measured as a function of wavelength in the water phase with the Laserball. It was assumed that the optics of the UPW inside and outside the AV were the same, and outer UPW has not changed since.

The measured values of the attenuation coefficients $\alpha_w(\lambda) = 1/l_{\text{H}_2\text{O}}^{\text{inner}}(\lambda)$ and their associated errors were taken from [85]. The wavelength range this Laserball data was taken over was 337–500 nm; only wavelengths in this range were considered in this analysis. For a given SMELLIE subrun with wavelength λ_j , $l_{\text{H}_2\text{O}}^{\text{inner}}(\lambda_j)$ was estimated by linearly interpolating between Laserball α_w data points, and then taking a reciprocal. The uncertainties of the Laserball measurements were dominated by the systematic uncertainty in the Laserball position, which are highly-correlated between data points. Therefore, the uncertainty in $\alpha_w(\lambda_j)$ was estimated by linearly interpolating the quadrature sum of the statistical and systematic uncertainties at each Laserball data point. The Fig. 5.3 shows this process in action from the wavelength 375 nm. At its largest, the uncertainty in $l_{\text{H}_2\text{O}}^{\text{inner}}$ is $\sim 50\%$. Fortunately, the impact of this large error is mitigated in Eq. 5.4 because $L_{ij,\text{H}_2\text{O}}^{\text{inner}}(\lambda_j) \ll l_{\text{H}_2\text{O}}^{\text{inner}}$.

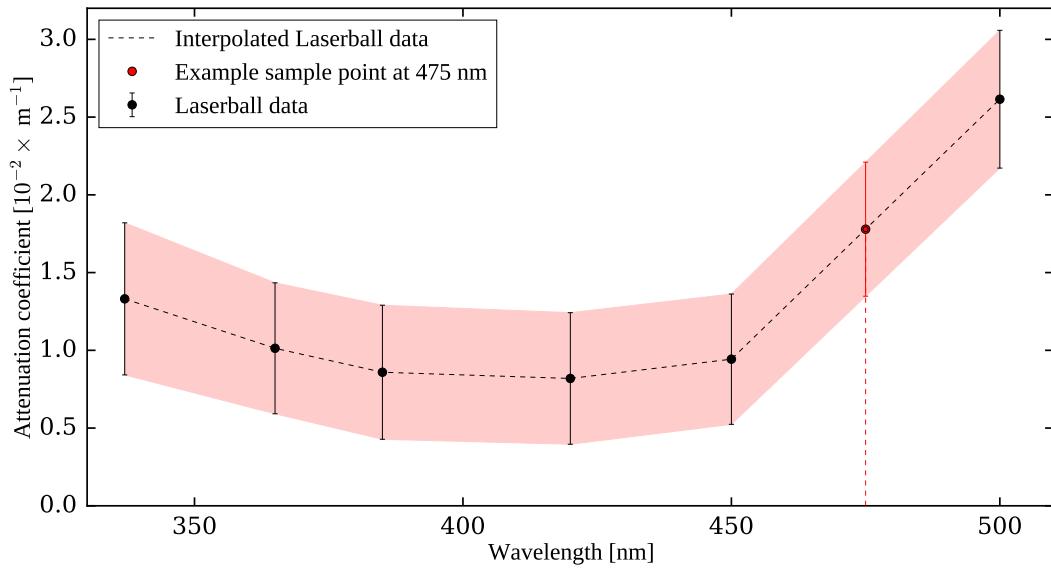


Fig. 5.3: Plot showing how linear interpolation has been used on the Laserball water phase UPW attenuation coefficient data. An example point at 475 nm is also shown.

Path Lengths and Transmission Coefficients

Section 2.3.2 also discusses the measured refractive indices as a function of wavelength for the UPW, acrylic, and LAB-PPO. For a given detector phase, subrun, and PMT, the Collaboration’s **Light Path Calculator** is able to determine the values of $L_{ij,p}^{\text{inner}}(\lambda_j)$ as well as the combined Fresnel transmission coefficient $T_{ij,p}(\lambda_j)$. It is assumed that there is negligible uncertainty in these values.

Measuring the Number of Photoelectrons

Critical to this analysis is the determination of the mean npe per event in both ‘far’ and ‘backscatter’ PMTs. These two PMT selections have to be approached slightly differently.

The ‘far’ PMTs were selected by first finding the intersection point on the PSUP with a line that passes through both the fibre emission point and the centre of the AV. The 20 PMTs closest to the intersection point were chosen. For a given analysis between a scintillator phase subrun j and a water phase subrun

with matching laser, wavelength, and fibre, only PMTs inside this selection which were identified as “good” (as defined in Section 4.2.1) in both subruns were used. For a given far PMT being used, direct light was isolated by calculating the time residuals of all hits on the PMT, using the “ $t_{\text{emm}} = t_{\text{med}}$ ” approach mentioned in Section 3.5. Then, the number of hits observed in a “tight” time residual window of $[-5, +5]$ ns was measured for the PMT of interest. By converting to occupancy and then using a multi-hit correction as described in Eq. 4.3, the total npe per event for that PMT was estimated. The uncertainty in this value was given by the Poisson error of the calculated npe. A final correction was made to account for noise hits: the measured noise rate of a given PMT was calculated using the PULSEGT triggers, as described in Section 3.6. After accounting for the width of the time window, the corresponding expected number of noise hits was subtracted off of the total measured npe to give the npe per event from direct light only. In order to minimise the statistical uncertainty, all individual far npe measurements for a given subrun were combined into one value of the ‘far’ light npe per event, $\mu_{j,p}^{\text{far}}$.

Backscattered PMTs were selected for each fibre by finding the 50 PMTs closest to the fibre emission point. A t_{res} window of $[-30, -10]$ ns was used for the isolation of backscattered light in each subrun. Like above, the total hits for each PMT was converted into a total npe with associated Poisson error, and had their estimated number of noise hits subtracted off. The npe from these PMTs were combined into one value of the backscattered light npe per event, $\mu_{j,p}^{\text{back}}$. The variables R_s and R_w are the ratios of the far light npe to the backscattered light npe for a given scintillator or water phase subrun, respectively. The subrun j subscript has been dropped for simplicity.

Fig. 5.4 shows the time residual distributions for both PMT selections of a simulation of the PQ407 laser being fired through fibre FS007 during the scintillator phase, with 2.2 g/L PPO loading. The simulation used the optical

model described in Section 2.3.2. The earliest photon track associated with a given PMT hit was classified by the optical processes it underwent. A hit associated with a photon that travelled unimpeded through the detector is classified as ‘direct’. For this fibre and wavelength combination, the time residual windows used allow for a signal-to-background ratio of 98.8% for the far PMTs, and 99.7% for the backscattered PMTs, ignoring noise hits.

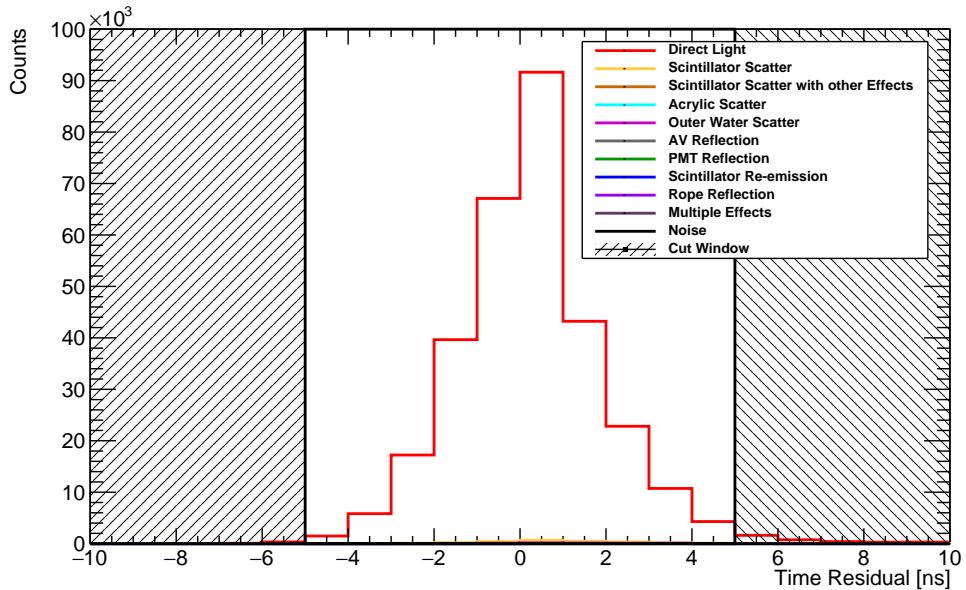
t_{med} was used in this analysis instead of t_2 because it was found to be far more robust to changes in emission intensity. Fig. 5.5 shows a comparison of the t_{res} distribution for backscattered PMTs between using t_2 and t_{med} , at different emission intensities. As can be seen, the t_2 distribution is biased towards positive t_{res} values as intensity increases. This is not the case when using t_{med} .

5.1.3 Results in Data

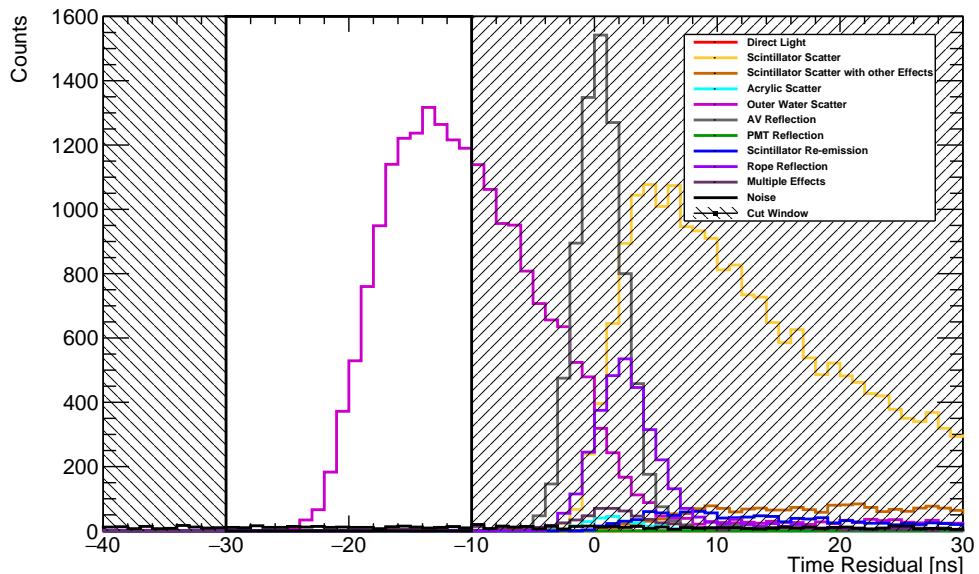
Datasets Used

The datasets used in this analysis are summarised in Table 5.1. The water phase data all came from run 114,018. Scintillator phase data was taken at five different points during the scintillator phase: two during the loading of PPO, and three afterwards. These datasets match those highlighted in Fig. 3.2.

‘Medium’ intensity subruns from the PQ407, PQ495, and SuperK lasers were used in this study. Results from the PQ446 laser are not shown, as many extinction length values produced had negative values (which are unphysical), likely due to the very substantial intensity instabilities seen in much of their water and scintillator phase subruns. The PQ375 laser was also not used, for reasons discussed shortly. Fibres FS193 and FS293 were not used as they failed to ever allow a substantial amount of light through them; FS207 was also not used as there was no beam profile distribution made for that fibre. Every scintillator phase subrun being used was matched to an equivalent water phase subrun.



(a) Far region



(b) Backscatter region

Fig. 5.4: Time residual distributions of the far and backscatter PMT regions for a simulation using laser PQ407 through fibre FS007, split by the different optical components. The time windows defining the signal regions are shown in both.

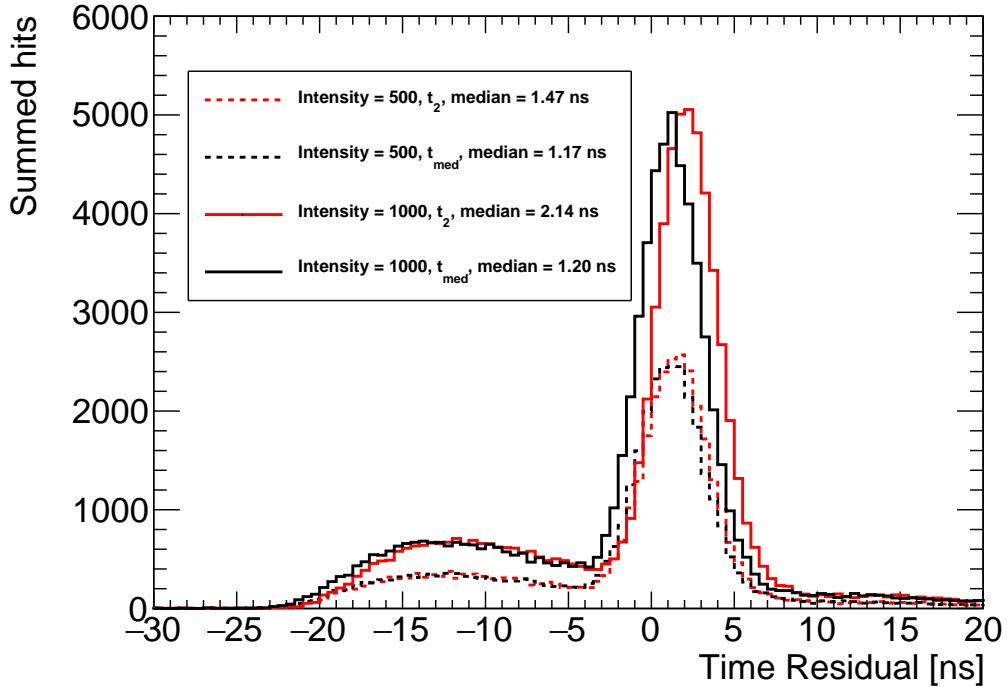


Fig. 5.5: Comparison of the backscatter light region time residual distribution, as a function of both emission intensity and emission time calculation choice. A clear intensity-dependence can be seen when using t_2 , whereas t_{med} is much more robust.

The sensitivity of the extinction length measurement is strongly dependent on the ratio of the extinction length to the length scale of the AV. Very long extinction lengths are difficult to measure because only a small amount of light is expected to be lost due to scattering or absorption. In contrast, at short extinction lengths only a very small amount of light will reach the other side without being extinguished; the dominant source of hits will come from scattered or re-emitted light.

As an example, Fig. 5.6 shows the expected optical component breakdown for hits from far PMTs from fibre FS007 due to laser PQ375, using the current RAT model of the 2.2 g/L LAB-PPO scintillator phase detector. Because the short extinction length at this wavelength relative to the diameter of the AV, typically there would rarely be more than one hit in the beamspot region per event. This meant that t_{med} would correspond to that single hit's time, and because there

Date	Detector Phase	Runs used
June 2018	Water Phase	114,018
May 2021	Scintillator Phase, 0.6 g/L PPO	270,856; 270,857; 270,858; 270,862
October 2021	Scintillator Phase, 1.1 g/L PPO	275,676; 275,680
May 2022	Scintillator Phase, 2.2 g/L PPO	300,708; 300,712; 300,715; 300,748
July 2022	Scintillator Phase, 2.2 g/L PPO	302,630; 302,634
June 2023	Scintillator Phase, 2.2 g/L PPO	310,294; 310,298; 310,303

Table 5.1: Datasets used in the SMELLIE extinction length analysis.

is an overlap between the beamspot and far PMTs for this fibre, a large spike at $t_{res} = 0$ can be observed. Importantly, the amount of direct light seen in the far PMTs in this simulation is entirely sub-dominant to the re-emitted light. Without accounting for this re-emitted light, this analysis would over-estimate the extinction length substantially. In contrast, it has already been shown in Fig. 5.4a that the contribution of re-emitted light to the PQ407 laser is expected to be negligible.

Although re-emitted light is sub-dominant in the far region, scattered light remains an important background. This becomes most relevant for the 10° and 20° fibres, as the direction of the far PMTs are somewhat distant from the beamspot, so the intensity of direct light is substantially less than that of the 0° fibres. A simulation of this is shown in Fig. 5.7 for fibre FS107 in the 2.2 g/L scintillator phase. The proportion of background components in the [−5, +5] ns window is substantial. As a result, any measurement of R_s would be systematically too large.

With similar simulations for all other fibres, it was found that all 0° fibres apart from FS093 achieved a signal-to-background ratio of over 95%. No other fibres were able to achieve a signal purity close to this value. Because of this,

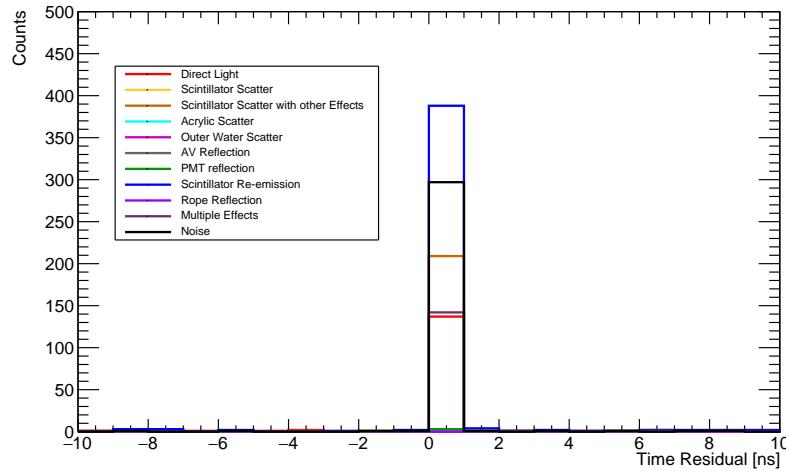


Fig. 5.6: Unstacked time residual distribution of a simulation of light from the PQ375 laser firing through fibre FS007, split by optical components. The direct light is clearly subdominant to light that has been absorbed and re-emitted by the scintillator. The spike at 0 ns arises from only one hit being observed in the beamspot for that event.

only these ‘pure’ fibres were used for actual extinction length calculations in this analysis. Data from the other fibres was still processed, but only used for comparison. Unlike all other fibres, FS093 was found to have a strong contribution to the far PMT hits from light reflected off of ropes. Although this rope reflection contribution should not be impacted by changes to the inner detector medium, it was decided to leave this fibre out of the extinction length calculations.

Results

The analysis results from the PQ407 laser are shown in Fig. 5.8. These have been split into a different plot for each time period. Each fibre is represented by an individual colour and marker shape pairing associated with the fibre’s pointing angle and node, respectively. Because 0° fibres had the far PMT region nearest the beamspot, they observed the largest values of R_w and R_s compared to the 10° and 20° fibres. This necessitated the use of log-log plots to effectively show the data. Lines of constant gradient, associated with scintillator extinction lengths in the range 5–120 m in 5 m steps, are shown as dashed grey lines. These lines

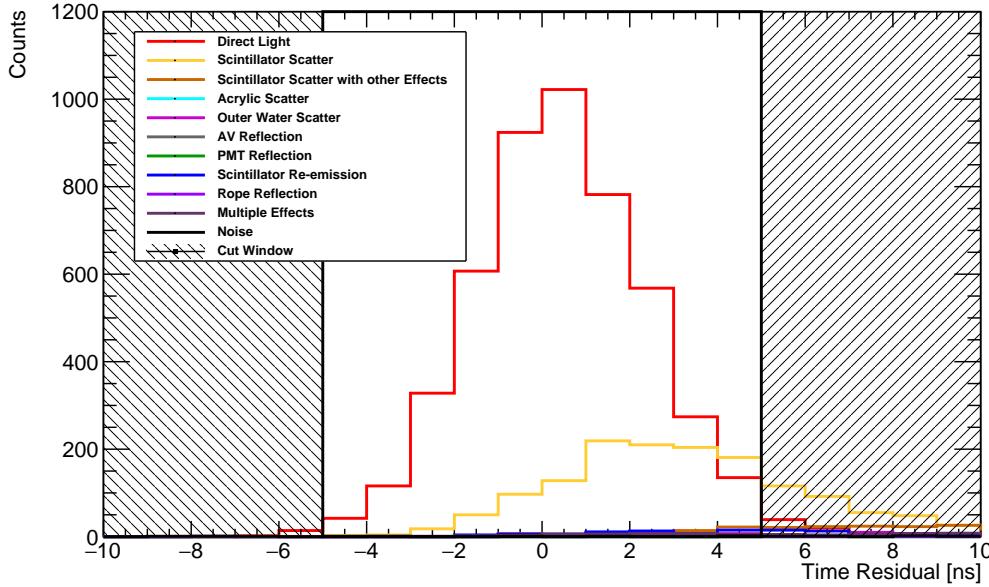


Fig. 5.7: Far PMT time window selection applied to a simulation of the PQ407 laser fired through laser FS107. The substantial background contribution from scattered light can be seen.

increase in extinction length from the top left down towards the bottom right, until there becomes a gradient at which the extinction length asymptotes. For gradients smaller than this asymptote value, a negative (unphysical) extinction length would be required. The change in distance between these lines show that longer extinction lengths are much harder to measure with the same precision as shorter extinction lengths.

As explained in the previous section, each plot had a straight line of the form $R_w = mR_s$ fit to the 0° data points, excluding fibre FS093. Because there were comparable uncertainties in both axes, this fit was performed by minimising the χ^2 term:

$$\chi^2(m) = \sum_j \frac{(R_{j,w} - mR_{j,s})^2}{\sigma_{R_{j,w}}^2 + m^2\sigma_{R_{j,s}}^2}. \quad (5.5)$$

An uncertainty on the gradient of the fit was found by determining the change in m required to increase the value of $\chi^2(m)$ by 1. Re-arranging Eq. 5.4, one can

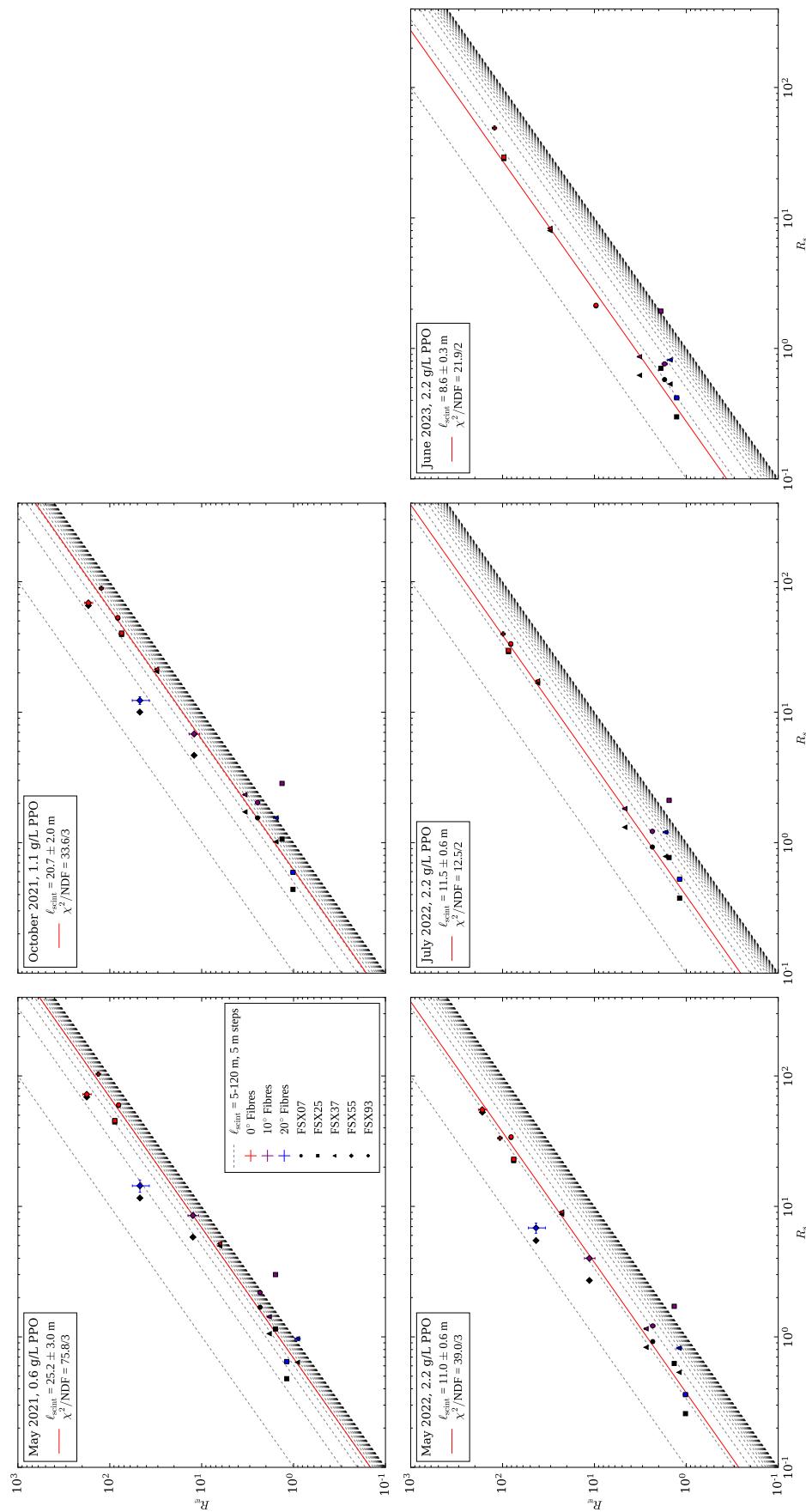


Fig. 5.8: Results of the extinction length analysis using the PQ407 laser.

derive the formula:

$$l_{\text{scint}} = \frac{L_s}{\ln m + \ln \left(\frac{T_s}{T_w} \right) + \frac{L_w}{l_w}}, \quad (5.6)$$

from which the extinction length in scintillator at a given wavelength at a given time period with propagated uncertainties was found.

As the concentration of PPO is increased in the detector, the extinction length at 407 nm appears to decrease, from (25.2 ± 3.0) m in May 2021 to (11.0 ± 0.6) m in May 2022. Then, the first two datasets taken in the 2.2 g/L scintillator phase appear consistent with one another. This is not true of the final dataset taken at the end of the scintillator phase, which measured a shorter extinction length again, of (8.6 ± 0.3) m.

Also shown in the plot are the results of fibres not included in the fit. As an example, fibre FS125 (shown as the purple square) is found to be generally in the non-physical region corresponding to negative extinction lengths, in the bottom-right half of the plots. This can be explained by the substantial scattering background expected in the far PMTs of this fibre. Assuming the simulation discussed in the previous section for this fibre in the 2.2 g/L scintillator phase, as well as matching simulations for the other PPO concentrations, this systematic effect in R_s was corrected for and shown with a black square. This was also shown for all other fibres. Because this systematic effect impacts only the values of R_s , the black points will always be shifted leftwards on the plot. As can be seen, by using the nominal correction from the existing scintillator optics models, the FS125 measurements are pushed into the physical regime, quite close to the line of best-fit.

Fig. 5.9 shows the data from the SuperK laser in the 400–410 nm range, for the May 2021, May 2022, and June 2023 datasets in which SuperK data was taken. Once again, there appears to be a substantial shortening of the scintillator’s extinction length as PPO was added, from (15.5 ± 1.4) m in May

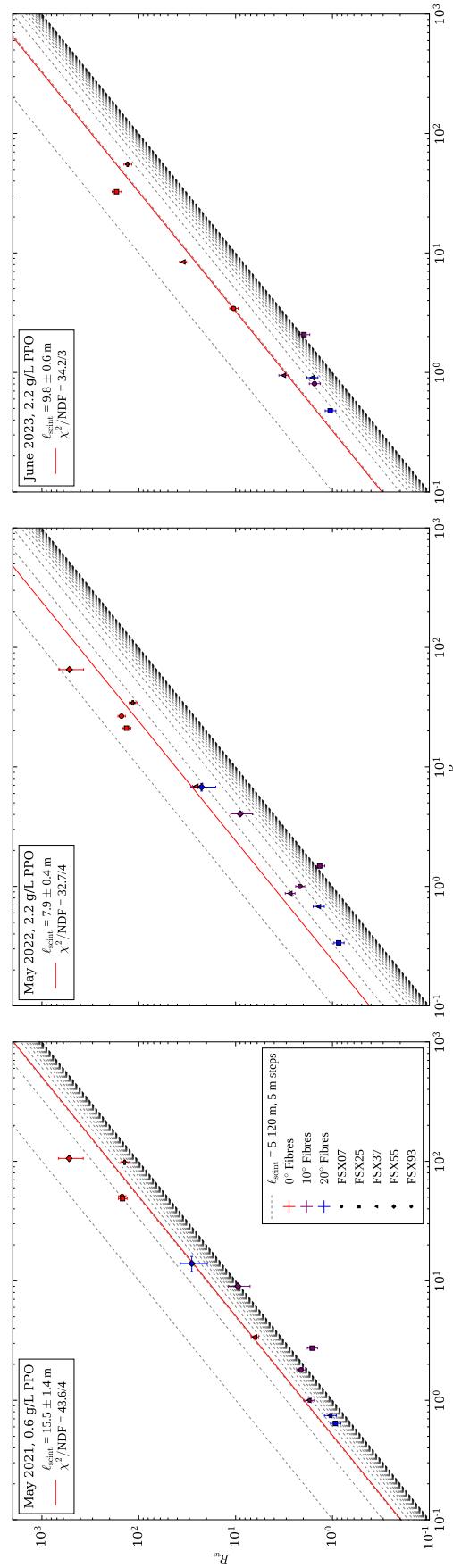


Fig. 5.9: Results of the extinction length analysis using the SuperK laser in the 400–410 nm range.

2021 to (7.9 ± 0.4) m in May 2022. These values differ from the PQ407 result, which could be explained by the SuperK laser measuring the extinction length averaged over a whole 10 nm range, weighted by the intensity spectrum of the laser. The third measurement result increases compared to the second here up to (9.8 ± 0.6) m, another contrast to the PQ407 laser.

One final thing shown in Figs. 5.8 & 5.9 is the χ^2 and number of degrees of freedom (NDF) in each of the fits. In all cases shown, the value of the ‘reduced χ^2 ’, equal to the ratio χ^2/NDF , is $\mathcal{O}(10)$. Because these values are much greater than 1, this indicates some inconsistency between the data and the fit model. The origin of this inconsistency is currently unknown; one possibility is a systematic effect that has not been considered in the analysis. The impact of this would be for the measured extinction lengths in this analysis to have substantially greater uncertainties than those currently quoted.

Fits to the SuperK data at wavelengths between 400–500 nm during May 2021, May 2022, and June 2023 were performed in the same manner as described above. The combined results for all of these measurements, as well as the PQ407 laser, are shown in Fig. 5.10. All three SuperK datasets show an increase in the extinction length with wavelength, as expected. There is also a substantial shortening in the extinction length going from the 0.6 g/L dataset to the 2.2 g/L dataset. This is matched qualitatively with the PQ data. Generally, data from May 2022 and June 2023 appear to be consistent with one another.

Comparisons with other Measurements

It is important to compare the measurements of the scintillator extinction lengths using SMELLIE to other methods used by the Collaboration. The optical model used in RAT for the 0.6 g/L scintillator was built from bench-top measurements prior to scintillator phase [95, 117]. The extinction length associated with this model is shown as a black line in Fig. 5.10. The results from this analysis with

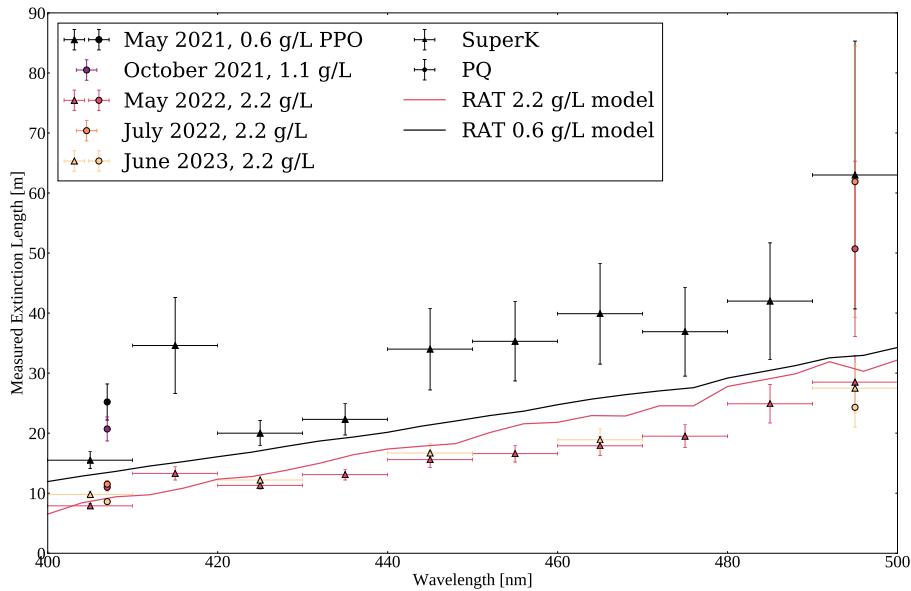


Fig. 5.10: Summary of scintillator extinction length measurements made by SMELLIE, with a comparison to the current 2.2 g/L optics model used in RAT.

SMELLIE data give extinction lengths up to 20 m longer than the existing model, for both the PQ and SuperK laser over all wavelengths in the 400–500 nm region.

For the 2.2 g/L scintillator, more techniques were used for measurement of the extinction length. One approach by Ben Tam was to use ‘UV-vis’ spectroscopy on samples of the scintillator extracted from the detector [151, 152]. Another method, used by Serena Riccetto, was to look at the `nhits` distribution of the radioactive background ^{210}Po within the detector: a discussion of this and other backgrounds in the detector can be found in Section 6.2. The peak `nhit` value associated with this background varies with the position of the event, because of the changing lengths of scintillator the event’s scintillation light must travel through. Different optical models could then be tested by comparing the peak `nhits` of the ^{210}Po in simulation to data [153].

The current model of the 2.2 g/L scintillator absorption and scattering used in RAT was developed by combining the measurements of B. Tam and S. Riccetto with

older bench-top measurements made before the scintillator phase: details about this model-building can be found in [96, 101]. The extinction lengths obtained from this model are shown as a red line in Fig. 5.10. There appears to be good agreement between these model values and the results from the analysis with SMELLIE data taken in May 2022 for wavelengths in the ranges 400–430 nm and 490–500 nm. At medium wavelengths, SMELLIE appears to measure extinction lengths up to 5 m shorter than the existing RAT model. These results appear consistent between the May 2022 and June 2023 datasets for the SuperK data; there is some disagreement remaining for the PQ lasers. This could be a result of underestimating the total uncertainty of the PQ measurements, as discussed in the previous section.

On its own, measurements of the extinction length of the scintillator cannot disentangle changes in scattering from absorption. A second, complementary analysis is needed.

5.2 Scattering Analysis

5.2.1 Historical Approaches and the Problem of Systematics

A method for measuring scattering lengths in the detector using SMELLIE was first developed by K. Majumdar [139]. This method was used by E. Turner to measure the scattering length of the UPW in the water phase [115]. A similar approach was built by S. Langrock, and tested on fake water phase data [141]. The general analysis technique was as follows:

1. Using simulations under nominal conditions, a region of PMTs and associated time windows was selected that optimised the sensitivity to hits from photons

that were scattered inside the AV. This signal region was defined by a geometric region of the 2D (t_{res}, α) parameter space.

2. Data was taken using the SMELLIE system.
3. A calibration curve of the fraction of hits in the scattered signal region to that of the whole detector as a function of the scattering length for a given fibre and wavelength, was built using simulations.
4. The fraction of hits in the signal region was determined for each subrun of SMELLIE data, and then the calibration curve was used to derive values of the scattering length at that wavelength.

The challenge with this analysis approach is that it relies on the use of SMELLIE simulations to derive results. There are substantial systematic uncertainties in the modelling of the SMELLIE beam profiles, which can readily propagate into the uncertainty of the scattering length measurement. Both S. Langrock and E. Turner found this in their analyses [115, 141]. Furthermore, the signal regions chosen historically had a substantial background contribution, which often appeared to not be modelled with sufficient accuracy.

As a demonstration of these problems, consider a comparison between PQ495 laser data taken in July 2022, through fibre FS055, to a simulation of the same setup. This is shown in Figs. 5.11a & 5.12. Also shown in these plots is the scattering signal region as determined by K. Majumdar and E. Turner’s method. One can qualitatively observe features that differ between the data and MC: there are differences in the distribution along the line $t_{res} = 0$ due to beam profile systematics, as well as differences in the features at large angles and times. Although the signal region was chosen to try and optimise the statistical significance, it failed to account for the systematic uncertainties present in the SMELLIE model. To further emphasise this, Fig. 5.11b shows the impact of

using a modified beam profile in the simulation. Instead of using the beam profile developed for that fibre in Section 4.2.1, one was used that had data from only one water phase subrun, as shown in Section 4.1.2. The fraction of hits in the signal region changes from 31.2% to 27.7%.

There are two major approaches to handling the problems described above. One approach is to build an optical model of the detector and SMELLIE that has negligible systematic uncertainties, so that the existing analysis approach can be used with minimal problems. An alternative approach, and the one used in this thesis, is to build a new analysis that is more robust to the existing systematic uncertainties in both the SMELLIE beam profile and detector optics.

5.2.2 New Methodology

Signal Region Selection

A new PMT and time region was found that had substantially higher scattered signal purity than previous selections, whilst maintaining large signal statistics. Fig. 5.13 shows a simple 2D model of the geometric optics of SMELLIE in the scintillator phase. As can be seen, refraction due to the larger refractive index of LAB and acrylic compared to UPW causes a substantial region of the PSUP to be unreachable by direct light-paths. In the full three dimensions of the detector, this region corresponds to a broad ring of PMTs.

For this analysis, this new scattering signal region will be known as the “bad light-path” region, because no direct light-paths can be found. By consequence, it is impossible to calculate a time residual for hits in this region, as no direct light-path time-of-flight calculation is possible. Instead, times are calculated simply by $t = t_{\text{hit}} - t_{\text{emmm}}$, where the usual t_{med} method is used for calculating the emission time as described in Section 3.5. For each fibre, the bad light-path region

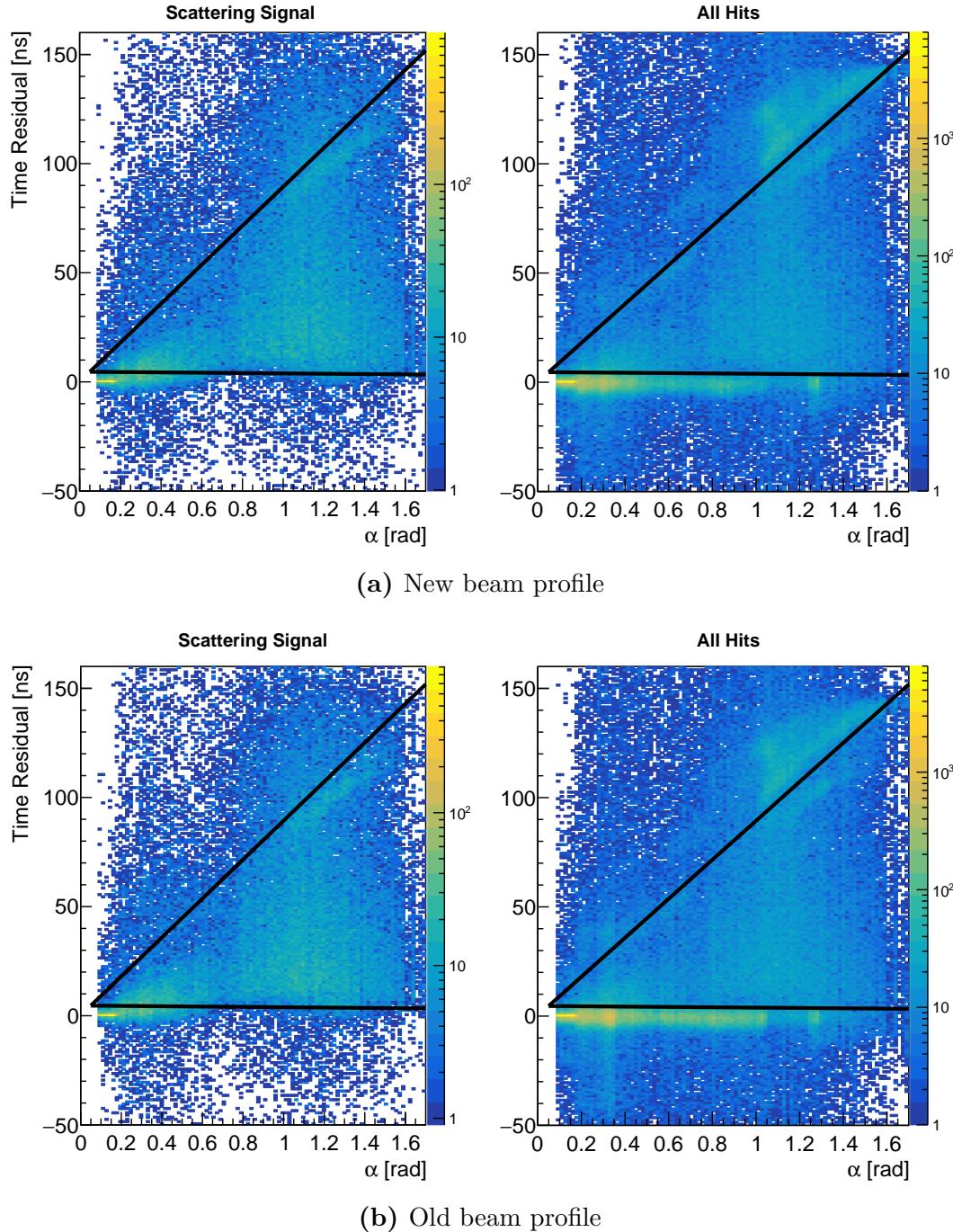


Fig. 5.11: Comparison of the signal and total hit distribution in the (t_{res}, α) -space in MC, using new and old beam profiles for laser PQ495 through fibre FS055. The triangular scattering signal region of K. Majumdar's analysis are shown, optimised for the new beam profile simulation. The fraction of all hits in this signal is 31.2% for the new beam profile simulation, and 27.7% for the old beam profile.

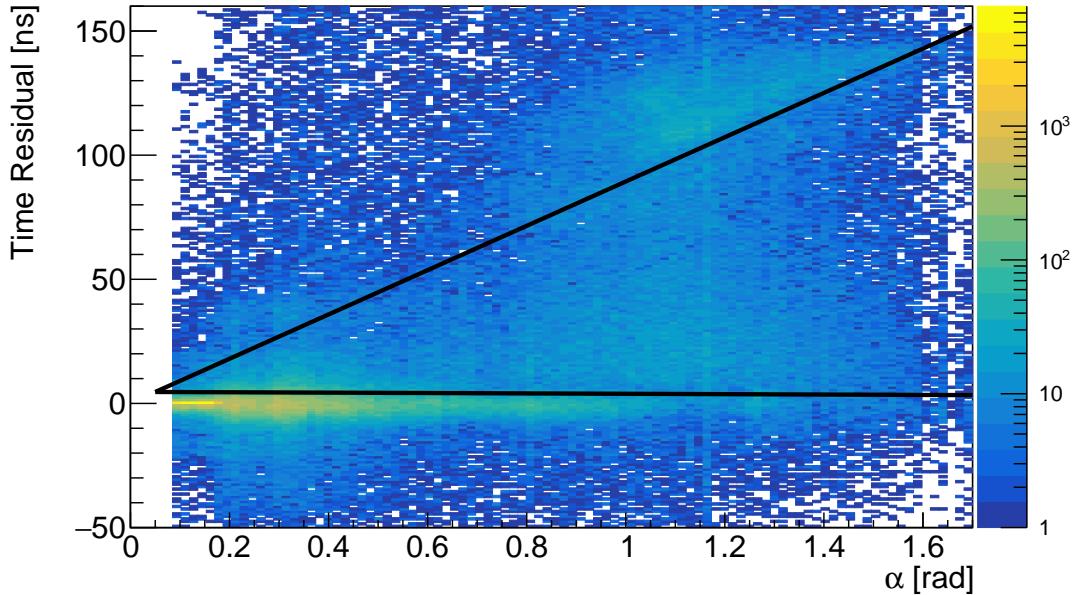


Fig. 5.12: Hit distribution of data taken in run 302,634, with laser PQ495 through fibre FS055, as a function of both t_{res} and α . The triangular scattering signal region optimised from Fig. 5.11a.

was defined by running SMELLIE simulations for each fibre, and selecting PMTs in the forward hemisphere that observed no hits from direct light.

Fig. 5.14 shows the time distribution of hits in the scattering signal region, split by the different components of a simulation, using the PQ495 through fibre FS055. As can be seen, there is a large peak in the $[-25, +30]$ ns time window that is dominated by light scattered by the scintillator. It is this time window that is used in this analysis to isolate the scattered signal for all fibres.

In addition to having a high signal purity, this new scattered signal region has the benefit of being robust to beam profile uncertainties. Although beam profile systematics impact the intensity distribution of the direct light, no such light is seen in the signal region. Moreover, because scattered light must have a large scattering angle to reach the signal region, the beam profile shape is ‘washed out’ by the scattering. As a demonstration of this effect, Fig. 5.15 compares the observed time distribution in the bad light-path region for the new and old beam

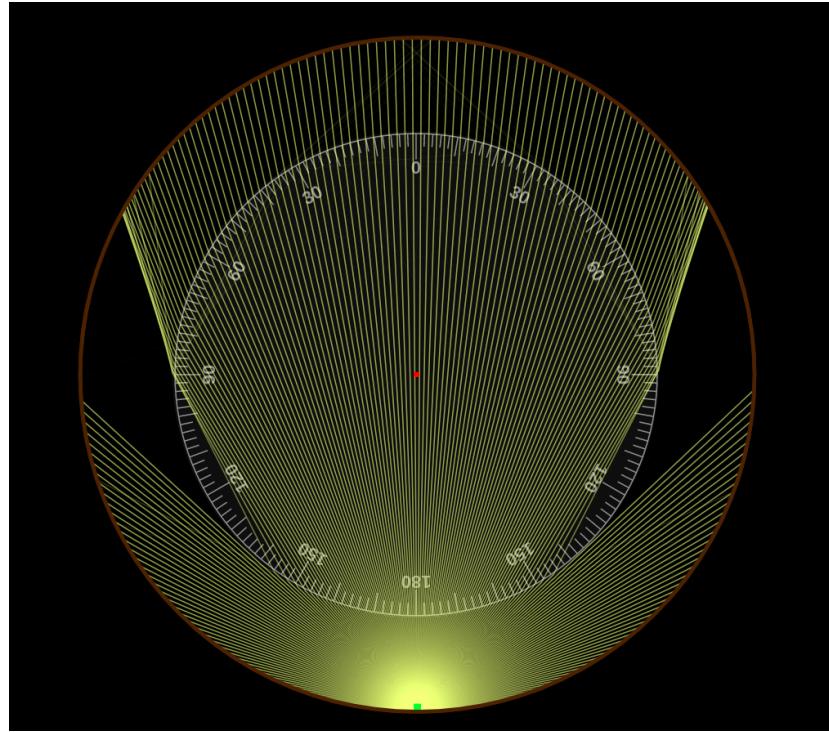


Fig. 5.13: Model of light rays propagating in the SNO+ detector during scintillator phase, emanating from an external source such as SMELLIE. Because of refraction, a large ringed region is formed where no direct light can reach the region from the source. This forms the basis of the new scattering signal region.

profiles of FS055, still using the PQ495 laser. There is only a change of 1.0% in the signal time window, compared to the more substantial differences seen in Fig. 5.11.

In the new signal region, some backgrounds to scattering are still present. These come from PMT noise, scintillator re-emission, reflections off of PMTs and their concentrators, and combinations of multiple optical effects. A substantial fraction of the paths which undergo multiple processes scatter in the scintillator, and so are treated as signal for the purposes of this analysis. The contribution due to noise hits can be readily determined through the use of PULSEGT triggers, through the same approach used in Section 5.1.2. In the example shown in Fig. 5.14, 82% of non-noise hits in the signal region were from photons which underwent scattering in the scintillator.

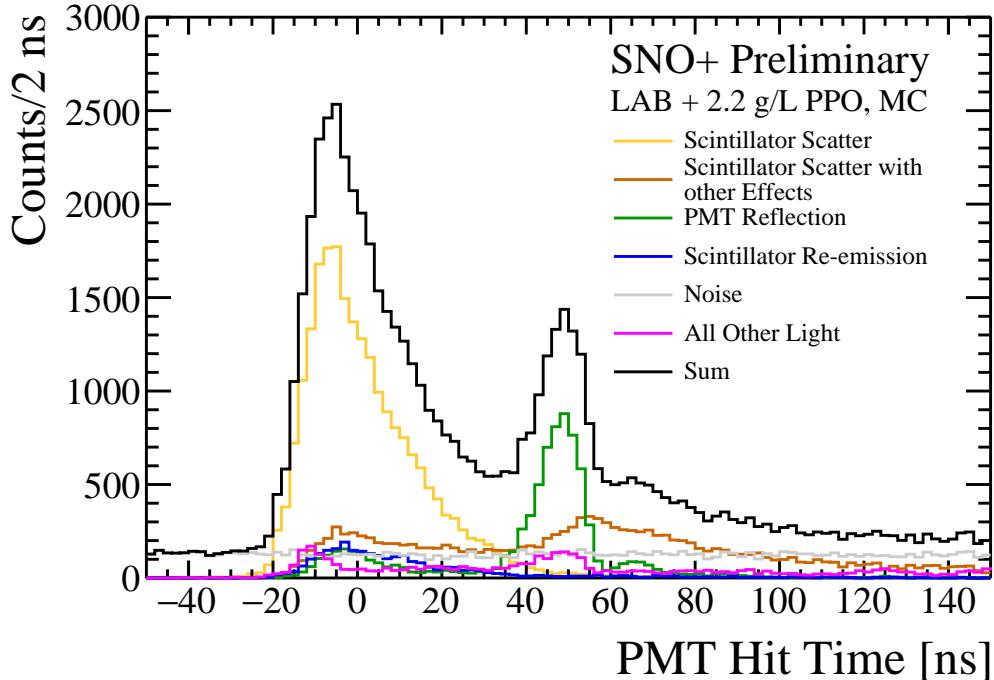


Fig. 5.14: Timing distribution of a simulation of the observed hits in the bad light-path PMT selection through fibre FS055, using the PQ495 laser. The distribution is split into the different optical components.

Another thing to check was whether natural changes in the relative position of the AV centre with respect to the PSUP over time could leak direct light into the signal region. This offset is determined through a calibration system stationed at the top of the neck of the detector, known as the Neck Sense Ropes, whose measurements have been validated during the partial fill phase with a laser survey [154, 155]. These AV offset measurements are saved in the RATDB, and used within any simulation of a given data run in RAT. Of the five SMELLIE scintillator datasets mentioned in Table 5.1, the largest recorded change in AV offset was between May and July 2022, in which the AV moved vertically by 39.5 mm. The bad light-path PMT region used simulations with run conditions from July 2022 data; for comparison, a simulation equivalent to the that of Fig. 5.15 was made using the run conditions of the May 2022 dataset. Only a negligible amount of hits from direct light leaked into the signal region in this alternative simulation.

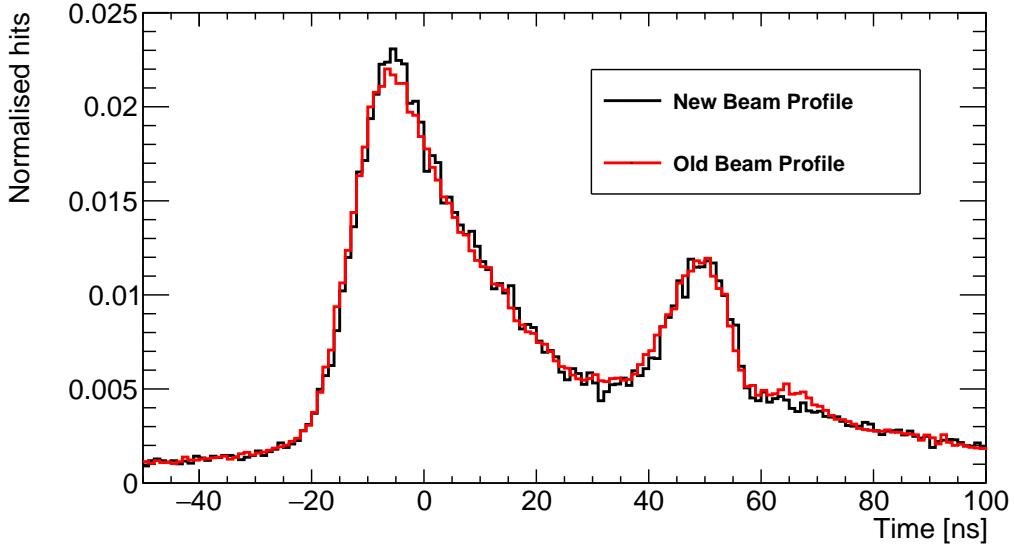


Fig. 5.15: Comparison of timing distributions in the bad light-path PMT region for simulations of laser PQ495 through fibre FS055, comparing the old and new beam profile. Both distributions have been normalised by the number of events. The fractional difference in the number of hits in the signal time window is 1.0%.

Observing Changes in Scattering

If the absolute event-to-event intensity of a SMELLIE subrun is known, and the models of the scintillator re-emission and PMT reflections are well-enough known for the wavelengths of interest, then the scattering length of the scintillator can be determined by comparing data in the signal region to matching simulations of varying scattering length. Sadly, at the time of writing, no method for determining the absolute intensity which is sufficiently robust to systematic uncertainties in the modelling of MC is known. Instead, it was decided for this analysis to use the back-scattered light region described in Section 5.1.2 as a measure of relative intensity.

Without knowledge of the absolute intensity, absolute scattering lengths cannot be measured. However, by calculating the ratio of npe in the signal region to that of the back-scatter region for two scintillator subruns using the same laser, wavelength range, and fibre, denoted R_1 and R_2 , any substantial change in the

scattering length between datasets can be observed. Another restriction that this analysis has is that changes in the scattering length cannot be distinguished from changes in the background processes, in particular PMT reflections and scintillator re-emission. The reflectivity of the PMT concentrators is known to be changing slowly with time as the reflective petals of the concentrators degrade in the outer water [85]. A. S. Inácio determined from comparing Laserball data taken in SNO and SNO+ that the PMT angular response had worsened over a period of 12 years by up to 8% in the incident angle range up to 42° [79]. Therefore, an upper estimate of the fractional change in the PMT bucket reflectivity over the course of the two-year period in which SMELLIE scintillator phase data was taken is 1.4%, assuming the same trend. Using the example shown in Fig. 5.14, 6.5% of non-noise hits in the signal region came from PMT reflections, meaning that the overall systematic effect of concentrator degradation is expected to be merely 0.09%.

The other dominant source of background is scintillator re-emission. This is expected to change as a function of PPO concentration. However, because the absorption length of the scintillator at 2.2 g/L is substantially longer than the scattering length in the 400–500 nm range according to the current RAT model, the proportion of this background component is expected to be small: see Figs. 2.2 & 2.4. According to the simulation shown in Fig 5.14, 7% of non-noise hits in the signal region come from scintillator re-emission. According to simulations which use the current RAT model, this fraction of re-emitted light stays roughly constant over the 400–500 nm range.

5.2.3 Results in Data

The datasets used in this scattering analysis are the subset of those described in Table 5.1, which were taken in the scintillator phase by the SuperK laser between

400 nm and 500 nm. Only three scintillator datasets had SuperK data taken: May 2021, May 2022, and June 2023. The npe ratio for the first dataset is denoted by R_1 , and used as the baseline to compare to the later two datasets, denoted R_2 .

Results from the same wavelength for the same dataset pair were combined onto a plot, as shown in Fig. 5.16. Because far more hits are observed in the signal region compared to the backscatter light region for all subruns, the dominant source of statistical uncertainty in R comes from the latter region. As a result, data points with large R values had much larger uncertainties, because these corresponded to subruns with lower statistics in the backscatter light region. In order to account for the non-negligible uncertainties along both axes, the same fitting procedure from Section 5.1.3 was used. This found the best-fit line $R_2 = m \cdot R_1$ for each plot with an associated uncertainty.

The results for the gradient of all the data considered in this analysis are shown in Fig. 5.17. All values for the ratio R_2/R_1 are found to be less than 1, varying almost linearly from 0.70 to 0.95 over the 400–500 nm wavelength range. This implies that, at all wavelengths in this range, less light was observed in the scatter signal region of the 2.2 g/L scintillator than the 0.6 g/L scintillator. Furthermore, there is a strong wavelength-dependence to this effect, such that shorter wavelengths observe a substantially lower fraction than at longer wavelengths. These results appear consistent to within statistical uncertainties for both 2.2 g/L scintillator datasets considered.

5.2.4 Interpretation

A naïve reading of the above results would say that the amount of scattering has decreased by up to 30% between May 2021 and May 2022. Given that the scattering length is inversely proportional to intensity of scattered light, this would correspond to an increase in the scattering length of up to 40%. However, that

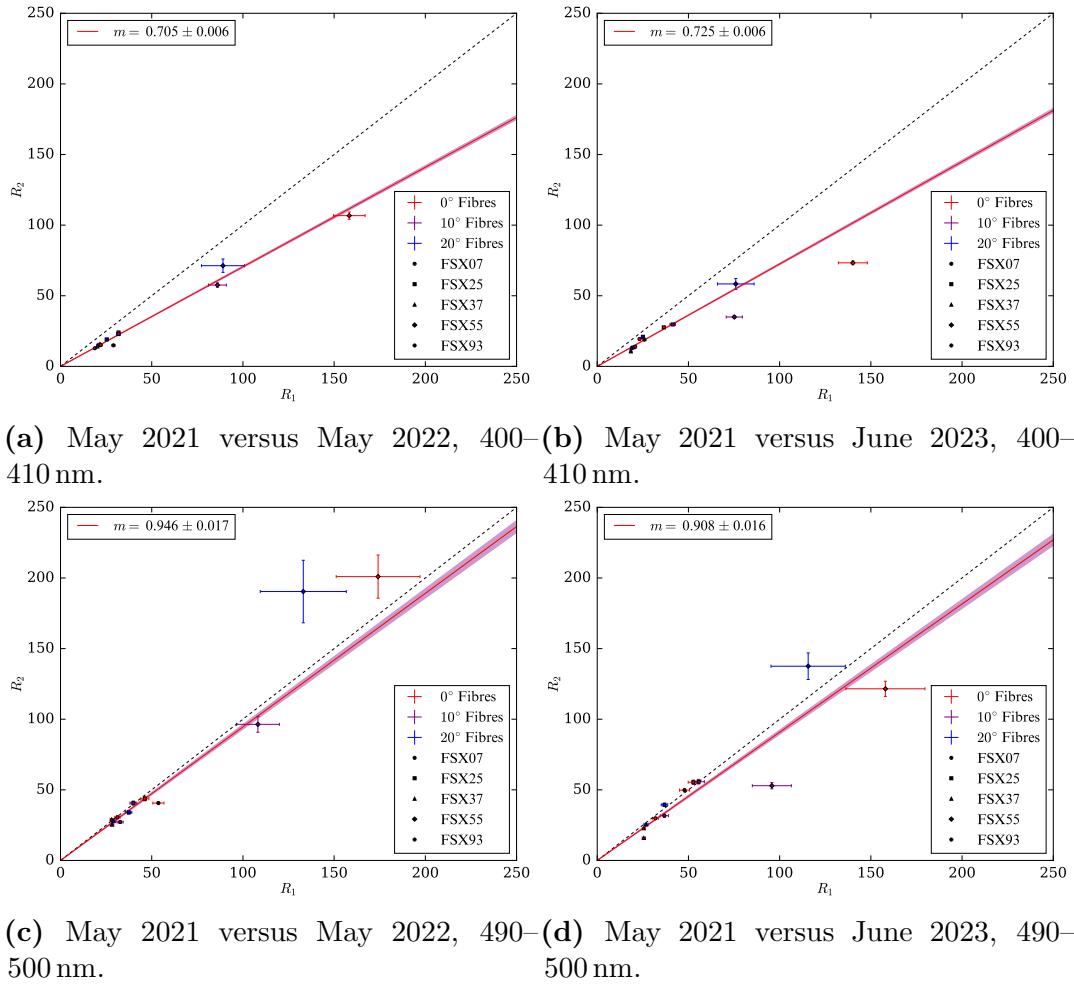


Fig. 5.16: Plots of R_1 against R_2 for four pairs of scintillator phase datasets, using the SuperK laser.

would be incompatible with the results of the analysis of the previous section, which clearly shows a significant decrease in the observed extinction length over the same time period. It is also very difficult to explain from a theoretical perspective how the addition of PPO to the scintillator cocktail could decrease the amount of Rayleigh scattering.

There is an alternative interpretation which fits the observations of both analyses. Suppose that the scattering length did not change with time, but there was an additional non-re-emitting component of the scintillator added during the PPO loading. Then, this new component of the scintillator would be capable of absorbing light, shortening the observed extinction length, but the

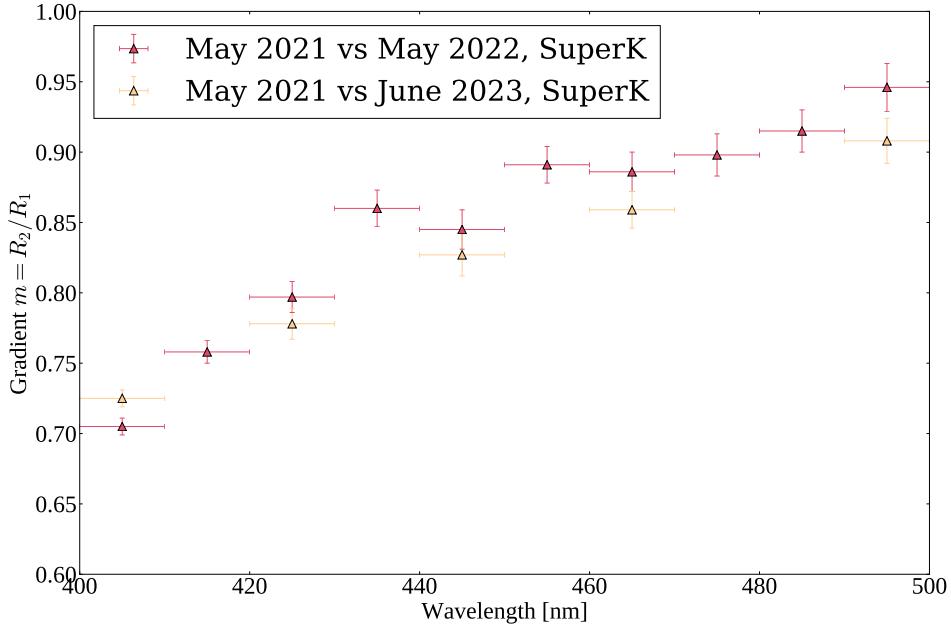


Fig. 5.17: Summary of results from the scattering analysis, shown as a function of wavelength.

amount of re-emitted light could be much less than expectation in the wavelength range considered in this work. This would lead to a decrease in the amount of light observed in the scattered signal region, as the background re-emitted light substantially decreased in quantity. The expected fraction of light arriving in this region due to scintillator re-emission is, as seen in Section 5.2.2, of the correct magnitude to explain the change in the scattering signal region.

In the process of building the current 2.2 g/L scintillator absorption and scattering model, S. Riccetto and B. Tam found that adding a non-re-emitting absorption component was necessary to fit their calibration data simultaneously [96]. The origin of this new component was investigated by B. Tam, who concluded that near the end of the PPO top-up campaign some contaminants were likely added from the accidental overheating of the LAB-PPO during one of the distillation processes [156]. There is a slight difference between their results and the implication from work in this thesis. S. Riccetto and B. Tam saw little evidence of this

new component impacting longer wavelengths, and instead preferred causing a significant impact at short wavelengths $\lambda \lesssim 400$ nm.

5.3 Summary and Suggestions for Further Work

In this chapter, two analyses were developed and performed on data taken with the SMELLIE calibration system, to understand the optical properties of the scintillator within SNO+. The first analysis measured the extinction length of the scintillator as a function both of time and wavelength, through comparison to the already-calibrated water phase data. The extinction lengths found had some differences with existing models, although both see an increase in the extinction length seen as a function of wavelength of at least 20 m between 400–500 nm. A decrease of at least 10 m in the extinction lengths were observed between the May 2021 and May 2022, when the PPO concentration of the scintillator increased from 0.6 g/L to 2.2 g/L. However, the change in extinction length was far greater than the expected change to the absorption length of PPO, ~ 5 m. Data taken one year later using the SuperK laser, in June 2023, shows no indication of any further change to the scintillator’s extinction length distribution. Data taken with the PQ laser does show some change, but it is possible that this is only because the uncertainties have been underestimated by some means.

In the second half of this chapter, another analysis was built, designed to be sensitive to changes in Rayleigh scattering. Simulations indicated that this new analysis was also sensitive to changes in the scintillator’s re-emission properties or changes in the PMT bucket reflectivity. By performing relative measurements between scintillator datasets, it was found that the amount of light seen in signal region decreased between May 2021 and May 2022 between 5% and 30%, with similar results being seen when comparing May 2021 to June 2023 data. This could only be explained by a decrease in the quantity of scattered light or scintillator

re-emission. A significant wavelength-dependence was also seen in these results, with shorter wavelengths observing greater decreases in the amount of light seen.

The best physical explanation that combines these results is the existence of a new component of the scintillator cocktail that absorbs, but does not scatter or re-emit light substantially in the 400–500 nm range. This would generate a significant shortening of the extinction lengths over these wavelengths, but decrease the amount of light seen in the signal region of the scattering analysis. This theory has similarities to the observations seen by B. Tam in *ex-situ* measurements, as well as the fits to the spectra of *in-situ* backgrounds performed by S. Ricetto. However, the magnitude of the extinction length shortening appears much more substantial in the SMELLIE data than the other measurements.

The key insight that enabled both of the analyses in this chapter is the building of analysis methods that were robust to the major systematics present in SMELLIE. By not comparing SMELLIE data to simulation, the discrepancies between the RAT model of SMELLIE and reality, described in Section 4.3, could be avoided. However, not using simulations in these analyses does come at a price. Firstly, the extinction length analysis becomes fundamentally dependent on the water phase data taken by both the Laserball and SMELLIE. The former had large systematic uncertainties in their UPW attenuation measurements that puts a firm limit on the precision that any extinction length measurement can make.

As an example, consider a measurement of the scintillator’s extinction length at 446 nm, with a ‘true’ value of the extinction length of 12 m. If the uncertainties from all other factors could be made negligible, because the derived attenuation length from the water Laserball data is (108 ± 49) m, this propagates to an uncertainty on the extinction length of the scintillator of 5%. This effect becomes even larger as the extinction length increases.

This limitation is further compounded by the limited statistics taken in the SMELLIE water phase subruns used in this analysis. The extinction length

analysis used ‘medium’ intensity subruns, with the rare UPW backscatter light being used for intensity calibration, where for multiple PQ lasers there were very large event-to-event intensity variations. In addition, by needing to determine the emission time for each event independently, instead of being able to rely on the EXTA trigger timing, a small fraction of events may have leaked into or out of the time windows used in both analyses. This effect would be most predominant in data with a low number of hits in the beamspot region.

The other major source of uncertainty in the results of both analyses comes from the scatter between fibres around the line of best fit for a given plot. For example, in Fig. 5.16d one of the points with the strongest pull on the fit value of the gradient is the purple triangle associated with fibre FS037. The backscattered light distributions that give rise to this data point are shown in Fig. 5.18. As can be seen, there is a prominent change of shape of the peaks near $t_{res} = 0$ ns, which appears to leak into the backscattered light time window. The origin of this shape change is not yet known, but only appears to affect this fibre in the most recent scintillator dataset. This problem could have minor impacts on the particular fit values which use this data, but all the qualitative results of this chapter will remain.

There are a series of improvements that can be made to these two SMELLIE analyses, based upon the comments above. Major progress could be achieved if a robust measure of a fibre’s absolute emission intensity could be found. In so doing, measurements of the extinction lengths and scattering properties of the detector media could be made independent of Laserball data or having to make merely relative measurements. This could become even more powerful if the beam profiles of the fibres for different wavelengths could be known with sufficient precision.

Another important set of considerations that will need to be made in the future in order to make precision scattering length measurements is the calibration of the scintillator re-emission and PMT reflections. It is worth noting that the analysis

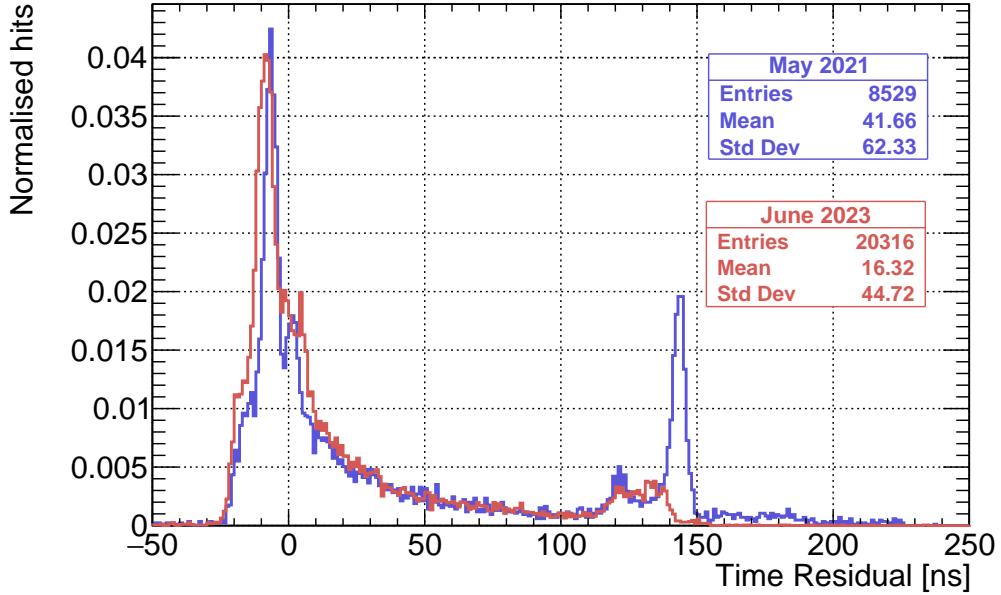


Fig. 5.18: Observed time residual distributions for the backscattered light region of PMTs, with the SuperK laser in the 490–500 nm range through fibre FS037. The distributions from both May 2021 and June 2023 are compared, with both of their distributions normalised to 1. The lack of the peak at 145 ns in the June 2023 data likely comes from the changing of the TUBii trigger delay during the hardware upgrades of Summer 2022: see Section 3.5 for more details.

of Section 5.2, despite being designed for measuring changes in scattering length in the scintillator, actually appears to have measured a change in the re-emission properties of the scintillator. The impact of PMT reflections over time can be constrained by isolating regions of PMTs and time for which there is high purity of reflected light, such as the peak seen at 50 ns in Fig. 5.14. Alternatively, these could be calibrated through a deployment of the Laserball during the scintillator phase.

A possible approach to disambiguating between re-emission and scattering in scintillator is through their different scattering angle distributions. Whilst the former emits light isotropically, Rayleigh scattering has the angular form $(1 + \frac{1-\delta}{1+\delta} \cos^2 \theta)$, as discussed in Section 2.3.2. To do this, a secondary measurement of the scattered and re-emitted light could be made in the far PMT region. Because it is known that it is this light which is the dominant background to the direct

light for the extinction length analysis, the amount of this background light which enabled the 10° and 20° fibres to best line up with the fit results of the 0° ones could be determined. Then, this amount could be compared to the amount seen in the bad light-path signal region. One expects a difference in the observed amounts of scattered and re-emitted light in these two regions, because of the anisotropy of the Rayleigh scattering. With the differing values of these two different datasets, one might be able to tease apart the individual contributions of scattering and re-emission, and hence measure both.

Chapter 6

Solar Oscillation Analysis

Driving out into the Sun

Let the ultraviolet cover me up

Looking for a Creation Myth

Ended up with a pair of black lips

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.

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 [157], which made an initial estimate of the sensitivity of SNO+. 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 [90], and more broadly from the general techniques used in the $0\nu\beta\beta$ analysis of Tereza Kroupova [137] and Jack Dunger [158].

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

6.1 Observational Principle

There are three key properties of solar neutrinos and their interactions that enable neutrino oscillations to be measured. The first of these is the dependence of the electron neutrino survival probability, P_{ee} , on both the individual oscillation parameters and the neutrino energy. As discussed in Section 1.2.2, P_{ee} is dependent on the oscillation parameters θ_{12} and θ_{13} when propagating through the vacuum of space (after averaging over very many oscillations), and the effective oscillation parameters θ_{12}^M and $\Delta m_{M,21}^2$ when passing through the Sun or Earth. By Eq. 1.14, this leads to an overall dependence of $P_{ee} = P_{ee}(\theta_{12}, \theta_{13}, \Delta m_{12}^2)$.

Fig. 6.1 shows the dependence of $P_{ee}(E)$ on each of these oscillation parameters, as well as the oscillation parameter Δm_{13}^2 for comparison. It is clear that only the parameters Δm_{21}^2 and θ_{12} have a substantial impact on $P_{ee}(E)$. θ_{13} does have a small effect on the survival probability, but the global fit measurement of $\sin^2 \theta_{13} = 0.02220^{+0.00068}_{-0.00062}$ [60] has uncertainties over an order of magnitude smaller than the range of $\sin^2 \theta_{13}$ values scanned over in Fig. 6.1. Because of this, only Δm_{21}^2 and θ_{12} will be varied in this analysis, with θ_{13} and Δm_{13}^2 being fixed at the global fit values shown in Table 1.1.

This dependence of P_{ee} on the oscillation parameters is not enough on its own to allow Δm_{21}^2 and θ_{12} to be measured. If a purely NC weak interaction

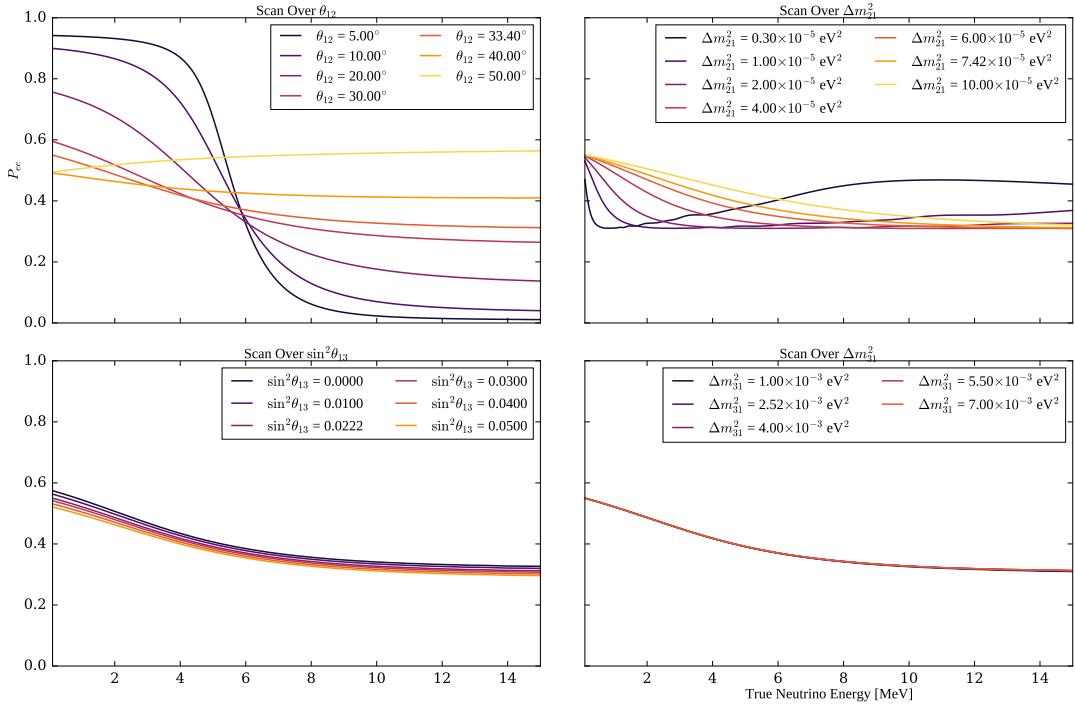


Fig. 6.1: P_{ee} as a function of true neutrino energy, scanned over a variety of oscillation parameter values. For a given oscillation parameter being scanned over, all other oscillation parameters are set at the NuFit 5.1 global fit values. P_{ee} values are calculated using the `PSelmaa` algorithm, as described in Section 6.3.6.

was used to detect the solar neutrinos, then there would be no way of telling the flavour-state of an interacting neutrino. The second property that enables solar neutrino oscillation measurements is that of neutrino-electron elastic scattering, the interaction mechanism used in this analysis. As discussed in Section 1.2.1, this interaction has an additional CC mode for electron neutrinos only, on top of an NC mode for all flavours of neutrino. This modifies the differential cross-section for electron neutrinos compared to other neutrino flavours. As a result, the resulting differential interaction rate R for solar neutrinos in the detector as a function of the neutrino energy E_ν , is given by:

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

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 for all flavours, 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, neutrino energies are not directly measured in the detector — only the associated scattered electron. The final property needed to observe solar neutrino oscillations is a correlation between the energies of the incident neutrino and the scattered electron. Using the formula in Eq. 1.4, the differential cross-sections of electron neutrinos at different incident energies, as a function of the scattered electron's kinetic energy is shown in Fig. 6.2. For a given value of E_ν , conservation of energy limits the maximum allowed kinetic energy of the scattered electron, T_{\max} , to be:

$$T_{\max} = \frac{2E_\nu^2}{m_e + 2E_\nu}. \quad (6.2)$$

This kinematic limit can be very clearly seen in the shapes of the differential cross-sections. As a result, a higher energy neutrino will generate a higher energy scattered electron, on average. Fig. 6.2 also demonstrates that the correlation between neutrino and electron energies is rather weak.

The combined effect that each of the above physical processes has on the energy spectrum of the solar neutrinos can be seen in Fig. 6.3. A broad energy distribution of ${}^8\text{B}$ electron neutrinos are generated in the Sun. These neutrinos then oscillate their flavour state as they propagate to the detector, in an energy-dependent manner. When neutrinos interact with the electrons in the detector, there is both an energy- and flavour-dependence on the cross-section. The scattered electrons gain a kinetic energy with some mild dependence on the inciting neutrino's energy, which is then measured by the detector to within some energy resolution.

The overall effect of changing the oscillation parameters on the observed signal distribution can be seen in Fig. 6.4. There is some shape change in the spectrum

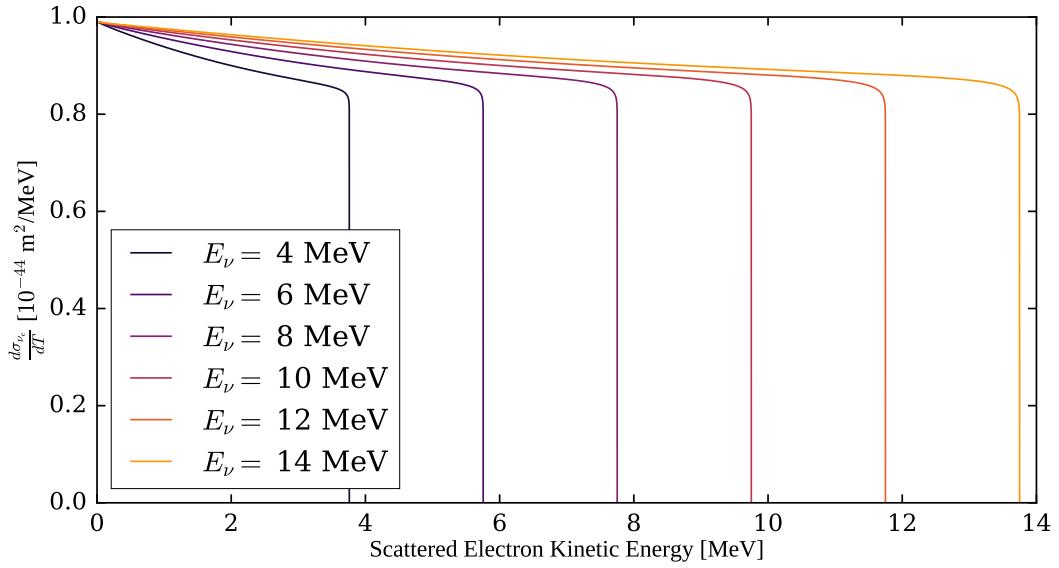


Fig. 6.2: Plot showing the differential cross-section for neutrino-electron elastic scattering of electron neutrinos, $\frac{d\sigma_{\nu_e}}{dT}$, as a function of the kinetic energy of the scattered electron, T . This is shown for a variety of different incident neutrino energies.

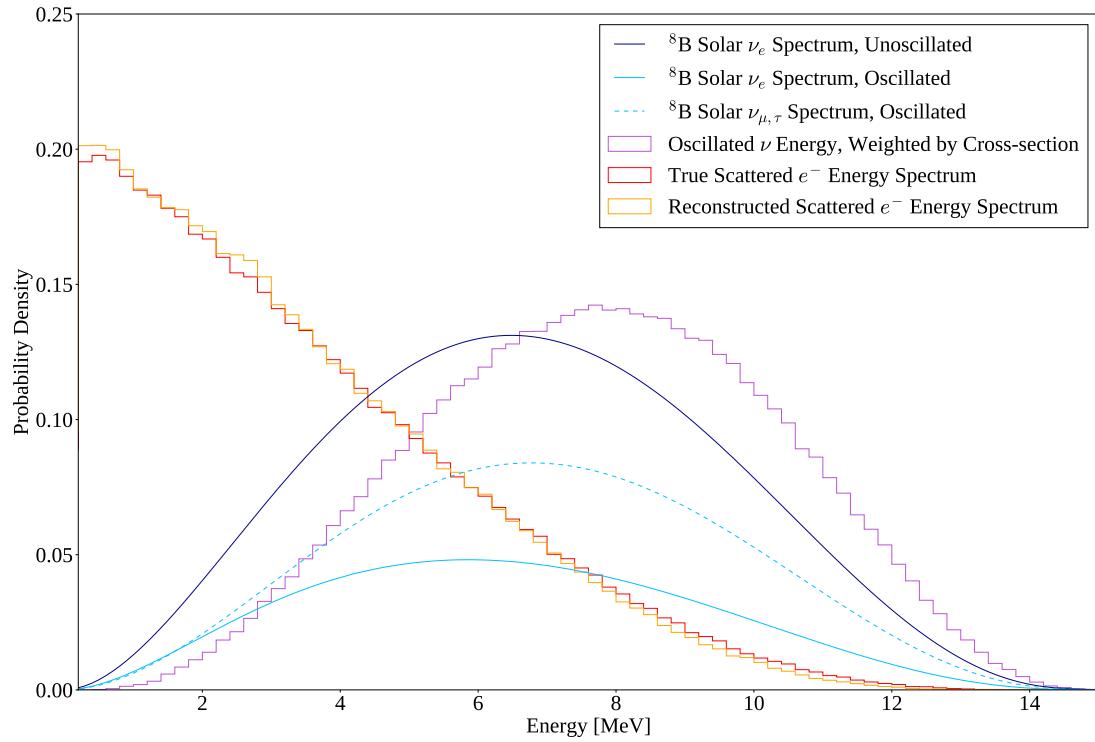


Fig. 6.3: The evolution of energy distributions related to ${}^8\text{B}$ solar neutrino detection. The unoscillated neutrino spectrum is taken from [30]; neutrino oscillations assume oscillation parameters from the current global fit results. The latter three distributions were obtained from MC production as described in Section 6.4.1.

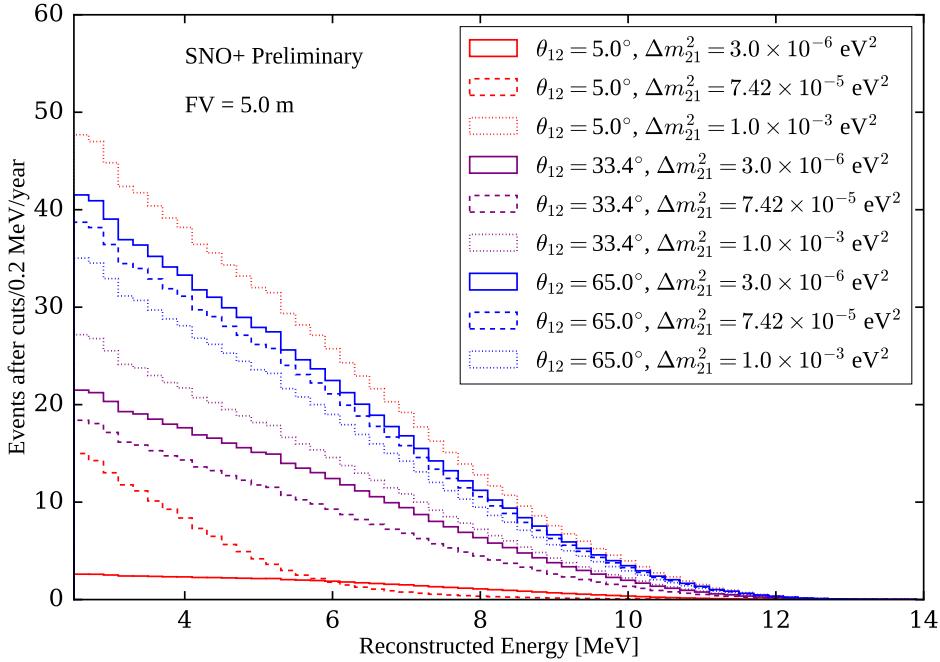


Fig. 6.4: Comparison of the reconstructed electron energy distribution expected for 9 different oscillation parameter combinations, after 1 year of data. The unoscillated rate was calculated in Appendix A, with the cuts described in Section 6.4.2 being applied.

as the oscillation parameters are modified. However, the correlation between the neutrino and electron energies is only weak, meaning the total rate of ${}^8\text{B}$ events observed changing is the more dominant effect.

The NuFit 5.1 global fit results, shown in Table 1.1, have the solar oscillation parameters measured as $\Delta m_{21}^2 = 7.42^{+0.21}_{-0.20} \times 10^{-5} \text{ eV}^2$ and $\theta_{12} = 33.44^{+0.77}_{-0.74} {}^\circ$ [60]. Fig. 6.5a shows the current allowed parameter regions associated with this fit. Thanks to the work of both solar neutrino and reactor anti-neutrino oscillation experiments, both of these oscillation parameters are known to a precision better than 3%. For comparison, the initial sensitivity study by J. Caravaca [157] gave an expected result for 1 year of data in the SNO+ scintillator phase shown in Fig. 6.5b. Note that this analysis assumed no backgrounds, systematics, or parameter constraints, and used a different statistical approach and set of analysis cuts to the ones described in the rest of this chapter. The study showed that,

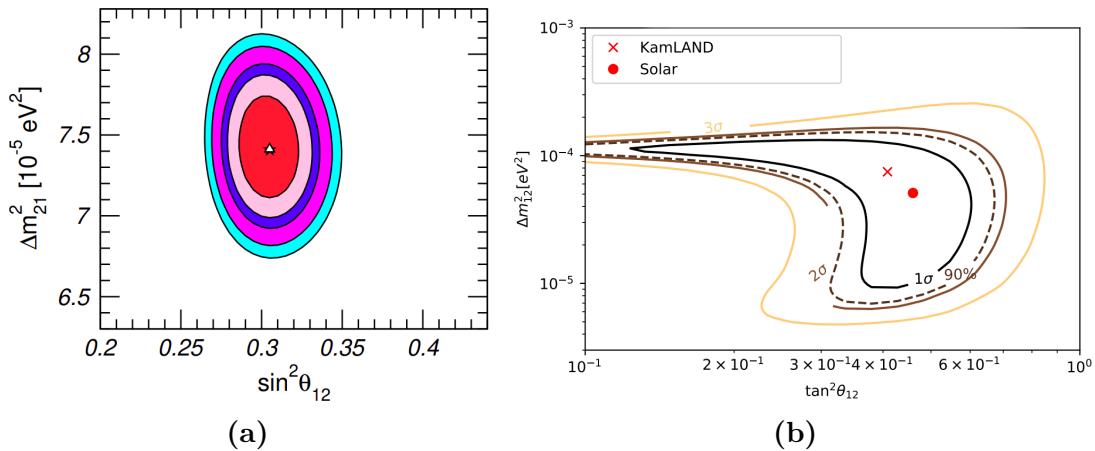


Fig. 6.5: (a): NuFit 5.1 global fit contours for the two solar neutrino oscillation parameters, after marginalisation over all other oscillation parameters [60]. Contours correspond to the allowed regions at 1σ , 90%, 2σ , 99%, and 3σ confidence levels. (b): Sensitivity contours produced by J. Caravaca [157] for 1 year of SNO+ data in a solar oscillation analysis sensitivity study. The analysis assumed no backgrounds, systematics, or uncertainty on the ${}^8\text{B}$ flux. The analysis took the NuFit 4.1 ‘solar’ combined fit oscillation parameters to be true, and compared them to the reactor anti-neutrino measurement made by KamLAND [46].

on its own, a solar oscillation analysis was unlikely to substantially improve the global fit measurement. However, it is possible that the solar analysis combined with a complementary reactor anti-neutrino analysis could do so. In any case, SNO+ has the capability to be the first experiment to measure θ_{12} and Δm_{21}^2 using solar neutrino and reactor anti-neutrino sources in the same detector. Moreover, it is valuable to perform a more detailed study of the solar analysis in SNO+, incorporating backgrounds, systematics, and constraints. This is the work of the rest of this chapter.

6.2 Background Processes

There are a number of background processes that the solar signal must compete against. Below a scattered electron reconstructed energy of ~ 2.5 MeV, it is known that low-energy backgrounds coming mainly from Uranium- and Thorium-chain isotopes coming from the natural radioactivity of the detector materials

completely dominate over the ${}^8\text{B}$ signal, and so for this analysis only processes that can generate reconstructed energies of at least $E_{\min} = 2.5 \text{ MeV}$ are considered. The following subsections explain each of the backgrounds present in the dataset above E_{\min} , as well as methods that have been used to mitigate them as much as possible.

Internal Uranium- and Thorium-Chain Backgrounds

Although every effort has been made to make the scintillator cocktail that fills SNO+ to be as radio-pure as possible, there inevitably remain trace amounts of the radioactive isotopes that derive from the decay chains of the ${}^{238}\text{U}$ and ${}^{232}\text{Th}$ isotopes. Fig. 6.6 shows these two decay chains. Only a fraction of the radioactive isotopes in these chains actually are capable of generating events in the detector with energies above E_{\min} : these have been highlighted in Fig. 6.6 in gold.

Of particular note are the decays of ${}^{212}\text{Bi}$ and ${}^{214}\text{Bi}$, both of which can either $\alpha - \beta$ decay via Tl, or $\beta - \alpha$ decay via Po. For the former, it is the subsequent β -decay of the Tl that can have a reconstructed energy above E_{\min} . For the latter, the Bi decay is the part of the pair of decays that can lie above E_{\min} . Although the α -decays here certainly have Q-values well above 2.5 MeV, the liquid scintillator quenches the observed energy to well below 2.5 MeV. The so-called “Bi–Po” decays are particularly special because the lifetimes of ${}^{212}\text{Po}$ and ${}^{214}\text{Po}$ are 300 ns and 164 μs , respectively, which are short enough to allow for highly-effective coincidence tagging of these events.

There are two main classes of Bi–Po event in the detector: “out-of-window” (OOW) events for which the Bi β -decay and Po α -decay events occur in separate triggered events, and “in-window” (IW) events whereby the Bi and Po occur within the same event. These lead to two distinct strategies for tagging these kinds of events. For out-of-window Bi–Pos, a delayed coincidence of two events is searched for in the same position: it is assumed that in the short time between the

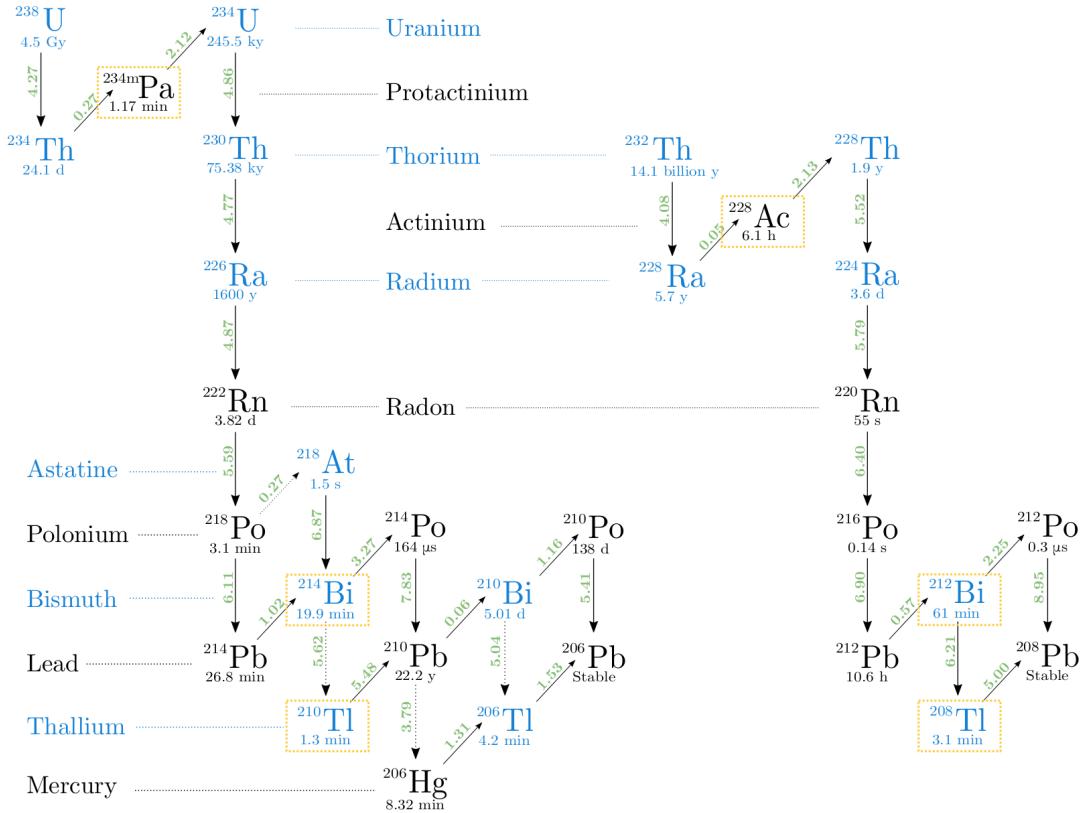


Fig. 6.6: The ^{238}U and ^{232}Th decay chains, taken from [137]. Isotopic half-lives are given below their symbol; the Q-values for each decay, in MeV, is given in green. Downward arrows indicate an α -decay; diagonal arrows indicate β -decay. Isotopes highlighted in gold are potential backgrounds for this solar analysis.

creation and decay of the daughter nucleus, there is negligible movement relative to the mother nucleus. Using the tagging algorithm suggested in [137, 159] as a starting point, the chosen procedure is described in Table 6.1. This choice of cuts was designed to be very broad, to ensure that the tagging was as *efficient* in rejecting Bi–Pos as possible, whilst negligibly impacting the solar signal. This is in contrast to the cuts chosen by Rafael Hunt-Stokes in [160], which uses a much tighter cut on position as well as cuts on the energies of the prompt and delayed events, to try and obtain a highly *pure* sample of Bi–Po tags.

The above delayed coincidence procedure cannot catch any of the IW Bi–Po events. For these, a different approach is used. Because two decays happened in the same event, two distinct peaks in the event’s time residual spectrum are expected

Property	Requirement
Prompt Event	Triggers detector, valid position reconstruction
Delayed Event	Triggers detector, valid position reconstruction, $\text{nHitsCleaned} \geq 100$
Δt	$< 4\text{ ms}$
ΔR	$< 2\text{ m}$

Table 6.1: Summary of cuts used for coincidence tagging.

to be seen. In order to look for this event topology, a likelihood-ratio classifier was run over events, first developed by Eric Marzec [161] and re-coordinated for the 2.2 g/L LAB-PPO scintillator optics by Ziping Ye [162]. This classifier calculates the likelihood ratio between the null hypothesis of a $0\nu\beta\beta$ event (a proxy in this analysis for single-decay events such as the ^8B signal) and the alternative hypothesis of an IW Bi–Po event. The results of this classifier are shown in Fig. 6.7. As can be seen, the more negative the value of the result, `alphabeta212`, the greater the evidence there is for rejecting the null hypothesis of a single-site event. Events with $\text{alphabeta212} < 0$, or the equivalent for $^{214}\text{Bi–Po}$ events, `alphabeta214`, were then rejected.

Combining both OOW and IW Bi–Po tagging, the impact on $^{212}\text{Bi–Po}$, $^{214}\text{Bi–Po}$, and $^8\text{B } \nu_e$ events can be seen in Fig. 6.8. Only events that pass all other cuts used in this analysis (other than the cuts for externals defined shortly) are considered: these will be explained in Section 6.4.2. Because of the different lifetimes of the decays relative to the length of the event trigger window, $^{214}\text{Bi–Po}$ 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 $^8\text{B } \nu_e$ signal events.

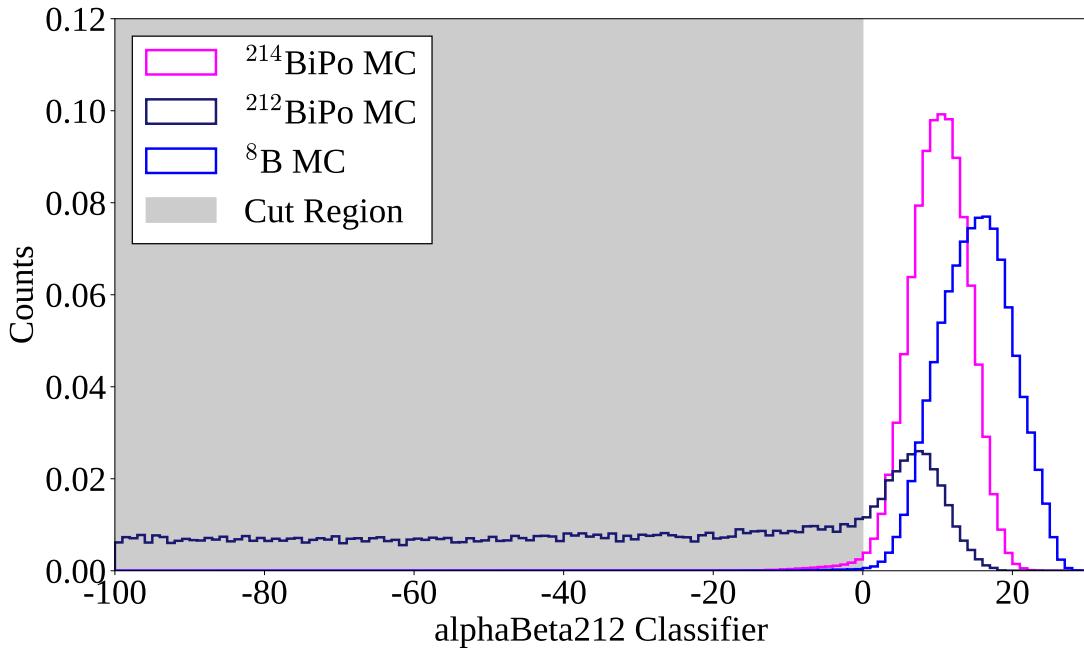


Fig. 6.7: The distributions of the IW ^{212}Bi –Po classifier for Bi–Po and ^8B events in simulation, within the analysis ROI. The cut of $\text{alphabeta212} < 0$ clearly removes a large fraction of ^{212}Bi –Po events, and 2% of ^{214}Bi –Po events, while keeping 99.89% of the ^8B signal. The relative normalisations before these cuts have been applied is arbitrary.

(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: $\alpha + ^{13}\text{C} \longrightarrow ^{16}\text{O} + \text{n}$. This is known as an (α, n) reaction.

The topology of this reaction in the detector is a delayed coincidence, as shown in Fig. 6.9. 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 Compton scatters to create the delayed signal in the detector.

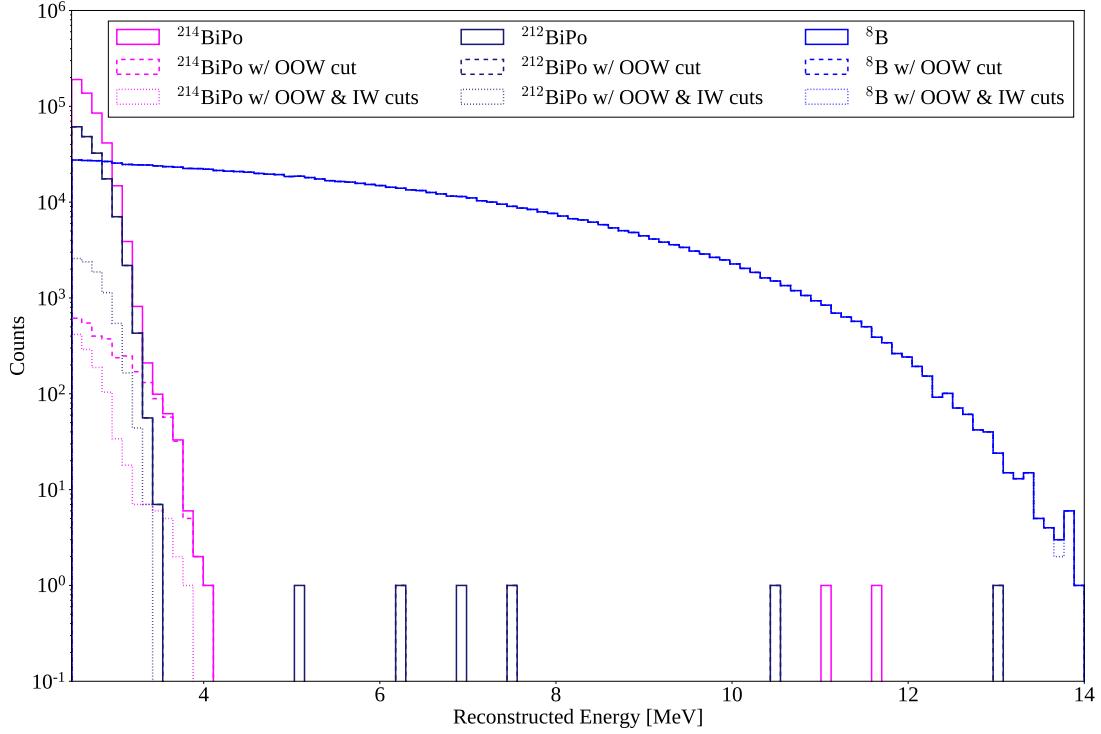


Fig. 6.8: $^{214}\text{Bi-Po}$, $^{212}\text{Bi-Po}$, and ^8B observed energy spectra in MC before and after OOW and IW cuts. The relative normalisations before these cuts have been applied is arbitrary.

As can be seen in Fig. 6.10, (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 $\sim 100 \text{ ns}$, the aforementioned out-of-window and in-window Bi–Po tagging algorithms also efficiently tag (α, n) events. Looking again at Fig. 6.10, simply by using the

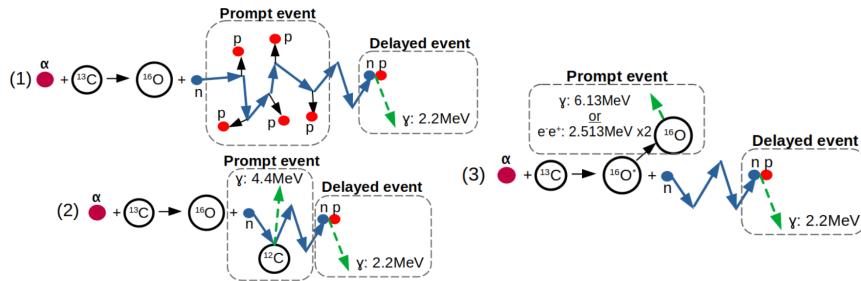


Fig. 6.9: Schematic of the three dominant modes of (α, n) interaction, taken from [90].

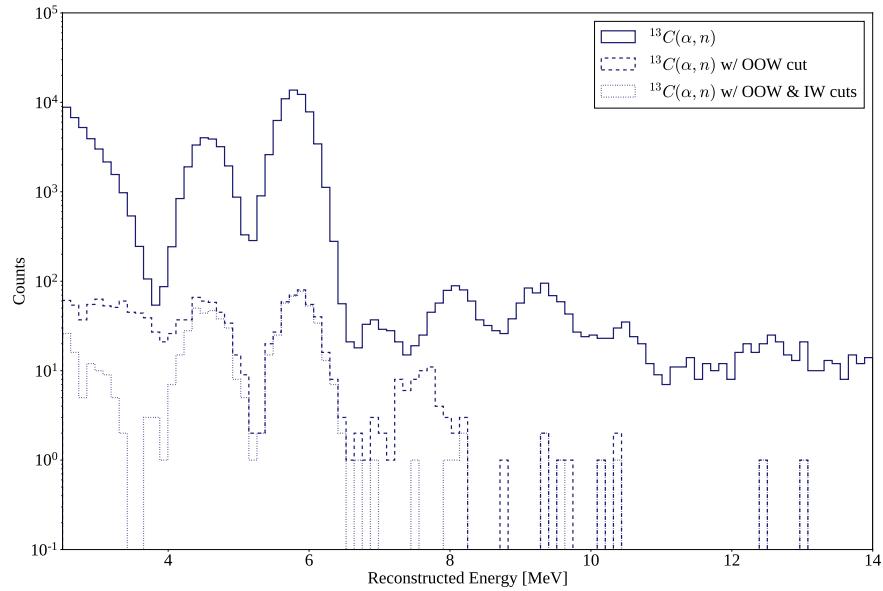


Fig. 6.10: (α, n) reconstructed energy spectrum in MC within the analysis ROI, before and after out-of-window cuts have been applied. The relative normalisation of the distribution before the additional cuts have been applied is arbitrary.

out-of-window and in-window Bi–Po taggers without any further modifications 99.37% of events in the ROI are cut.

External Backgrounds

To distinguish between the inherent backgrounds within the scintillator, and the backgrounds from materials at larger radii, the terminology “internal” and “external”, is used, respectively. External backgrounds can come from the acrylic, ropes, external water, and PMTs. These components have had their radiopurity measured throughout the detector’s lifetime, often back to the construction of the original SNO detector itself. The materials other than the liquid scintillator are known to have far higher background levels, especially in the important ^{238}U - and ^{232}Th -chain backgrounds [163].

Although there are numerous external backgrounds, with an accurate and precise position reconstruction algorithm they can be efficiently rejected. The simplest approach is with a “fiducial volume” (FV) cut: all events that reconstruct

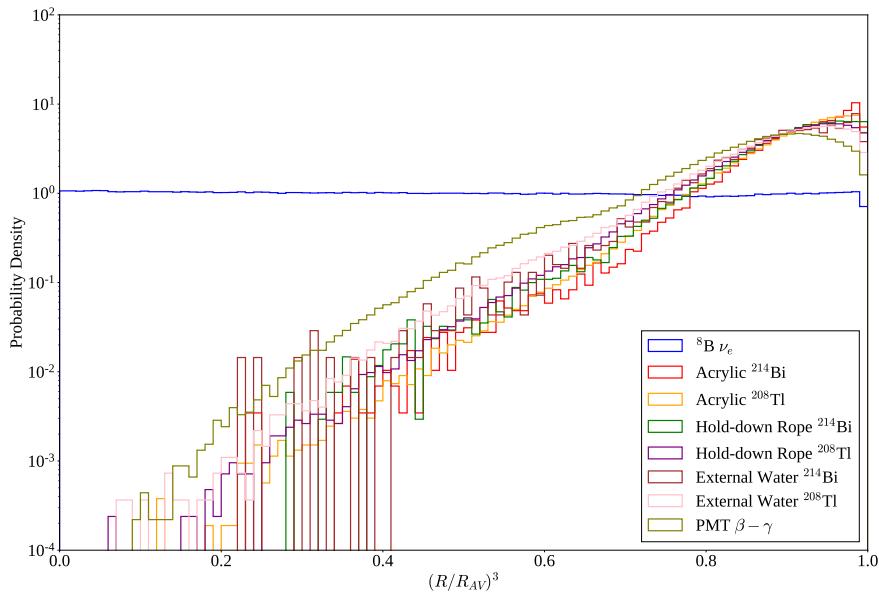


Fig. 6.11: Radial dependence of external backgrounds relevant in this analysis, as compared to the ${}^8\text{B}$ signal. Events for all processes shown here had to pass the first 6 cuts of Table 6.2.

beyond some radius are cut. The only external background events that will reach within the FV are those that have reconstructed very poorly, or have some long-distance radiation that manages to deposit radiation towards 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 with the Compton scattering length towards the centre of the detector. This strong radial-dependence can be seen in Fig. 6.11. One example of such a background are ‘PMT $\beta - \gamma$ ’ events, in which the ${}^{238}\text{U}$ - and ${}^{232}\text{Th}$ -chain radio-isotopes present in the PMT photocathodes generate γ s from the decays.

What this figure also demonstrates is that the 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 [137] allows for additional means of distinguishing external backgrounds from the solar signal. The underlying assumption in the reconstruction of SNO+ events is that photons were created from an electron which lost its kinetic energy in a near-point-like region of the scintillator. This is a reasonable assumption 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 an event with energy depositions in multiple different positions: a “multi-site” event. Because the `scintFitter` position reconstruction algorithm is not designed to deal with multi-site events in the scintillator, the time residual distribution will change. 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, external backgrounds should have the earliest PMT hits associated with PMTs closest to the main site of energy deposition. Because external events that reconstruct at small radii typically have a γ that travelled a long distance towards the centre of the detector, the direction from the centre of the detector to the reconstructed position should point in the direction of the deposition from the γ , the main decay site, and hence the PMTs which were hit first. A distribution of PMT hits for a given event as a function of their angular distribution relative to this direction 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`).

Fig. 6.12 shows the correlation between the two classifier results for both ^{208}Tl decays in the acrylic, and $^8\text{B} \nu_e$ events, using MC. The other external backgrounds

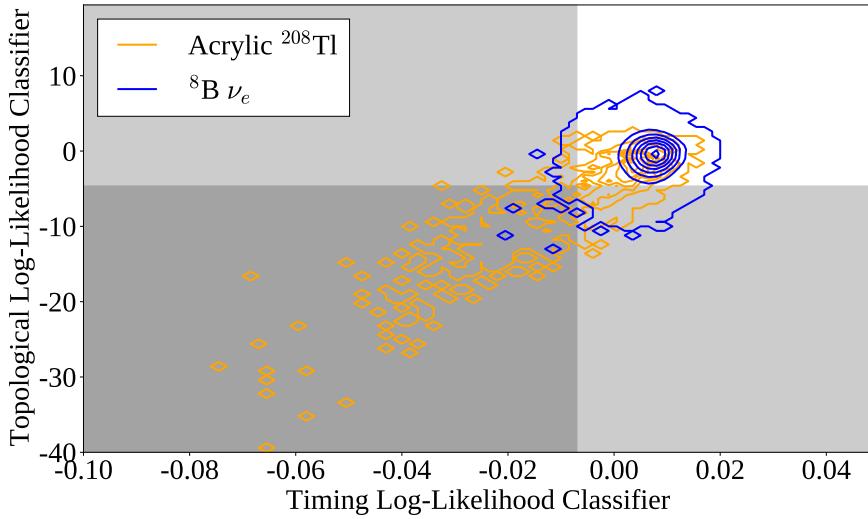


Fig. 6.12: Distributions of the timing and topological external classifiers, for both the ^8B signal and ^{208}Tl decays in the acrylic. Contours indicate lines of equal probability density. All events shown were in the analysis ROI, and below 5 MeV. Grey regions indicate the cuts used in this analysis.

in MC have a similar distribution in the external classifiers to that of the ^{208}Tl decays in the acrylic.

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

6.3 Statistical Approach

6.3.1 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 constitute the observed data with those same cuts

applied. Using these, the combined energy and radial distributions of the MC are fit to that of the data. Given a set of PDFs, a number of parameters can be modified to try and match the distribution of observables in data. 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.

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

$$\mathcal{L}(\mathcal{N}, \boldsymbol{\theta} | \mathbf{n}) = - \sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i + \sum_{j=1}^{N_{\text{bins}}} n_j \ln \left(\sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i P_{ij}(\boldsymbol{\theta}) \right), \quad (6.3)$$

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

6.3.2 The Bayesian Statistical Approach

There are two main schools of statistical inference, “Frequentist” and “Bayesian”. In the former, probabilities describe the fraction of times a situation can be found within the whole ensemble of possible worlds. For the latter, what matters instead is our degree of belief about matters in this current one. Our beliefs are updated as knowledge is acquired about the world through Bayes’ Theorem:

$$P(\boldsymbol{\mu}|\mathbf{x}) = \frac{\mathcal{L}(\boldsymbol{\mu}|\mathbf{x}) P(\boldsymbol{\mu})}{P(\mathbf{x})}. \quad (6.4)$$

Here, $\boldsymbol{\mu}$ is the set of parameters that model a given system, $P(\boldsymbol{\mu})$ is our *prior* (pre-existing) distribution for those model parameters, and \mathbf{x} is the data taken in an experiment. The updated, *posterior* distribution $P(\boldsymbol{\mu}|\mathbf{x})$ is then the prior

multiplied by the likelihood of parameters $\boldsymbol{\mu}$ given observations \mathbf{x} , $L(\boldsymbol{\mu}|\mathbf{x})$, and divided by the total probability $P(\mathbf{x})$ of observing \mathbf{x} under any circumstance. It is the Bayesian approach that has been used in this analysis.

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 a 1σ Credible Interval (CI): this is a set of values for a given parameter which has a total posterior probability of 68.3%, and contains the best-fit value. There are an infinite number of CIs that satisfy this property; for this analysis, the values have been chosen in decreasing order of marginalised posterior probability density.

6.3.3 Markov Chain Monte Carlo

The above discussion of Bayesian statistics assumes that one can accurately determine the posterior density distribution. Whilst the likelihood and prior distribution are straightforward 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 very numerically complex.

An alternative approach 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 distribution of possible positions that the system can be in converges [164]. MCMC uses such a Markov Chain which attempts to converge towards the posterior density distribution in particular. In MCMC, after choosing the initial position in the parameter space, successive states are chosen at random with a probability dependent only on the properties of the current position in parameter space and the proposed position. It is possible to show that, because of the convergence property of Markov chains, the set of steps made in the parameter space will have a distribution that converges to that of the posterior density distribution [164].

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

$$\begin{aligned} S(\boldsymbol{\mu}'|\boldsymbol{\mu}) &= \min\left(1, \frac{P(\boldsymbol{\mu}'|\mathbf{x})}{P(\boldsymbol{\mu}|\mathbf{x})}\right) = \min\left(1, \frac{L(\boldsymbol{\mu}'|\mathbf{x}) P(\boldsymbol{\mu}')}{L(\boldsymbol{\mu}|\mathbf{x}) P(\boldsymbol{\mu})}\right) \\ &= \min\left(1, \exp\left[\mathcal{L}(\boldsymbol{\mu}'|\mathbf{x}) - \mathcal{L}(\boldsymbol{\mu}|\mathbf{x}) + \ln \frac{P(\boldsymbol{\mu}')}{P(\boldsymbol{\mu})}\right]\right). \end{aligned} \quad (6.5)$$

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

The specific implementation of MCMC used for this analysis is that of **OXO**, a C++ analysis framework first developed by Jack Dunger [158] and extended for this analysis. **OXO** is able to run the Metropolis algorithm on multidimensional binned data, using the log-likelihood defined in Section 6.3.1. 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.

6.3.4 Choosing Priors

For this analysis, the suggestions made by Biller & Oser in [165] 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:

$$\ln P(\mu) = \mathcal{A} - \begin{cases} \frac{(\mu-\beta)^2}{2\sigma_+^2} & \text{if } \mu \geq \beta, \\ \frac{(\mu-\beta)^2}{2\sigma_-^2} & \text{if } \mu < \beta. \end{cases} \quad (6.6)$$

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. The oscillation parameters are restricted to be in the very broad range $3.0 \times 10^{-6} \text{ eV}^2 \leq \Delta m_{21}^2 \leq$

1.0×10^{-3} eV² and $5^\circ \leq \theta_{12} \leq 65^\circ$. The prior density beyond these limits is zero, meaning that any proposed step outside the allowed region can be immediately rejected.

6.3.5 Including Systematics in the Fit

One important implementation detail is how systematics are applied within the MCMC fitting process. Once systematics are added to the fit, at every step the binned PDFs for all the processes considered in the fit must get modified appropriately, which can become extremely computationally-intensive if not approached carefully. The strategy used in the `OXO` framework starts by thinking of the contents of a binned PDF as a vector of bin probabilities, $\mathbf{p} = (p_1, p_2, \dots, p_{N_{\text{bins}}})^T$. Then, a systematic acting on the PDF can be thought of as a linear transformation, and hence a matrix S acting on this vector: $\mathbf{p}' = S\mathbf{p}$. This matrix only needs to be calculated once for a given set of systematic parameter values, and can then use the same matrix on all the PDFs in the fit. Furthermore, when multiple systematics are applied, the matrix for each systematic can then be combined via matrix multiplication into one single “detector response” matrix. `OXO` uses the `Armadillo` [166, 167] linear algebra package for efficient matrix manipulation.

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

This is exemplified by the impact of an energy scale systematic on the energy distribution of PMT $\beta - \gamma$ events in the detector, shown in Fig. 6.13. Because

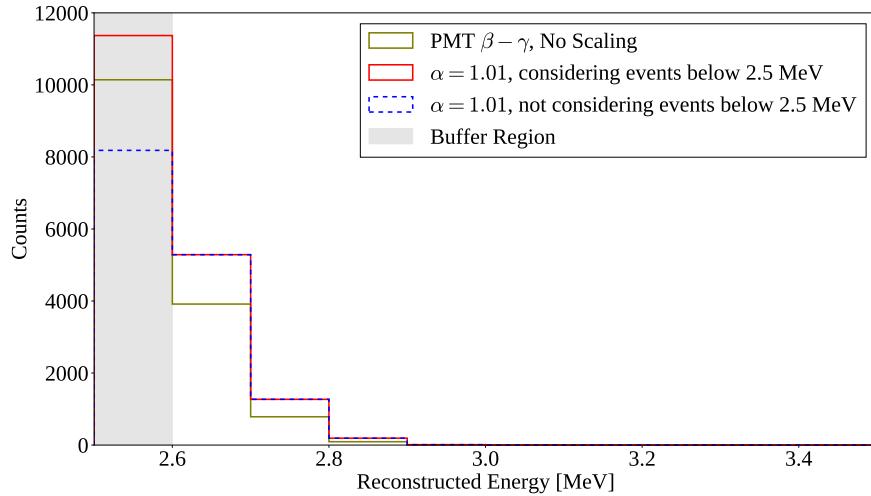


Fig. 6.13: Demonstration of how to handle an energy scaling systematic correctly. The distribution of PMT $\beta - \gamma$ events in the analysis ROI is shown versus energy, both before and after an energy scaling of $\alpha = 1.01$ has been applied. If the distribution below 2.5 MeV is not known, then the scaled distribution is biased systematically. This is solved through the use of a buffer region, as indicated.

the events seen for this process in the ROI are merely the high-energy 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 the edges of the ROI, which allow for tracking of events in and out of the ROI due to systematics, but are not considered when calculating the likelihood. This is also shown in Fig. 6.13. After the scaling systematic is applied, although incorrect bin values are found in the buffer region, this is fine because the likelihood is no longer being calculated with those bins. Note that because of this modification, the normalisation parameters 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.

6.3.6 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 $\varphi_{\text{sB}} = \Phi_{\text{sB}}^{\text{meas}} / \Phi_{\text{sB}}$, the unoscillated ${}^8\text{B}$ neutrino 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 to correctly apply oscillations, as the oscillation probability is a function of the neutrino’s true energy, not the measured energy. 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 $\boldsymbol{\theta} = (\Delta m_{21}^2, \theta_{12}, \varphi_{\text{sB}})$ the following is performed to oscillate the signal PDFs. Firstly, the normalisations are scaled by the factor φ_{sB} . 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. Within the structure of the `OXO` framework these bin-by-bin scaling are not 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 $\varphi_{\text{sB}} \cdot P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12})$ or $\varphi_{\text{sB}} \cdot (1 - P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12}))$ 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 SNO [168]. 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 described in Chapter 1 is assumed with neutrinos obeying the Normal Hierarchy, with the Sun following the **B16_GS98** metallicity model [31] and the **PREM** model being used for the Earth [169].

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 vertically above the detector, both vectors would be along the vertical direction $\hat{\mathbf{z}}$ 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 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

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

6.4 Analysis on Scintillator-Phase data

6.4.1 Dataset and Livetime

The data used in this analysis was scintillator phase data after the end of the PPO top-up campaign that completed in April 2022. An initial validation of the analysis tools was performed on data between 29th April and 10th May 2022 [170]. For the full analysis, data was chosen from runs taken after these dates. Not all data taken during this time was considered usable for this analysis, however. The Collaboration’s ‘Preliminary Scintillator Gold’ run selection list was used as the basis for this analysis. This run list requires:

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

The livetime of the data actually used in the analysis ends up being somewhat less than this raw value, because of the muon tagger and the **high-nhit** event tagger described in Section 6.4.2. This is because these tagging algorithms veto events for a time period following a tagged event. 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 6.3.6, 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. 6.14. Also shown in this plot is the zenith distribution expected if data was taken uniformly over the time period between 17th May–30th November 2022. While both distributions have peaks at $\cos \theta_z = -0.3$, 0.4, and 0.9, the data distribution

has a far more complex structure due to the times of day and year of the runs selected for this analysis.

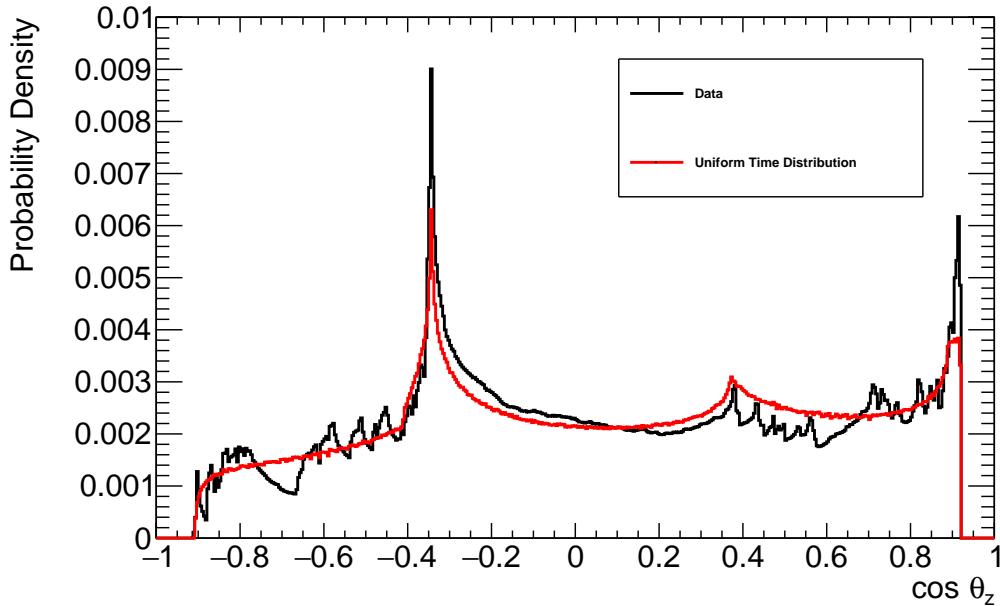


Fig. 6.14: Distribution of the cosine of the solar zenith angle, $\cos \theta_z$, for the events within this dataset. If an event has $\cos \theta_z > 0$, then at the time of that event the position of the Sun is above the detector's equator. Also shown is the zenith distribution expected if data was taken uniformly in the period 17th May–30th November 2022.

6.4.2 Event Selection

Once the runs of processed data and matching MC have been selected, the next step was to perform cuts on both in order to obtain the analysis ROI. The full list of cuts used are shown in Table 6.2, most of which have already been explained in Section 6.2. The energy and position cuts are contingent on a valid reconstruction having been performed on a given event; however, the `ScintFitter` is only run on events with a minimum number of hits, and even then the results may not be valid. Therefore, two important cuts in this analysis are to confirm that the `ScintFitter` was actually run, and that the results are valid.

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 Table 6.1
Remove BiPo In-window tags	Tagged if $\text{alphaBeta212} > 0$ and $\text{alphaBeta214} > 0$ Tagged if $E < 5.0 \text{ MeV}$,
Remove External tags	<code>ext0NuTimeT1208AVNaive > -0.007</code> , and <code>ext0NuAngleT1208AV > -4.7</code>

Table 6.2: Cuts used in this analysis.

Most of the other cuts used in this analysis that have not yet been discussed come from handling differences between data and MC. In MC, it is possible to simulate an event but not have it trigger the detector. In this case, the stored index of that event (`evIndex`) will be less than zero. These ‘events’ are never seen in data, but they do need to be removed from simulation. On the flip side, there are types of events that are observed in the detector but do not get simulated. Many of these events are caught during ‘data cleaning’, which looks for a wide variety of problems in the data when the data is being processed. Any subset of the checks made during data cleaning can be chosen to consider in a given 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 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 that occurs when the GTID rolls over.

- **Junk cut:** Tags events that have a channel with multiple recorded hits, which is not possible for a normal event. These are associated with so-called ‘orphaned’ hits that the event building process does not 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.
- **Muon cut:** Tags muon-like events.
- **Muon follower cut:** Tags all events coming up to 20 s after a tagged muon event.

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

The results of all these cuts can be seen for the data in Table 6.3. 654 out of the 18,112,495,770 events within the dataset pass all cuts. These same cuts were also run over the MC, with results summarised in Table 6.4. Important to note was the difference between the number of physics events simulated for a given

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 6.3: Impact of each cut on the quantity of events in data. The efficiency for a given cut is defined here as the fraction of events which have survived all previous cuts, which also survive that cut.

process and the number of triggered events. The latter ends up being larger than the former for all processes discussed here, because a given physics event in the detector will often generate a retrigger event following the primary triggered event. Also shown in this table are the combined cut efficiencies for each process. This efficiency is defined as the number of triggered events that pass all cuts, divided by the number of simulated physics events for that process. It is this value that can be used to convert between the number of physics events expected in a given time period, and the number of observed triggered events observed.

Variable bin widths in energy were chosen for this analysis: 0.1 MeV bins between 2.5 MeV–5.0 MeV, 0.25 MeV bins between 5.0 MeV–13.0 MeV, and a single energy bin between 13.0 MeV–14.0 MeV. Four equally-spaced bins were used in the 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 was used to allow for equal volume weighting for each radial bin. This binning for energy and r_3 was chosen as a balance between wanting to get as much information about the distributions of each process as possible, whilst ensuring that there were sufficient statistics for the MC PDFs in every bin.

MC process	# Physics Events Simulated	# Triggered Events Simulated	# Remaining after cuts	Overall Cut Efficiency (%)
$^8\text{B} \nu_e$, unoscillated	2,830,425	5,511,568	929,551	32.8
$^8\text{B} \nu_{\mu,\tau}$, unoscillated	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 6.4: 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: (unoscillated) signal, external backgrounds, and internal backgrounds. A final category corresponds to other processes that were considered for this analysis, but either all events are cut out, or Section 6.4.3 will show a negligible number of these events are expected in our dataset.

6.4.3 Expected Rates and their Constraints

Because this analysis is dependent on deriving the shape and normalisation of the ${}^8\text{B}$ spectrum from the data observed, it is important to know the number of events expected to be seen within the dataset for each process, both signal and background. This is done in two stages: firstly, the expected rates for each process before any cuts are applied is determined, along with any constraints. Then, by using the cut efficiencies calculated in the previous section, estimates for the expected number of events after cuts can be derived. When combined with a constraining uncertainty, this can be used as a prior within the MCMC fit. The resulting expected rates before and after cuts, along with any constraints being used, are shown in Table 6.5. Details of how the rates and constraints were calculated can be found in Appendix A.

6.4.4 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, Section 6.5 will show that the analysis is statistically-limited for this dataset, 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

MC process	Expected # Events Pre-Cuts	Overall Cut Efficiency (%)	Expected # Events Post-Cuts	Con- straint (%)
$^8\text{B} \nu_e$, unoscillated	572.58	32.8	187.94	+2.5% -1.7%
$^8\text{B} \nu_{\mu,\tau}$, unoscillated	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, along with any constraints. As in Table 6.4, 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.

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 $^{214}\text{BiPo}$ events tagged by R. Hunt-Stokes [171] in the dataset to the equivalent production MC, in the same 5 m FV. For a given α , the energies of the tagged ^{214}Bi in MC were scaled by α , and then compared to the equivalent (unscaled) distribution in data, via a log-likelihood ratio \mathcal{L}_{Bi} :

$$\mathcal{L}_{\text{Bi}} = \sum_{i=1}^{N_{\text{bins}}} n_i^{\text{data}} \ln \frac{n_i^{\text{MC}}}{N_{\text{MC}}}. \quad (6.7)$$

Here, n_i^{data} and n_i^{MC} are the number of events for data and (scaled) MC in energy bin i , with N_{bins} being the total number of bins and N_{MC} being the total number of tagged events in MC. To account for the edge effects discussed in Section 6.3.5, the first and last bins in energy were used as buffers and not considered in calculation of \mathcal{L}_{Bi} .

Values of α were scanned over, generating a log-likelihood distribution. In the usual way, a constraint on α was obtained by obtaining the energy scaling factor which maximises the log-likelihood, followed by looking for the α values with values of \mathcal{L}_{Bi} less than the maximum by 1/2. The constraint was found to be $\alpha = 0.9969 \pm 0.0022$, with the data and scaled MC at the best fit value of α shown in Fig. 6.15.

To handle energy scaling as a systematic in this analysis, the energy scale is floated within the MCMC fit, with the constraint given above. The first and last energy bins of the MCMC fit, 2.5–2.6 MeV and 13–14 MeV, were used as buffer bins to let this floating occur. The energy scale was then allowed to float in the range $\frac{13}{14} \leq \alpha \leq \frac{2.6}{2.5}$, which ensures that scaling could never go beyond the

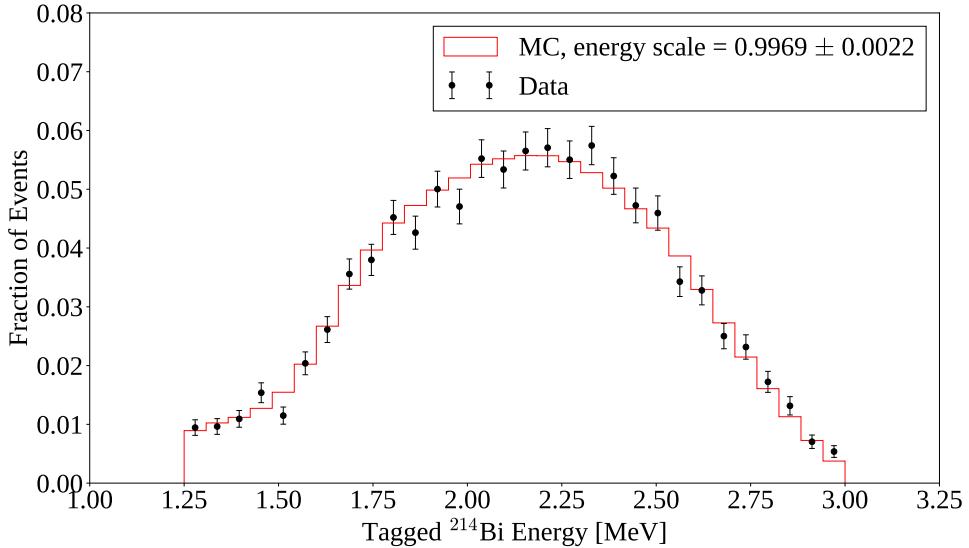


Fig. 6.15: Comparison of tagged ^{214}Bi reconstructed energy distributions between data and MC, after MC has been scaled by the best-fit factor $\alpha = 0.9969$.

maximum or minimum allowed by the size of the buffer bins. This allowed region is much larger than the constraint found above.

In addition to a global energy scaling, it is possible that the resolution of the energy reconstruction could be systematically off. Because of the effects discussed in Section 6.1, changes to the oscillation parameters only impact the observed signal distribution over a broad range of energies. This means that mismodelling the energy resolution of the detector is only liable to have an impact on the measurement if that mismodelling is substantial. However, by looking again at Fig. 6.15 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.

Two other possible energy systematics could also exist. There could be an offset in the reconstructed energy, and even some degree of non-linearity. This would lead to a mismodelling of the form $E_{\text{reco}} \rightarrow E_0 + \alpha E_{\text{reco}} + \beta E_{\text{reco}}^2$, where E_0 and β are the energy offset and non-linearity systematic parameters. Calibrating all of these systematics together would require at least three calibration sources.

This has not been done in this thesis, and so for now these two systematics have been ignored.

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 all normalisations are being floated within the MCMC fit, and the normalisation constraints of almost all PDFs are relatively large, the impact of these systematics is expected 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 assumptions made to model them (such as the use of simulation shells) are not fully accurate. These shape systematics can also arise from the lack of statistics in the PDFs used to model them: as an example, Table 6.4 shows that only 62 external water ^{214}Bi events are used in the creation of the associated PDF. Handling this type of modelling uncertainties for the external backgrounds is not considered in this thesis.

6.4.5 Results

Fit Validation

In theory, the convergence of the sampled distribution within an MCMC fit to the posterior density distribution is guaranteed to eventually occur, regardless of the specific shape of the multivariate Gaussian proposal distribution. However, the speed of this convergence is highly dependent on the widths of the proposal

distribution. This is very important as in practice the MCMC fit can only be run for a finite amount of time. For a given parameter within the fit, if the associated width (also known as the step size) in the proposal distribution is too large, then the vast majority of proposals will be of positions in the parameter space where the log-likelihood is less than before, and so these proposals will be consistently rejected. Alternatively, if the width is chosen too small, the acceptance rate of the chain will be very high, but too much time will be spent for the chain to simply explore the space.

By choosing sensible width parameters for the proposal distribution, as well as running the chain for as long as possible, one can attempt to maximise the “effective sample size” of the MCMC fit. The effective sample size is the theoretical number of independent samples that would need to be drawn from the posterior density distribution to have equivalent statistics to that of the MCMC sampling process. It has been shown that [172], for a wide variety of situations, an appropriate size for the width of a given parameter is the standard deviation of the true posterior density distribution. Because knowing this 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 continued until two conditions were met simultaneously:

- The sampled distribution marginalised over every fit parameter appeared smooth;
- The marginalised distributions also appeared to fully explore their parameter spaces many times.

Such a result indicated 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. 6.16 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. These initial samples are not representative of the actual posterior density distribution, and are called the “burn-in” period. Fortunately, enough tuning of the width parameters was performed that there was minimal burn-in observed in any of the chains: this can be observed in Fig. 6.16. To be very safe, for all chains the first 100,000 steps were declared as the burn-in period. The overall set of samples used to approximate the posterior density distribution corresponds to collection of all sampled points on all the chains, after burn-in.

In addition to the above, the MCMC fitting code was also tested on a series of ‘fake’ datasets, including those described in Section 6.5. These were able to validate that if a ‘fake’ dataset was generated with a given set of parameters, the MCMC fitter would correctly generate posterior densities consistent with those input parameters.

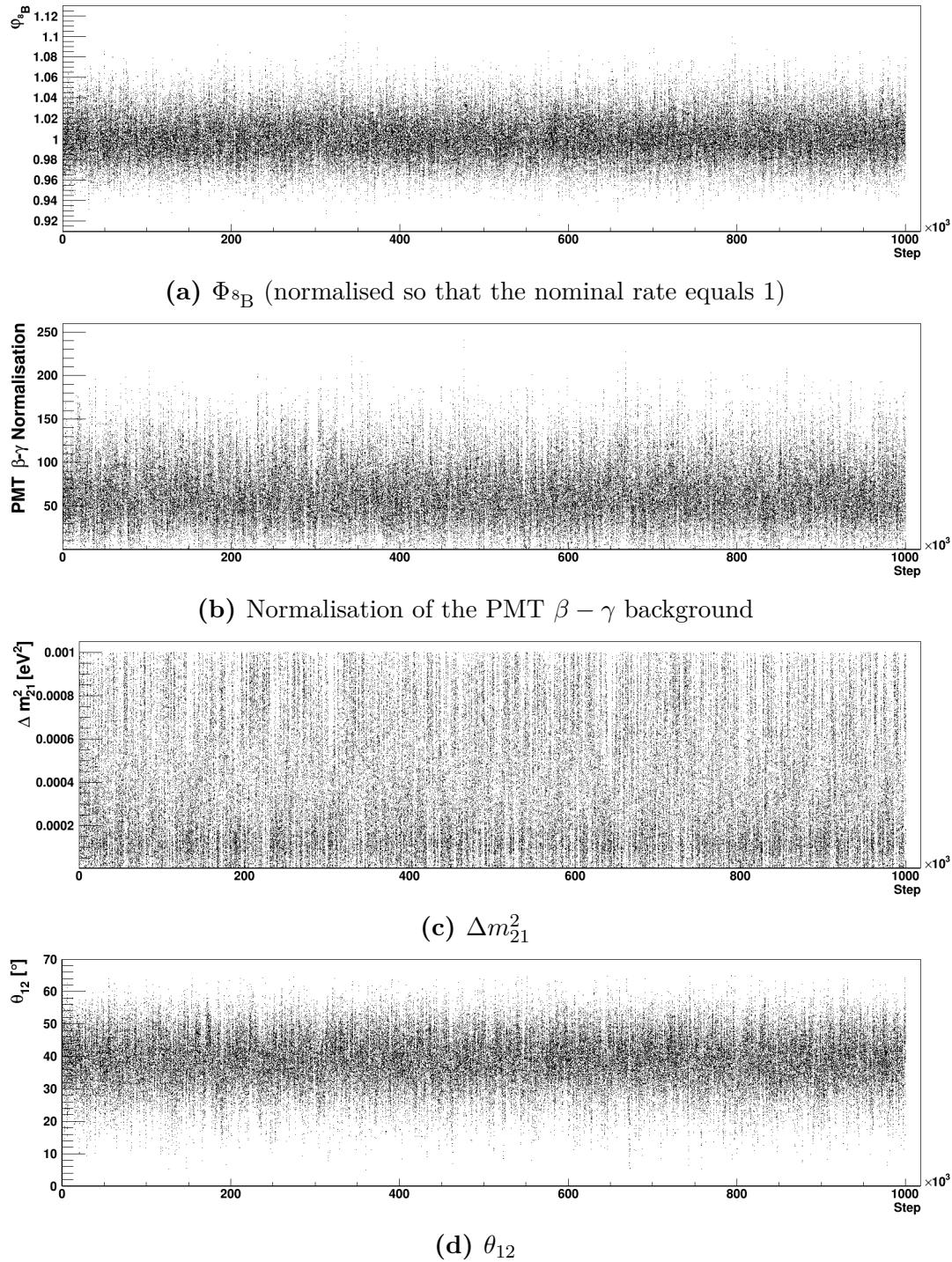


Fig. 6.16: Examples of the sampled values for parameters within a given MCMC chain as a function of the step number, after tuning the proposal distribution width parameters.

Oscillation Fit Results

Now that there is confidence in the convergence of the MCMC fit to the posterior density distribution, the results of this analysis can be obtained. Fig. 6.17 shows the posterior density distribution, marginalised onto each parameter of the fit other than the two oscillation parameters. Each marginalised distribution appears smooth, another qualitative indication of a sufficient effective sample size in the MCMC fit. For each 1D distribution, the bin centre associated with the HPD has been indicated, and the vertical band indicates the Bayesian 1σ CI for that parameter. Also shown in these plots is the prior distribution for each of the parameters, for comparison.

For a number of the parameters, the posterior distributions are more strongly peaked than their associated priors. This indicates that there was enough information in the dataset to further constrain those parameters. One such parameter is the rate of the external water ^{214}Bi . The resulting measurement of this is $34.1_{-18.9}^{+18.5}$ events in the dataset. This is 10 times larger than the expected rate shown in Table 6.5, providing some 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. These slices have been numbered radially, with the innermost radial slice being slice 1. 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 information 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.

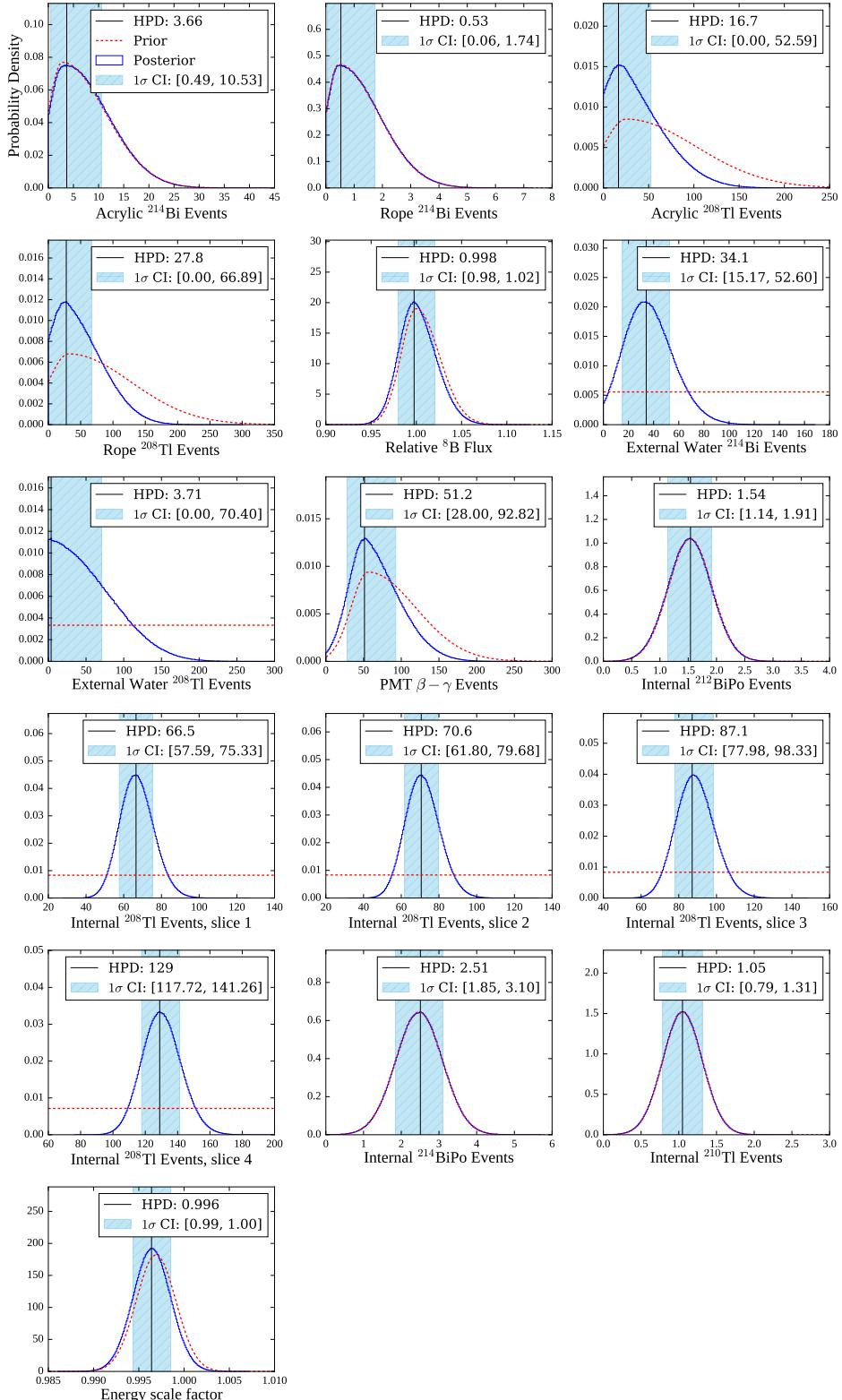


Fig. 6.17: Marginalised 1D posterior density distributions for all non-oscillation parameters, shown in blue. The HPD and 1 σ Bayesian CI are shown for each parameter. Also shown in dashed red are the associated prior distributions.

Fig. 6.18 shows the correlation coefficients associated with the 2D marginalised distributions, for each pair of parameters. From this, clear anti-correlation between a number of the external backgrounds can be seen. This seems intuitive: the PDFs for many of these external backgrounds are similar, and so the fit struggles to distinguish between them.

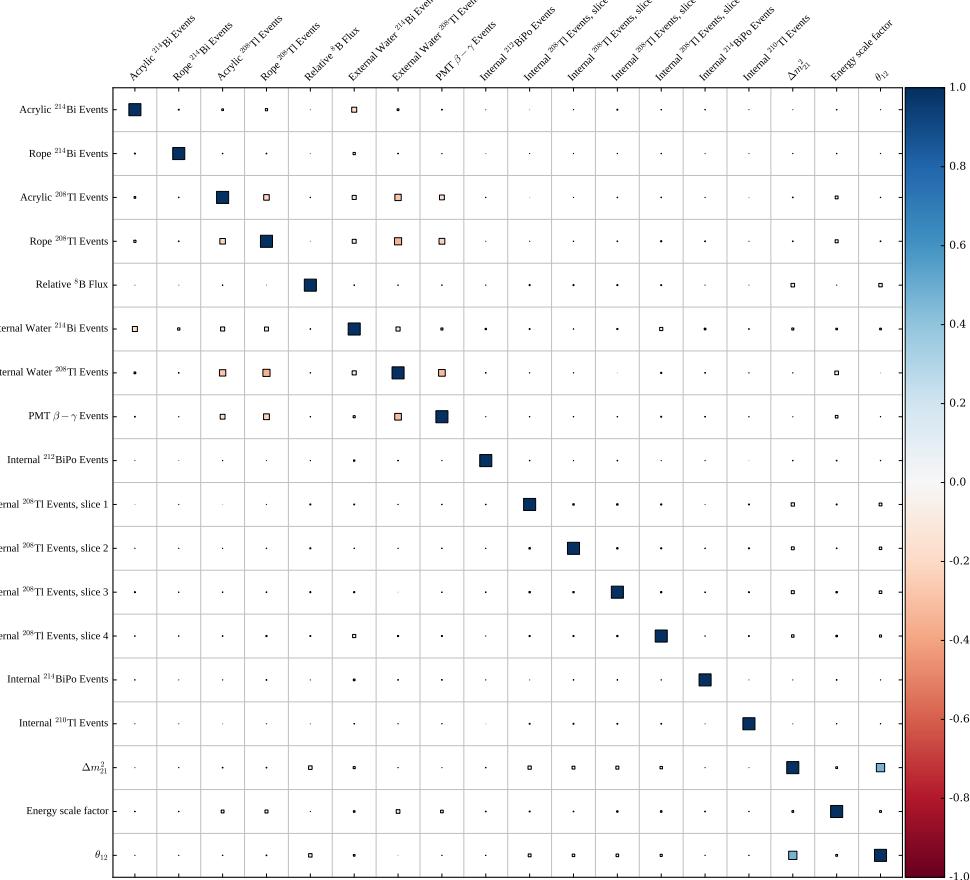


Fig. 6.18: Correlation matrix between all parameters in the MCMC fit.

The only other pair of parameters which have a substantial correlation between one another are the two oscillation parameters. Fig. 6.19 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 $\theta_{12} = 34.1^\circ$, $\Delta m_{21}^2 = 1.80 \times 10^{-5} \text{ eV}^2$.

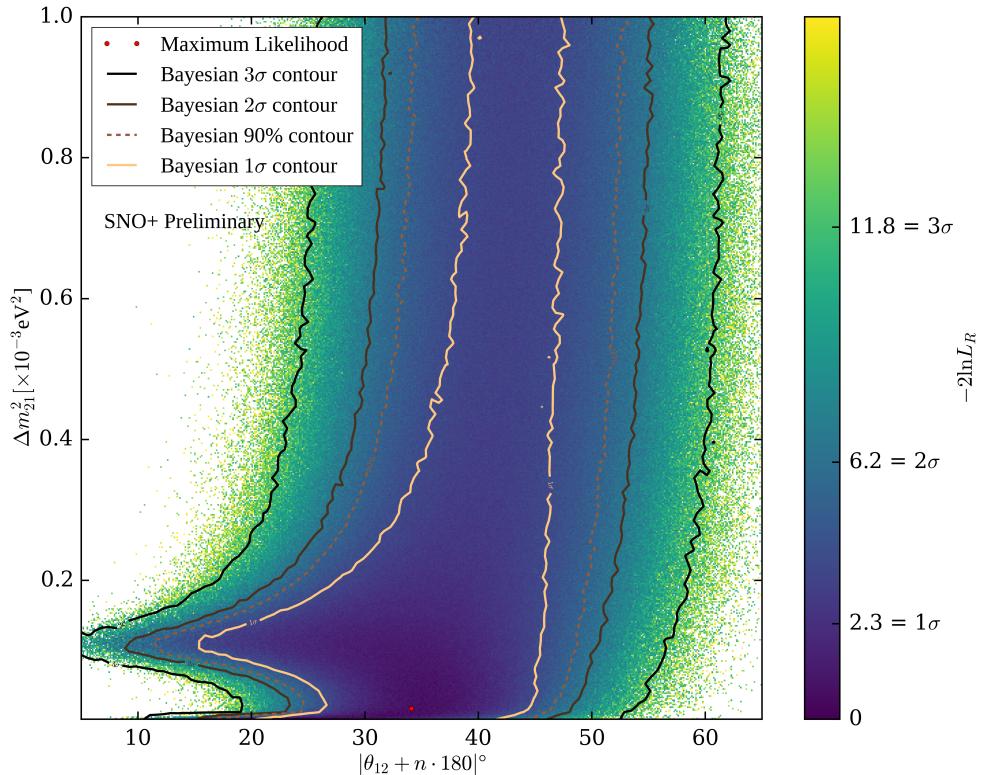


Fig. 6.19: 2D posterior density distribution marginalised onto Δm_{21}^2 and θ_{12} . The colour axis shows $-2 \ln L_R$, proportional to the natural logarithm of the posterior density, with labels corresponding to the values associated with the $N\sigma$ Frequentist CIs. Also shown, in shades of brown, are the Bayesian CIs.

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 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. 6.20. From this, we can obtain a measurement of:

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

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 [60]:

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

As can be seen, the measurement made in this thesis is consistent with the global fit value, although with a substantially larger uncertainty, due to the limited statistics of the dataset.

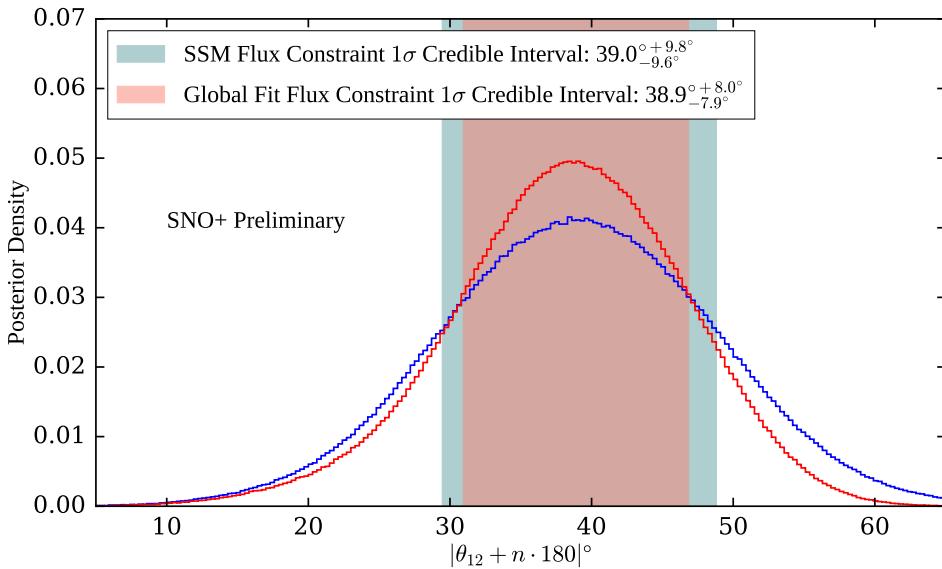


Fig. 6.20: Comparison of 1D posterior density distributions marginalised onto θ_{12} , using both forms of ${}^8\text{B}$ flux constraint. The 1σ Bayesian CI are shown for both.

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. 6.19, the preferred values of θ_{12} for large Δm_{21}^2 are somewhat greater on average than for Δm_{21}^2 values close

to the global fit value of 7.4×10^{-5} eV 2 . As a result, the final value quoted for θ_{12} is dependent on the choice of the prior distribution for Δm_{21}^2 : this is an inevitable result of using a Bayesian framework for this analysis. The impact of choosing a less conservative prior for Δm_{21}^2 can be seen by ignoring all sampled points with $\Delta m_{21}^2 \geq 5 \times 10^{-4}$ eV 2 , as an example. The resulting measurement of θ_{12} becomes:

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

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

Also shown in Fig. 6.19 is the posterior density for θ_{12} if the looser SSM flux constraint described in Appendix A is used. Under these conditions, the measurement becomes:

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

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

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

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.

Fit Parameter	1D HPD value	Maximum likelihood value
Relative ^8B 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}Bi Events	3.66	8.49
Ropes ^{214}Bi Events	0.53	0.52
AV ^{208}Tl Events	16.71	19.08
Ropes ^{208}Tl Events	27.81	31.20
External Water ^{214}Bi Events	34.09	45.93
External Water ^{208}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}Tl Events, Slice 1	66.52	68.78
Internal ^{208}Tl Events, Slice 2	70.60	66.62
Internal ^{208}Tl Events, Slice 3	87.09	87.33
Internal ^{208}Tl Events, Slice 4	128.92	125.25
Internal $^{214}\text{BiPo}$ Events	2.51	2.56
Internal ^{210}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.

Using the maximal likelihood fit parameters, the comparison of data to MC is shown in Fig. 6.21. The ^8B 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}Tl shows a clear peak well above the ^8B signal. Qualitatively, the MC appears to give a good fit to the data.

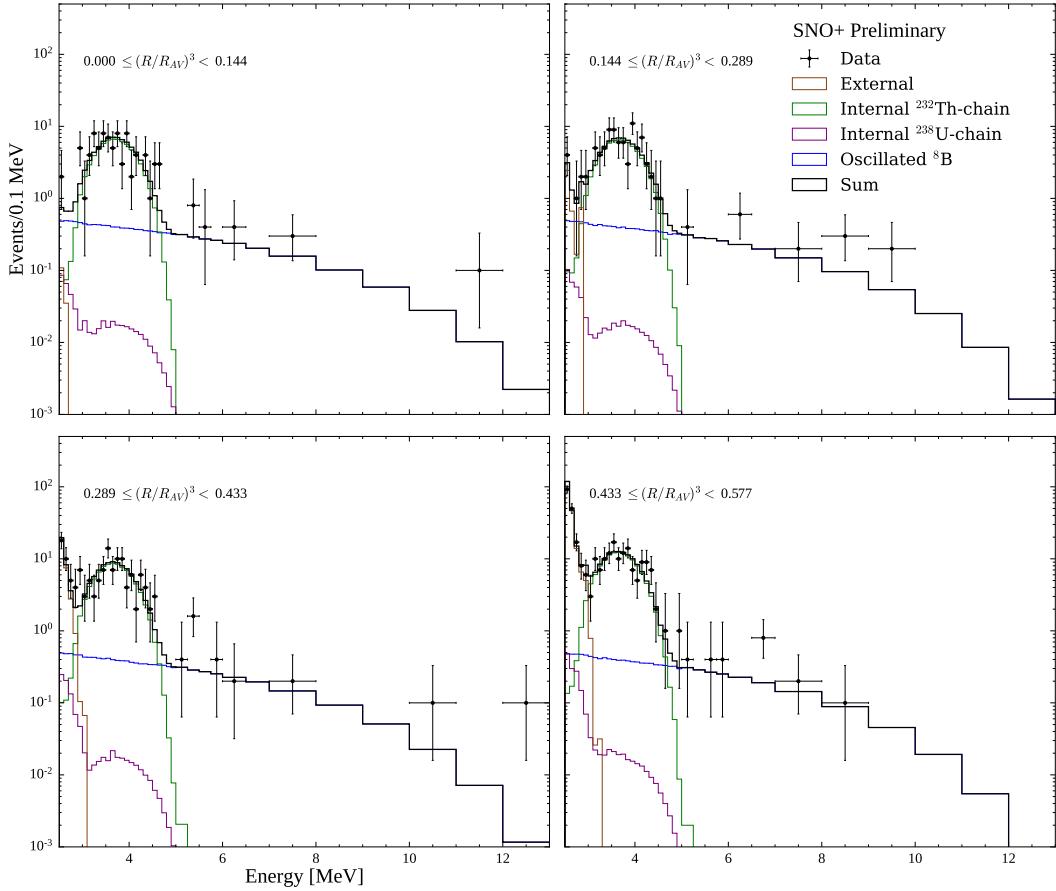


Fig. 6.21: Comparison of data to MC for each radial slice, using the best-fit parameter values derived from the MCMC results.

6.5 Sensitivity Projections

Given the limited statistics used in the dataset analysed in Section 6.4, the question of how much better the solar oscillation parameters could be measured with more data in SNO+ 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 ${}^8\text{B}$ 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 [173].

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 ${}^8\text{B}$ 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 ${}^{238}\text{U}$ - and ${}^{232}\text{Th}$ -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 ${}^{214}\text{Bi-Po}$, 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 ${}^{208}\text{Tl}$, 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 the fractional change in rate between the two time periods, 0.89/0.94.

External backgrounds were in general set to the rates expected from T. Zummbo, as described in Section A.4. The exceptions were the backgrounds from the external water: for the ${}^{208}\text{Tl}$ component, the target rate of this background as described in [163] was used. For the ${}^{214}\text{Bi}$ 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. 6.22 and 6.23. Some chains took much longer to converge than in the main analysis; as a result, a burn-in of 250,000 steps was found to be sufficient. As can be seen, by ~ 2 years of livetime, there is sufficient evidence to confidently reject large Δm_{21}^2 values above 2×10^{-4} eV². 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} made by SNO+, as well as the beginnings 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 bis-MSB and Te into the detector that is likely to occur in the coming years. For bis-MSB, 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 [84]. Once the Te is also loaded, the observed light yield will likely decrease back down to levels similar to that of the current LAB-PPO. One of the biggest expected impact of the Te phase on this analysis will instead

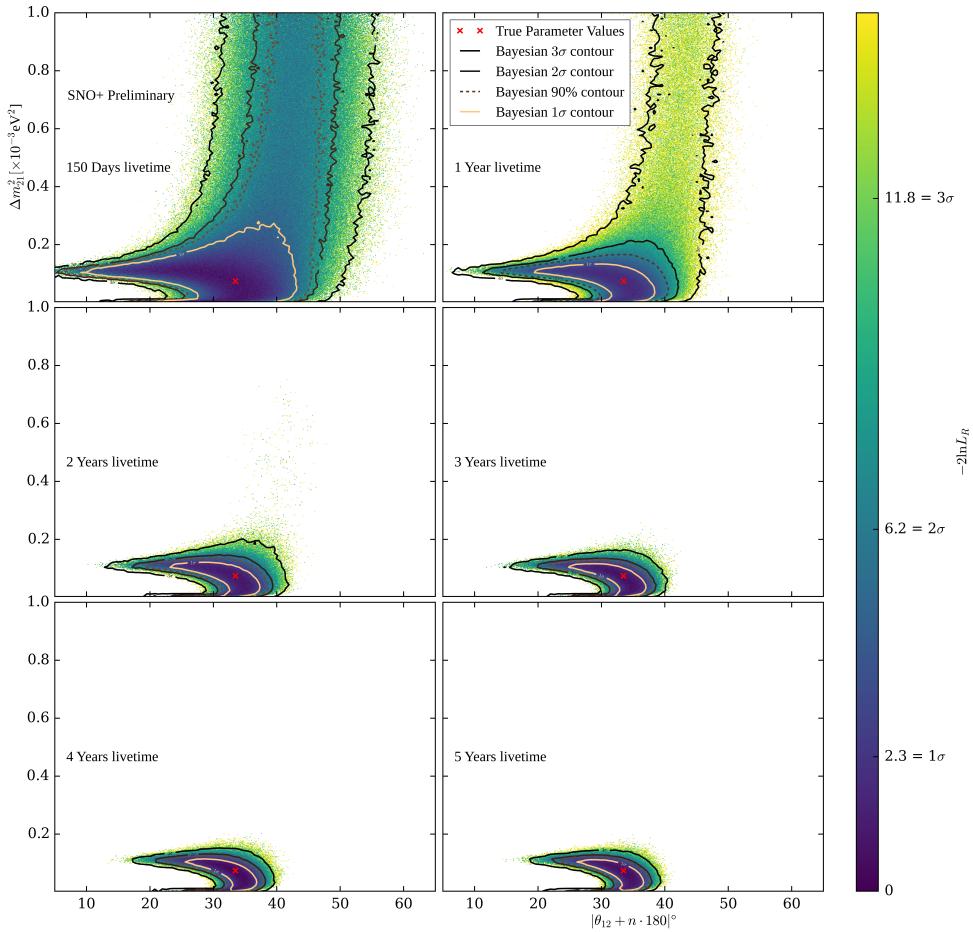


Fig. 6.22: Posterior density distributions marginalised onto the two solar oscillation parameters, for each livetime scenario in the pure scintillator phase under the current backgrounds.

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 [163].

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 [137]. 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

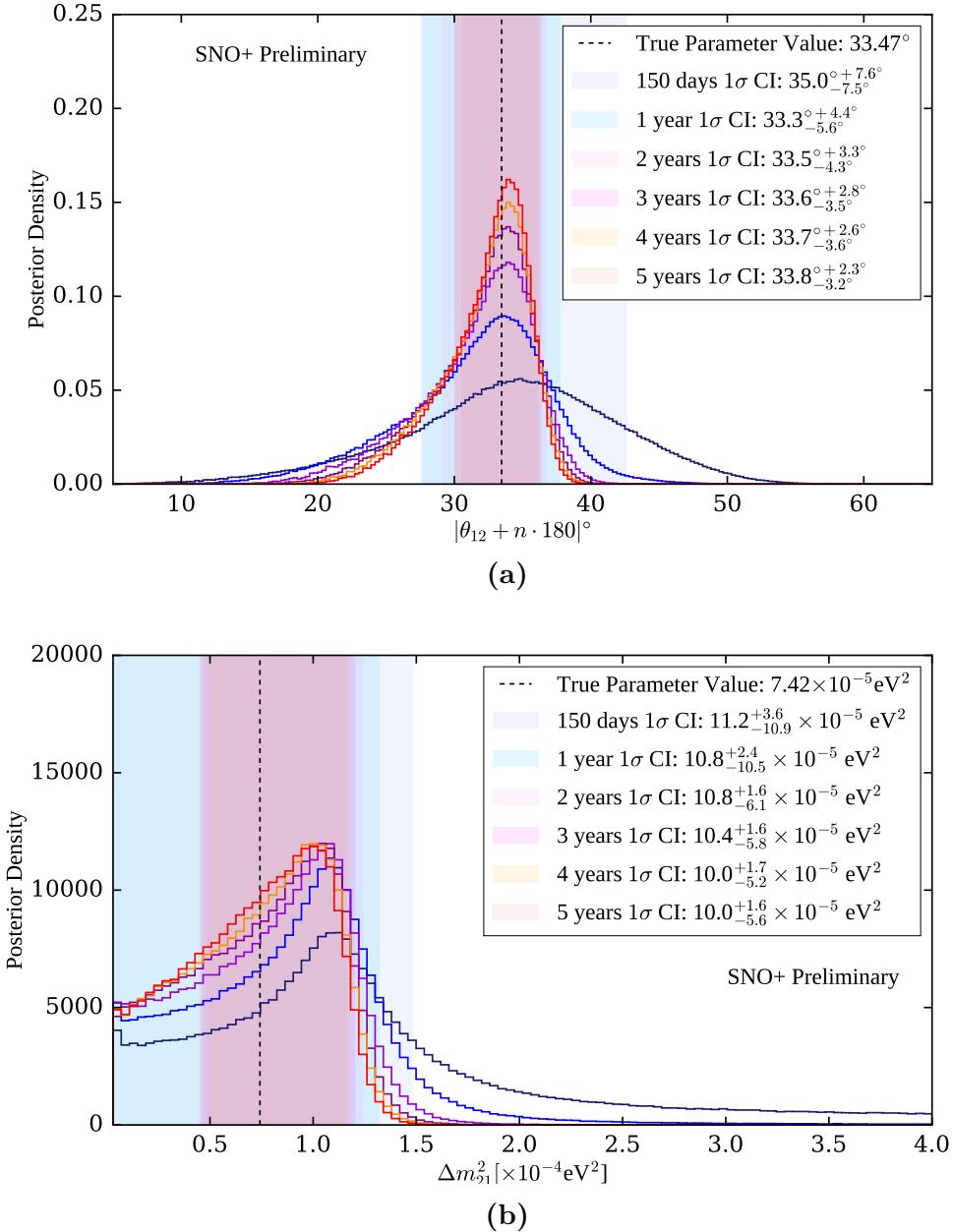


Fig. 6.23: Posterior density distributions marginalised onto each of θ_{12} and Δm^2_{21} , for each livetime scenario in the pure scintillator phase under the current backgrounds.

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 6.2 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 [137]. 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 lower level measured by T. Zummo in the water phase, and the internal ^{208}Tl were reduced by a factor of 10. This latter scenario reflects the possibility of better control of the radioactivity in the external water, as well as a way of using new analysis methods to cut out 90% of the internal ^{208}Tl events.

The results of running these two scenarios can be seen in Figs. 6.24 & 6.25. Unsurprisingly, the 1σ CI for θ_{12} is wider when backgrounds are greater, and thinner in the lower background scenario. However, the magnitudes of the change in width are somewhat different. For the high backgrounds case, the width only increases by 2%. This implies that the higher internal backgrounds expected during the Te phase should not, on their own, substantially hamper the sensitivity of the measurement of θ_{12} . With lower external water and internal ^{208}Tl backgrounds, it appears that sensitivity improvements $\mathcal{O}(10\%)$ are achievable. This is because, over 1 year of livetime, the statistical uncertainties present in the signal process still dominate.

6.6 Summary and Suggestions for Further Work

In this chapter, a full analysis for measuring the solar neutrino oscillation parameters via ^8B neutrinos was built and demonstrated on SNO+ scintillator-phase

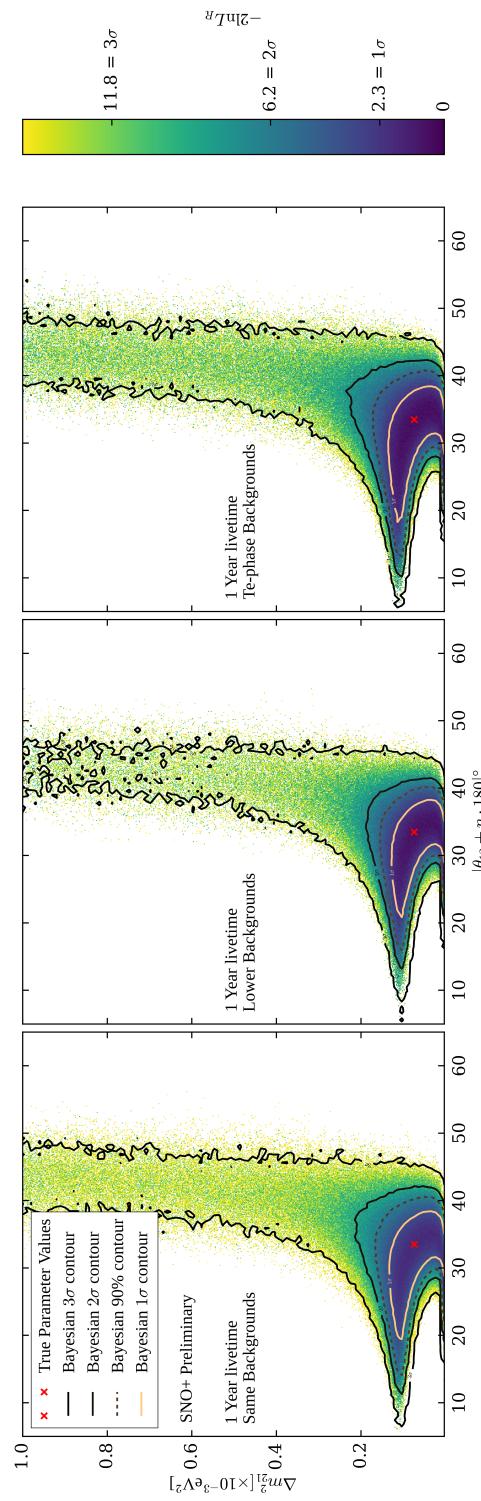


Fig. 6.24: Posterior densities marginalised onto the two oscillation parameters, for projections over 1 year of livetime, with different background expectations.

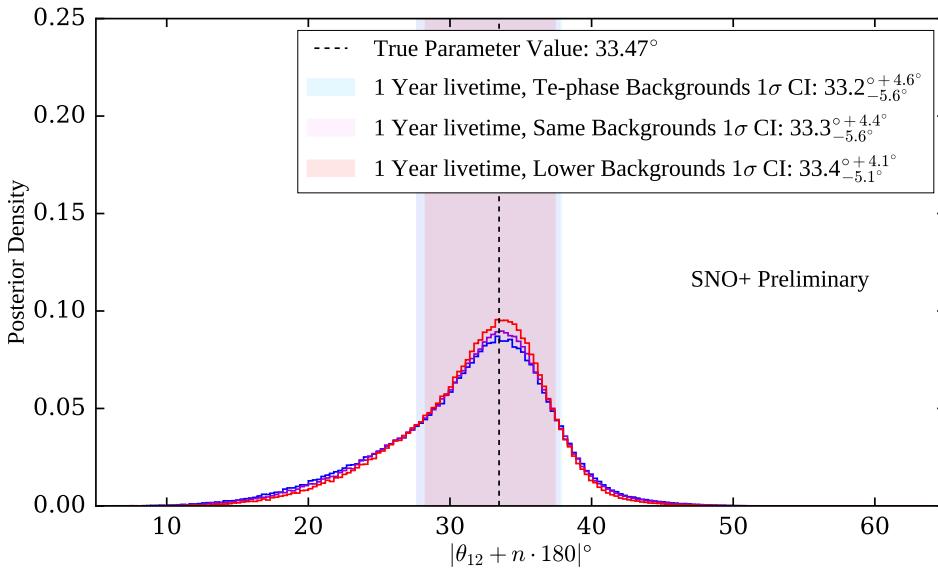


Fig. 6.25: Posterior densities marginalised onto θ_{12} , for projections over 1 year of livetime, with different background expectations.

data. Using 80.6 days of livetime, assuming the global fit constraint on the ${}^8\text{B}$ flux of $\Phi_{8\text{B}} = (5.16^{+2.5\%}_{-1.7\%}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ [174], the oscillation parameter θ_{12} was measured to be $|\theta_{12} + n \cdot 180|^\circ = 38.9^{+8.0\circ}_{-7.9\circ}$.

Furthermore, assuming an identical analysis approach and similar detector conditions to those seen in the above dataset, the precision of this measurement is liable to improve by a factor of 2 in under 2 years of total livetime. This is partly because there will be enough statistics to make a measurement of Δm_{21}^2 .

Of course, major changes to the detector's scintillator cocktail are expected in that timeline. The negative impacts expected from the higher internal backgrounds of the Te-loaded phase should have a small impact on the sensitivity to the oscillation parameters over a livetime of 1 year. The sensitivity obtained in this scenario is expected to be somewhat conservative, as the scintillation emission time profile is expected to become substantially shorter [93], leading to substantial improvements in the position resolution of the detector as well as the discrimination power of various time-based classifiers.

If the external water backgrounds can be returned to the levels observed during the water phase, and the event selection of internal ^{208}Tl events improves substantially, there is some scope for improvement to the measurement of θ_{12} when considering 1 year of livetime. There are three main possibilities for removal of internal ^{208}Tl events. Firstly, these β -decays come as the delayed partner to the α -decay branch of ^{214}Bi nuclei. Therefore, a coincidence tagging approach to remove these events could be possible. Furthermore, alongside the β -decay of ^{208}Tl are typically γ s, which should modify the observed time residual spectrum of these events compared to a ^8B signal event. As a result, creating a custom multi-site classifier to exclude internal ^{208}Tl events from single-site events could help further. Finally, solar neutrino and ^{208}Tl events can in theory be distinguished by their direction. Via a method such as the one developed in [91, 92], it might be possible to gain some discrimination power of the solar signal over this background.

On the subject of classifiers, the choice of cuts used for the externals classifiers were chosen to be deliberately conservative. By tuning those cuts, it is possible that the sensitivity could be improved through the increased rejection of external events. An even more sophisticated approach would be to include these classifiers as dimensions of the PDFs. This method yielded substantial improvements in the $0\nu\beta\beta$ study performed in [137].

Of course, what has been seen in this analysis is that the greatest barrier to 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 increasing the FV used for this analysis. For example, if the maximum radius used in this analysis was moved from 5.0 m out to 5.7 m, then the increased volume leads to an increased rate of signal events of 48%. Admittedly, for energies below ~ 3.0 MeV at larger radii external backgrounds are expected to completely dominate over the signal, so the benefits in the signal statistics will mostly be seen above 5 MeV.

Because the time of day impacts the extent to which solar neutrinos will be travelling through the Earth, there is expected to be some additional power in considering the difference between events observed during the day versus the night. This approach was used to boost the expected sensitivity for the solar oscillation analysis to be used on the JUNO experiment, for example [175, 176]. Also considered in that sensitivity study were additional interaction modes of solar neutrinos with a liquid scintillator, such as the CC interaction of ν_e on ^{13}C within the detector. These modes could provide further sensitivity.

Finally, analysis of neutrino oscillations from reactor anti-neutrinos in SNO+ is expected to lead to a world-leading limit of Δm_{21}^2 in under 4 years of livetime [90]. Although both analysis approaches have sensitivity to Δm_{21}^2 and θ_{12} , the reactor anti-neutrino method is more sensitive to Δm_{21}^2 whilst the solar analysis shown above is more sensitive to θ_{12} . Because of this complementarity, it is possible that a combined analysis using both types of oscillating neutrino signal incident on SNO+ will lead to even better measurements of both oscillation parameters.

Chapter 7

Conclusions

I believe we've reached the end of our journey.

All that remains is to collapse the innumerable possibilities before us.

Are you ready to learn what comes next?

SOLANUM, *Outer Wilds*

This thesis covered two major aspects of the SNO+ experiment during the scintillator phase. One of these was the first ever neutrino oscillation analysis built and performed on solar neutrino data from the scintillator phase. Using 80.6 days of livetime, the neutrino oscillation parameter θ_{12} was measured to be $38.9^{\circ +8.0^{\circ}}_{-7.9^{\circ}}$, assuming the global fit measurement of the ${}^8\text{B}$ flux, $\Phi_{^8\text{B}} = (5.16^{+2.5\%}_{-1.7\%}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$. In order to make this measurement, a sophisticated Bayesian Analysis framework was developed, using MCMC. This also required building a model of the various background events present in the data, and calibrate the energy scale systematic parameter.

This oscillation analysis is limited by the statistics of the ${}^8\text{B}$ neutrinos. A sensitivity study showed that, assuming the same conditions seen in the dataset analysed, the measurement precision can be improved by a factor of two in under two years of livetime. There is some scope for improvement to this sensitivity with the addition of bis-MSB to the detector, as well as optimisations to the analysis

described in Section 6.6. It may be possible to provide a much stronger result if combined with the reactor anti-neutrino oscillation analysis also being performed with SNO+.

The other major focus of this thesis has been the optical calibration system SMELLIE. Major hardware upgrades were made, improving the stability of the PQ lasers' emission intensities, the triggering system used for SMELLIE, and the Monitoring PMT Unit. A new algorithm for SMELLIE event generation was developed, increasing the speed of simulations by many orders of magnitude. A method for combining SMELLIE data to determine the beam profiles was built and implemented.

One consequence of the building of these new beam profiles is the more precise understanding of existing systematics in the optical model of RAT and SMELLIE. To still make progress despite these issues, a new set of analyses for SMELLIE were developed. One compared water phase and scintillator phase data to measure the extinction length of scintillator as a function of both wavelength and time, and another compared changes in scattering and re-emitted scintillation light. From these measurements, there is evidence of an optical component added to the scintillator between May 2021 and May 2022 which shortens the observed extinction lengths measured in the 400–500 nm range, but decreases the total amount of scattering and re-emitted light. There does not appear to be any substantial change to the observed optics following this.

It is an exciting time for the SNO+ experiment: multiple interesting analyses of scintillator phase data are underway, with calibrations allowing us to get an increasingly more accurate model of the detector. In the near future, tellurium will be loaded into the scintillator cocktail, allowing for the $0\nu\beta\beta$ campaign to begin in earnest. Only time will tell how successful this will be!

References

- [1] The Nobel Prize in Physics 2015, <https://www.nobelprize.org/prizes/physics/2015/summary/>, 2015, Online; accessed 20 September 2023.
- [2] S. L. Glashow, Partial-symmetries of weak interactions, Nuclear Physics **22**, 579 (1961).
- [3] S. Weinberg, A Model of Leptons, Physical Review Letters **19**, 1264 (1967).
- [4] A. Salam, Weak and electromagnetic interactions, in *Selected Papers of Abdus Salam*, edited by A. Ali, C. Isham, T. Kibble, and Riazuddin, World Scientific Series in 20th Century Physics Vol. 5, pp. 244–254, World Scientific, 1994.
- [5] **UA1 Collaboration**, G. Arnison *et al.*, Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ GeV, Physics Letters B **122**, 103 (1983).
- [6] **UA2 Collaboration**, M. Banner *et al.*, Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider, Physics Letters B **122**, 476 (1983).
- [7] **UA1 Collaboration**, G. Arnison *et al.*, Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider, Physics Letters B **126**, 398 (1983).
- [8] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Detection of the Free Neutrino: A Confirmation, Science **124**, 103 (1956).
- [9] F. Reines and C. L. Cowan, The Neutrino, Nature **178**, 446 (1956).
- [10] G. Danby *et al.*, Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos, Physical Review Letters **9**, 36 (1962).
- [11] **DONUT Collaboration**, K. Kodama *et al.*, Observation of tau neutrino interactions, Physics Letters B **504**, 218 (2001).
- [12] **Gargamelle Collaboration**, F. J. Hasert *et al.*, Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment, Physics Letters (1973).
- [13] **Gargamelle Collaboration**, F. Hasert *et al.*, Search for elastic muon-neutrino electron scattering, Physics Letters **46** (1973).

- [14] **Gargamelle Collaboration**, F. J. Hasert *et al.*, Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment, Nuclear Physics B **73**, 1 (1974).
- [15] **Gargamelle Collaboration**, J. Blietschau *et al.*, Evidence for the leptonic neutral current reaction $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$, Nuclear Physics B **114**, 189 (1976).
- [16] M. Goldhaber, L. Grodzins, and A. W. Sunyar, Helicity of Neutrinos, Physical Review **109**, 1015 (1958).
- [17] **The ALEPH, DELPHI, L3, OPAL and SLD Collaborations**, Precision electroweak measurements on the Z resonance, Physics Reports **427**, 257 (2006).
- [18] P. Janot and S. Jadach, Improved Bhabha cross section at LEP and the number of light neutrino species, Physics Letters B **803**, 135319 (2020).
- [19] F. F. Deppisch, Chapter 2: Neutrinos in the Standard Model, in *A Modern Introduction to Neutrino Physics*, IOP Concise Physics, pp. 1–25, Morgan & Claypool Publishers, 2019.
- [20] F. Englert and R. Brout, Broken Symmetry and the Mass of Gauge Vector Mesons, Physical Review Letters **13**, 321 (1964).
- [21] P. W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, Physical Review Letters **13**, 508 (1964).
- [22] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Global Conservation Laws and Massless Particles, Physical Review Letters **13**, 585 (1964).
- [23] **ATLAS Collaboration**, G. Aad *et al.*, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Physics Letters B **716**, 1 (2012).
- [24] **CMS Collaboration**, S. Chatrchyan *et al.*, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Physics Letters B **716**, 30 (2012).
- [25] **Particle Data Group**, R. L. Workman *et al.*, Review of Particle Physics: Review 14 - Neutrino Masses, Mixing, and Oscillations, PTEP **2022**, 083C01 (2022).
- [26] **KATRIN Collaboration**, M. Aker *et al.*, Direct neutrino-mass measurement with sub-electronvolt sensitivity, Nature Physics **18**, 160 (2022).
- [27] E. Di Valentino, S. Gariazzo, and O. Mena, Most constraining cosmological neutrino mass bounds, Physical Review D **104**, 083504 (2021).
- [28] J. N. Bahcall, *Neutrino Astrophysics* (Cambridge University Press, 1989).
- [29] **Borexino Collaboration**, M. Agostini *et al.*, Comprehensive measurement of pp-chain solar neutrinos, Nature **562**, 505 (2018).

- [30] W. T. Winter, S. J. Freedman, K. E. Rehm, and J. P. Schiffer, The ${}^8\text{B}$ Neutrino Spectrum, *Physical Review C* **73**, 025503 (2006), arxiv:nucl-ex/0406019.
- [31] N. Vinyoles, A. Serenelli, and F. L. Villante, The B16 Standard Solar Models, *Journal of Physics: Conference Series* **1056**, 012058 (2018).
- [32] N. Grevesse and A. Sauval, Standard Solar Composition, *Space Science Reviews* **85**, 161 (1998).
- [33] M. Asplund, N. Grevesse, A. J. Sauval, and P. Scott, The Chemical Composition of the Sun, *Annual Review of Astronomy and Astrophysics* **47**, 481 (2009).
- [34] R. Davis, D. S. Harmer, and K. C. Hoffman, Search for Neutrinos from the Sun, *Physical Review Letters* **20**, 1205 (1968).
- [35] J. N. Abdurashitov *et al.*, Measurement of the solar neutrino capture rate with gallium metal. III. Results for the 2002–2007 data-taking period, *Physical Review C* **80**, 015807 (2009).
- [36] M. Altmann *et al.*, Complete results for five years of GNO solar neutrino observations, *Physics Letters B* **616**, 174 (2005).
- [37] J. N. Bahcall, M. Kamionkowski, and A. Sirlin, Solar neutrinos: Radiative corrections in neutrino-electron scattering experiments, *Physical Review D* **51**, 6146 (1995).
- [38] B. T. Cleveland *et al.*, Measurement of the Solar Electron Neutrino Flux with the Homestake Chlorine Detector, *The Astrophysical Journal* **496**, 505 (1998).
- [39] K. Abe *et al.*, Solar neutrino measurements in Super-Kamiokande-IV, *Physical Review D* **94**, 052010 (2016).
- [40] **SNO Collaboration**, J. Boger *et al.*, The Sudbury Neutrino Observatory, *Nuclear Instruments and Methods in Physics Research Section A* **449**, 172 (2000).
- [41] **SNO Collaboration**, B. Aharmim *et al.*, Electron energy spectra, fluxes, and day-night asymmetries of ${}^8\text{B}$ solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory, *Physical Review C* **72**, 055502 (2005).
- [42] **IMB Collaboration**, R. Becker-Szendy *et al.*, Electron- and muon-neutrino content of the atmospheric flux, *Physical Review D* **46**, 3720 (1992).
- [43] **Kamiokande Collaboration**, Y. Fukuda *et al.*, Atmospheric ν_μ/ν_e ratio in the multi-GeV energy range, *Physics Letters B* **335**, 237 (1994).
- [44] **Super-Kamiokande Collaboration**, Y. Fukuda *et al.*, Evidence for Oscillation of Atmospheric Neutrinos, *Physical Review Letters* **81**, 1562 (1998).

- [45] **Super-Kamiokande Collaboration**, Y. Ashie *et al.*, Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations, *Physical Review Letters* **93**, 101801 (2004).
- [46] **LamLAND Collaboration**, A. Gando *et al.*, Reactor on-off antineutrino measurement with KamLAND, *Physical Review D* **88**, 033001 (2013).
- [47] **Daya Bay Collaboration**, D. Adey *et al.*, Measurement of the Electron Antineutrino Oscillation with 1958 Days of Operation at Daya Bay, *Physical Review Letters* **121**, 241805 (2018).
- [48] **Double Chooz Collaboration**, H. de Kerret *et al.*, Double Chooz θ_{13} measurement via total neutron capture detection, *Nature Physics* **16**, 558 (2020).
- [49] **RENO Collaboration**, G. Bak *et al.*, Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO, *Physical Review Letters* **121**, 201801 (2018).
- [50] **T2K Collaboration**, K. Abe *et al.*, Improved constraints on neutrino mixing from the T2K experiment with 3.13×10^{21} protons on target, *Physical Review D* **103**, 112008 (2021).
- [51] **T2K Collaboration**, K. Abe *et al.*, Observation of Electron Neutrino Appearance in a Muon Neutrino Beam, *Physical Review Letters* **112**, 061802 (2014).
- [52] **NOvA Collaboration**, P. Adamson *et al.*, Constraints on Oscillation Parameters from ν_e Appearance and ν_μ Disappearance in NOvA, *Physical Review Letters* **118**, 231801 (2017).
- [53] **NOvA Collaboration**, P. Adamson *et al.*, First Measurement of Electron Neutrino Appearance in NOvA, *Physical Review Letters* **116**, 151806 (2016).
- [54] **T2K Collaboration**, K. Abe *et al.*, Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations, *Nature* **580**, 339 (2020).
- [55] **NOvA Collaboration**, M. A. Acero *et al.*, First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by NOvA, *Physical Review Letters* **123**, 151803 (2019).
- [56] **OPERA Collaboration**, N. Agafonova *et al.*, Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam, *Physical Review Letters* **120**, 211801 (2018).
- [57] Z. Maki, M. Nakagawa, and S. Sakata, Remarks on the Unified Model of Elementary Particles, *Progress of Theoretical Physics* **28**, 870 (1962).
- [58] B. Pontecorvo, Neutrino Experiments and the Problem on Conservation of Leptonic Charge, *Soviet Physics JETP* **26**, 1717 (1968).
- [59] F. F. Deppisch, Chapter 3: Neutrino Oscillations, in *A Modern Introduction to Neutrino Physics*, IOP Concise Physics, pp. 1–27, Morgan & Claypool Publishers, 2019.

- [60] 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).
- [61] L. Wolfenstein, Neutrino oscillations in matter, *Physical Review D* **17**, 2369 (1978).
- [62] S. P. Mikheyev and A. Yu. Smirnov, Resonant amplification of ν oscillations in matter and solar-neutrino spectroscopy, *Il Nuovo Cimento C* **9**, 17 (1986).
- [63] C. Giunti and C. W. Kim, Chapter 9: Neutrino Oscillations in Matter, in *Fundamentals of Neutrino Physics and Astrophysics*, pp. 322–351, Oxford University Press, 2007.
- [64] C. Giunti and C. W. Kim, Chapter 13: Phenomenology of Three-Neutrino Mixing, in *Fundamentals of Neutrino Physics and Astrophysics*, pp. 452–483, Oxford University Press, 2007.
- [65] **LSND Collaboration**, A. Aguilar *et al.*, Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam, *Physical Review D* **64**, 112007 (2001).
- [66] **MiniBooNE Collaboration**, A. A. Aguilar-Arevalo *et al.*, Updated Mini-BooNE neutrino oscillation results with increased data and new background studies, *Physical Review D* **103**, 052002 (2021).
- [67] **MicroBooNE Collaboration**, P. Abratenko *et al.*, First Constraints on Light Sterile Neutrino Oscillations from Combined Appearance and Disappearance Searches with the MicroBooNE Detector, *Physical Review Letters* **130**, 011801 (2023).
- [68] S. Weinberg, Baryon- and Lepton-Nonconserving Processes, *Physical Review Letters* **43**, 1566 (1979).
- [69] P. Minkowski, $\mu \rightarrow e\gamma$ at a rate of one out of 109 muon decays?, *Physics Letters B* **67**, 421 (1977).
- [70] E. Majorana, Teoria simmetrica dell'elettrone e del positrone, *Il Nuovo Cimento (1924-1942)* **14**, 171 (1937).
- [71] W. H. Furry, On Transition Probabilities in Double Beta-Disintegration, *Physical Review* **56**, 1184 (1939).
- [72] M. Goeppert-Mayer, Double Beta-Disintegration, *Physical Review* **48**, 512 (1935).
- [73] **CUORE Collaboration**, D. Q. Adams *et al.*, Measurement of the $2\nu\beta\beta$ Decay Half-Life of ^{130}Te with CUORE, *Physical Review Letters* **126**, 171801 (2021).
- [74] **GERDA Collaboration**, M. Agostini *et al.*, Results on $\beta\beta$ decay with emission of two neutrinos or Majorons in ^{76}Ge from GERDA Phase I, *The European Physical Journal C* **75**, 416 (2015).

- [75] **KamLAND-Zen Collaboration**, A. Gando *et al.*, Precision Analysis of the ^{136}Xe Two-Neutrino $\beta\beta$ Spectrum in KamLAND-Zen and Its Impact on the Quenching of Nuclear Matrix Elements, *Physical Review Letters* **122**, 192501 (2019).
- [76] **NEMO-3 Collaboration**, R. Arnold *et al.*, Measurement of the $2\nu\beta\beta$ decay half-life of ^{150}Nd and a search for $0\nu\beta\beta$ decay processes with the full exposure from the NEMO-3 detector, *Physical Review D* **94**, 072003 (2016).
- [77] F. Kondev, Nuclear Data Sheets for $A = 206$, *Nuclear Data Sheets* **109**, 1527 (2008).
- [78] J. Schechter and J. W. F. Valle, Neutrinoless double- β decay in $SU(2) \times U(1)$ theories, *Physical Review D* **25**, 2951 (1982).
- [79] 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.
- [80] M. Agostini *et al.*, Final Results of GERDA on the Search for Neutrinoless Double- β Decay, *Physical Review Letters* **125**, 252502 (2020).
- [81] **CUPID-Mo Collaboration**, E. Armengaud *et al.*, New Limit for Neutrinoless Double-Beta Decay of ^{100}Mo from the CUPID-Mo Experiment, *Physical Review Letters* **126**, 181802 (2021).
- [82] **CUORE Collaboration**, D. Q. Adams *et al.*, Search for Majorana neutrinos exploiting millikelvin cryogenics with CUORE, *Nature* **604**, 53 (2022).
- [83] **KamLAND-Zen Collaboration**, S. Abe *et al.*, Search for the Majorana Nature of Neutrinos in the Inverted Mass Ordering Region with KamLAND-Zen, *Physical Review Letters* **130**, 051801 (2023).
- [84] **SNO+ Collaboration**, V. Albanese *et al.*, The SNO+ experiment, *Journal of Instrumentation* **16**, P08059 (2021).
- [85] **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).
- [86] **SNO+ Collaboration**, M. Anderson *et al.*, Measurement of the ^8B solar neutrino flux in SNO+ with very low backgrounds, *Physical Review D* **99**, 012012 (2019).
- [87] **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.
- [88] **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).
- [89] **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.

- [90] 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.
- [91] SNO+ Collaboration, A. Allega *et al.*, Event-by-Event Direction Reconstruction of Solar Neutrinos in a High Light-Yield Liquid Scintillator, 2023, arxiv:2309.06341.
- [92] J. Paton, *Directional Reconstruction in Liquid Scintillator Neutrino Detectors*, DPhil Thesis, University of Oxford, 2023.
- [93] 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.
- [94] I. Frank and Ig. Tamm, Coherent visible radiation of fast electrons passing through matter, Compt. Rend. Acad. Sci. URSS **14**, 109 (1937).
- [95] T. Kaptanoglu, L. Segui, A. Mastbaum, N. Barros, and I. T. Coulter, Optics Overview and Proposed Changes to RAT, 2016, SNO+ Internal Document DocDB #3461.
- [96] T. Kaptanoglu, Documentation for Attenuation Studies, 2022, SNO+ Internal Document DocDB #7608.
- [97] J. B. Birks, Chapter 3: The Scintillation Process in Organic Materials I, in *The Theory and Practice of Scintillation Counting*, pp. 39–67, Pergamon Press, 2nd ed., 1967.
- [98] R. Hunt-Stokes, Emission Timing Tuning - 2.2 g / L Optical Model, 2022, SNO+ Internal Document DocDB #7621.
- [99] A. LaTorre, New Measurements for the Timing Spectrum of Electrons and Alphas in LAB+PPO, 2016, SNO+ Internal Document DocDB #4081.
- [100] J. B. Birks, Chapter 6: The Scintillation Process in Organic Materials II, in *The Theory and Practice of Scintillation Counting*, pp. 185–234, Pergamon Press, 2nd ed., 1967.
- [101] S. Riccetto, RAT optics 2.2g/L - after CM_Aug2022, 2022, SNO+ Internal Document DocDB #7571.
- [102] **Particle Data Group**, R. L. Workman *et al.*, Review of Particle Physics: Review 34 - Passage of Particles Through Matter, PTEP **2022**, 083C01 (2022).
- [103] L. Rayleigh, On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **47**, 375 (1899).
- [104] G. Mie, Contributions to the Optics of Turbid Media, particularly of Colloidal Metal Solutions, Annalen der Physik **25**, 377 (1908).

- [105] L. Lorenz, Sur la lumière réfléchie et réfractée par une sphère (surface) transparente, Oeuvres scientifiques de L. Lorenz , 403 (1898).
- [106] A. Einstein, Theory of Opalescence of Homogeneous Fluids and Liquid Mixtures near the Critical State, Annalen der Physik **338**, 1275 (1910).
- [107] M. S. Smoluchowski, Molecular kinetic theory of gases in critical opalescence state, as well as some related phenomena, Annalen der Physik , 205 (1908).
- [108] 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).
- [109] X. Zhang and L. Hu, Estimating scattering of pure water from density fluctuation of the refractive index, Optics Express **17**, 1671 (2009).
- [110] J. T. Zhao, C. Qi, G. Li, and M. A. Schmidt, An improved spectrophotometric method tests the Einstein–Smoluchowski equation: A revisit and update, Physical Chemistry Chemical Physics **22**, 21784 (2020).
- [111] D. J. Coumou and E. L. Mackor, Isotropic light-scattering in binary liquid mixtures, Transactions of the Faraday Society **60**, 1726 (1964).
- [112] J. G. Kirkwood and R. J. Goldberg, Light Scattering Arising from Composition Fluctuations in Multi-Component Systems, The Journal of Chemical Physics **18**, 54 (1950).
- [113] X. Zhou *et al.*, Rayleigh scattering of linear alkylbenzene in large liquid scintillator detectors, Review of Scientific Instruments **86**, 073310 (2015), arxiv:1504.00987.
- [114] B. Moffat, *The Optical Calibration of the Sudbury Neutrino Observatory*, PhD thesis, Queen's University, Kingston, Canada, 2001.
- [115] E. Turner, *A Measurement of Scattering Characteristics of the Detection Medium in the SNO+ Detector*, DPhil Thesis, University of Oxford, 2022.
- [116] M. Chen and I. T. Coulter, Optical properties in RAT, 2012, SNO+ Internal Document DocDB #1664.
- [117] L. Segui, Scintillator Model: Comparison between new data and old model and its performance in RAT, 2015, SNO+ Internal Document DocDB #2774.
- [118] Y. Liu, Attenuation and Scattering of TeBD and the Cocktail, 2016, SNO+ Internal Document DocDB #3880.
- [119] Q. Liu *et al.*, Rayleigh scattering and depolarization ratio in linear alkylbenzene, Nuclear Instruments and Methods in Physics Research Section A **795**, 284 (2015).
- [120] M. Yu *et al.*, Measurements of Rayleigh Ratios in Linear Alkylbenzene, Review of Scientific Instruments **93**, 063106 (2022).

- [121] N. G. Jerlov and E. S. Nielsen, *Optical Aspects of Oceanography* (Academic Press / Symposium on Optical Aspects of Oceanography (1972 : Copenhagen), London ; New York, 1974).
- [122] R. Boardman, *The Detection of Cherenkov Radiation from Neutrino Interactions*, DPhil Thesis, University of Oxford, 1992.
- [123] 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.
- [124] E. Hecht, Section 4.6.2: The Fresnel Equations, in *Optics*, Pearson, 4th ed., 2014.
- [125] M. E. Moorhead, *Reflectors in Cherenkov Detectors*, DPhil Thesis, University of Oxford, 1992.
- [126] S. D. Biller, N. A. Jelley, M. D. Thorman, N. P. Fox, and T. H. Ward, Measurements of photomultiplier single photon counting efficiency for the Sudbury Neutrino Observatory, (1999).
- [127] J. Dunger and E. Turner, Occupancy Analysis for Multiple-Hit Corrections in Large Scintillation Detectors, 2016, NuPhys 2016 poster; SNO+ Internal Document DocDB #4107.
- [128] M. Howe *et al.*, Sudbury neutrino observatory neutral current detector acquisition software overview, *IEEE Transactions on Nuclear Science* **51**, 878 (2004).
- [129] J. C. Loach, *Measurement of the Flux of 8B Solar Neutrinos at the Sudbury Neutrino Observatory*, DPhil Thesis, University of Oxford, 2008.
- [130] M. Dragowsky *et al.*, The ^{16}N calibration source for the Sudbury Neutrino Observatory, *Nuclear Instruments and Methods in Physics Research Section A* **481**, 284 (2002).
- [131] S. Valder *et al.*, A Laserball Calibration Device for the SNO+ Scintillator Phase, in preparation.
- [132] P. G. Jones, *Background Rejection for the Neutrinoless Double Beta Decay Experiment SNO+*, DPhil Thesis, University of Oxford, 2011.
- [133] I. T. Coulter, *Modelling and Reconstruction of Events in SNO+ Related to Future Searches for Lepton and Baryon Number Violation*, DPhil Thesis, University of Oxford, 2013.
- [134] W. Parker, MultiPDF Method for Position Reconstruction, 2021, SNO+ Internal Document DocDB #7257.
- [135] W. Parker, Recoordinating ScintEffectiveSpeed & MultiPDF with Updated 2.2 g/l Optics, 2023, SNO+ Internal Document DocDB #7636.
- [136] M. Mottram, Updated functional form energy reconstruction: EnergyR-ThetaFunctional, 2015, SNO+ Internal Document DocDB #3488.

- [137] T. Kroupová, *Improving the Sensitivity to Neutrinoless Double Beta Decay in SNO+*, DPhil Thesis, University of Oxford, 2020.
- [138] J. R. Sinclair, *Positioning and Timing Calibration of SNO+*, PhD thesis, University of Sussex, 2015.
- [139] K. Majumdar, *On the Measurement of Optical Scattering and Studies of Background Rejection in the SNO+ Detector*, DPhil Thesis, University of Oxford, 2015.
- [140] L. Cavalli *et al.*, The SMELLIE Hardware Manual, 2016, SNO+ Internal Document DocDB #3511.
- [141] S. Langrock, *Measurement of the Rayleigh Scattering Length and Background Contributions during Early Data Taking Phases at SNO+*, PhD thesis, Queen Mary, University of London, 2016.
- [142] J. Lidgard, Supercontinuum addition to SMELLIE.RATDB, 2018, SNO+ Internal Document DocDB #4816.
- [143] D. Fibreoptics, 30dB Precision Variable Attenuator Specifications, 2017, URL: <https://www.diconfiberoptics.com/products/scd0069/SCD-0069C.pdf>, accessed 28/07/2023.
- [144] Corning, Corning Infinicor 50 μm Optical Fibres Product Information, 2007, URL: <http://www.princtel.com/datasheets/Infinicor50.pdf>, accessed 28/07/2023.
- [145] K. J. Clark, ELLIE Fibre Radioactivity, 2011, SNO+ Internal Document DocDB #1223.
- [146] C. Jones, *Future Searches for Rare Astrophysical Signals and Detector Commissioning in SNO+*, DPhil Thesis, University of Oxford, 2016.
- [147] S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, *The Annals of Mathematical Statistics* **9**, 60 (1938).
- [148] T. Kroupová, PMT Transit Time Tuning, 2018, SNO+ Internal Document DocDB #5353.
- [149] **SNO+ Collaboration**, M. Anderson *et al.*, Development, characterisation, and deployment of the SNO+ liquid scintillator, *Journal of Instrumentation* **16**, P05009 (2021).
- [150] A. Zummo, Residual Energy Correction, 2023, SNO+ Internal Document DocDB #7895.
- [151] B. Tam, Benchtop Attenuation Measurements, 2022, SNO+ Internal Document DocDB #7448.
- [152] B. Tam, *Enabling Neutrinoless Double Beta Decay in The SNO+ Experiment Through the Deployment and Study of Liquid Scintillator*, PhD thesis, Queen's University, Kingston, 2023.

- [153] S. Riccetto, Full Fill 210 NHits and Radial Profile, 2022, SNO+ Internal Document DocDB #7489.
- [154] P. Khaghani, Neck sense rope system and leaching studies for SNO+, Master's thesis, Laurentian University of Sudbury, 2015.
- [155] P. Khaghani, *¹⁶N Calibration and Background Studies for SNO+*, PhD thesis, Laurentian University of Sudbury, 2022.
- [156] B. Tam, Attenuation Measurements, 2022, SNO+ Internal Document DocDB #7481.
- [157] J. Caravaca, SNO+ Sensitivity to Standard Solar Neutrino Oscillations, 2020.
- [158] J. Dunger, *Topological and Time Based Event Classification for Neutrinoless Double Beta Decay in Liquid Scintillator*, DPhil Thesis, University of Oxford, 2018.
- [159] J. R. Wilson, BiPo Rejection Factors Summary, 2017, SNO+ Internal Document DocDB #4183.
- [160] R. Hunt-Stokes, Uranium and Thorium Background Analysis, 2022, SNO+ Internal Document DocDB #7664.
- [161] E. Marzec, BiPo Event Reduction using Likelihood Ratio Cut, 2013, SNO+ Internal Document DocDB #1994.
- [162] Z. Ye, Tagging in-window BiPo events with AlphaBetaClassifier – Recoordinated for 2.2 g/L LABPPO new optics, 2022, SNO+ Internal Document DocDB #7659.
- [163] **SNO+ Collaboration**, S. Andringa *et al.*, Current Status and Future Prospects of the SNO+ Experiment, *Advances in High Energy Physics* **2016**, 1 (2016).
- [164] 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).
- [165] 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* **774**, 103 (2015).
- [166] C. Sanderson and R. Curtin, Armadillo: A template-based C++ library for linear algebra, *Journal of Open Source Software* **1**, 26 (2016).
- [167] 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.

- [168] N. Barros, *Precision Measurement of Neutrino Oscillation Parameters: Combined Three-Phase Results of the Sudbury Neutrino Observatory*, PhD thesis, Universidade de Lisboa, 2011.
- [169] A. M. Dziewonski and D. L. Anderson, Preliminary reference Earth model, Physics of the Earth and Planetary Interiors **25**, 297 (1981).
- [170] D. Cookman, Solar Oscillation Analysis Update, 2023, SNO+ Internal Document DocDB #7707.
- [171] R. Hunt-Stokes, Private Communication, 2023.
- [172] A. Gelman *et al.*, Chapter: 11: Basics of Markov chain simulation, in *Bayesian Data Analysis*, 3 ed., 2013.
- [173] 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.
- [174] J. Bergström *et al.*, Updated determination of the solar neutrino fluxes from solar neutrino data, Journal of High Energy Physics **2016**, 132 (2016).
- [175] **JUNO Collaboration**, A. Abusleme *et al.*, Feasibility and physics potential of detecting 8B solar neutrinos at JUNO, (2020), arxiv:2006.11760.
- [176] **JUNO Collaboration**, J. Zhao *et al.*, Model Independent Approach of the JUNO 8B Solar Neutrino Program, (2022), arxiv:2210.08437.
- [177] M.-P. Lebeuf, LAB Certificate of Analysis, 2020, Example of document provided by CEPSA for each tanker of LAB used to fill SNO+. SNO+ Internal Document DocDB #7572.
- [178] J. Caravaca, Validation of SNO+ Production Signal, 2019, SNO+ Internal Document DocDB #5519.
- [179] J. D. Wilson, Thermally-driven scintillator flow in the SNO+ neutrino detector, Nuclear Instruments and Methods in Physics Research Section A **1055**, 168430 (2023).
- [180] M. J. Martin, Nuclear Data Sheets for $A = 208$, Nuclear Data Sheets **108**, 1583 (2007).
- [181] M. Shamsuzzoha Basunia, Nuclear Data Sheets for $A = 210$, Nuclear Data Sheets **121**, 561 (2014).
- [182] S. Riccetto, Private Communication, 2023.
- [183] S. Riccetto, Full Fill Internal Po-210 Collaboration Meeting Update, 2023, SNO+ Internal Document DocDB #7733.
- [184] V. Lozza and O. Chkvorets, Neutron sources and backgrounds in SNO+, 2015, SNO+ Internal Document DocDB #2497.
- [185] S. Yu, AV Surface Po210 Analysis, 2023, SNO+ Internal Document DocDB #7733.

- [186] A. Zummo, External Background Box Analysis Plot/Results, 2022, SNO+ Internal Document DocDB #7329.
- [187] T. Kaptanoglu, First look at external backgrounds in the scintillator phase, 2023, SNO+ Internal Document DocDB #7817.
- [188] M. J. Berger *et al.*, XCOM: Photon Cross Sections Database, NIST (2009).
- [189] W. J. Heintzelman, Survival to Radii Inside the PSUP of γ s from PMT β - γ Events, 2012, SNO+ Internal Document DocDB #1729.
- [190] W. J. Heintzelman, Angular distribution of surviving γ s from PMT β - γ events, 2013, SNO+ Internal Document DocDB #1712.
- [191] A. S. Inácio, Using a Smaller Shell when Simulating Bi214 and Tl208 in the External Water and Ropes, 2019, SNO+ Internal Document DocDB #5154.

Appendix A

Expected Rates and Constraints in the Solar Analysis

A.1 ${}^8\text{B}$ Signal

Numerous theoretical predictions and experimental measurements have been made of the ${}^8\text{B}$ flux. For this analysis, two different values have been used. The SSM predicts a relatively loose constraint of $\Phi_{\text{B}} = (5.46 \pm 12\%) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ [31], whereas a recent global fit of neutrino oscillation experiments by Bergström *et al* leads to a much stronger constraint of $\Phi_{\text{B}} = (5.16^{+2.5\%}_{-1.7\%}) \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ [174]. We shall mainly use the latter result to both calculate the expected rate of the signal events in the detector, and the fractional uncertainties used to constrain this rate. The looser constraint coming from the SSM will be used for comparison.

To calculate the number of electrons within the liquid scintillator, the method used in [79] is followed, which uses the formula:

$$n_e = \frac{(f_{\text{LAB}} n_{\text{LAB}} + f_{\text{PPO}} n_{\text{PPO}}) N_A M_{\text{LAB}}}{m}. \quad (\text{A.1})$$

Here, f_{LAB} and f_{PPO} are the fraction by weight of the LAB and PPO within the scintillator cocktail, respectively. Because this analysis uses 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 $\text{C}_{15}\text{H}_{11}\text{NO}$, 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 the breakdown provided here has been used [177], 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:

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

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

$$R_i = \Phi_{^8\text{B}} n_e \int S_\nu(E_\nu) \sigma_{\nu_i}(E_\nu) dE_\nu. \quad (\text{A.2})$$

Using the cross-section formula from Eq. 1.4, the ${}^8\text{B}$ spectral shape from [30], and $\Phi_{^8\text{B}} = 5.46 \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$ ($5.16 \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$), the expected rate before

considering oscillations or cuts is:

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

In the above rates, an additional volume correction factor of 1.0139 has been included, because in MC events are simulated within the neck of the AV in addition to the main spherical bulk [178].

A.2 Internal Uranium- and Thorium-Chain Backgrounds

As mentioned in Section 6.2, $^{214}\text{Bi-Po}$ and $^{212}\text{Bi-Po}$ decays, coming from the ^{238}U and ^{232}Th decay chains respectively, are capable of generating distinctive delayed coincidence events. Rafael Hunt-Stokes was able to use a series of cuts to isolate both types of coincidence signals [160], looking at coincidence signals within the same 5.0 m FV as this analysis. After correcting for his cut efficiencies, he obtained a derived rate in the whole detector of [171]:

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

These rates assume a uniform concentration of ^{222}Rn and ^{220}Rn throughout the detector: this is how internal backgrounds are simulated by default in **RAT**. However, it has been shown that this is not at all the case within SNO+; radon ingress from the AV neck propagates through the bulk of the detector via convection currents induced by the thermal gradient present throughout the detector [179]. This radon ingress also breaks the secular equilibrium of the two decay chains, so

that one can no longer make straightforward predictions about rates for decays in the chains above radon. Fortunately for this analysis, Table 6.4 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 Bi–Po events that survive all the cuts is 1.5 and 2.5 for ^{212}Bi –Po and ^{214}Bi –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% [180, 181], after considering the cut efficiencies in this analysis 512.4 internal ^{208}Tl and 1.1 internal ^{210}Tl events are expected.

In theory, the non-uniformity of all the radon-induced backgrounds could impact the results of the oscillation analysis. However, the small rates of each of these backgrounds after cuts have been applied indicated 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}Bi –Po, and ^{214}Bi –Po) did not have this PDF 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.

A.3 (α, n) Reactions

As discussed in Section 6.2, (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 surface ^{210}Po events have been tracked throughout the scintillator phase by Serena Riccetto and Shengzhao Yu, respectively.

Within the 5 m FV, S. Riccetto measured a total of 1.60×10^8 ^{210}Po events over the runs considered for this analysis [182, 183]. Assuming the internal rate of these events are uniform throughout the scintillator, this leads to a total rate of 2.78×10^8 events for all the scintillator. After using a conversion rate of 6×10^{-8} neutrons generated per α on ^{13}C in liquid scintillator [90, 184], the expected number of internal (α, n) events before cuts is 17. However, because a coincidence cut is used in this analysis, the vast majority of these should be removed, leading to merely 0.009 events expected after cuts.

Similarly, S. Yu measured the surface ^{210}Po rate to vary between 2–5 Bq m^{-2} over the time period of this analysis as well as a function of height in the detector [185]. Using the midpoint value of 3.5 Bq m^{-2} , one can derive an expected rate of 554.4 surface (α, n) events occurring over the dataset’s livetime, before cuts. Once again, the coincidence cut as well as the FV cut removes the vast majority of these events, so that only 0.07 events are expected after all cuts have been applied. Because both classes of (α, n) events have negligible expected rates in the dataset, they were not included within the MCMC fit.

A.4 External Backgrounds

During the water phase of SNO+, an analysis of the rates of the external backgrounds was performed by Tony Zummbo [186]. The results of this analysis are shown in Table A.1, giving the measured rates as a fraction of the nominal values given in [163]. Although the statistical uncertainty for these measurements were quite small, the total systematic uncertainty was substantial because T. Zummo’s

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 A.1: Measured rates of the external backgrounds during the water phase of SNO+, by Tony Zummo [186].

analysis involved looking at events in the tail of energy distributions, so that any uncertainty in the energy scale systematic had a strong impact on the number of events observed within the tail. These measured rates and their 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 have changed over the years since T. Zummo’s analysis dataset was taken [187]. 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 assuming the same rate of decays per unit volume, 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 [188, 189], 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, PMT $\beta - \gamma$ events are modelled as 2.6 MeV γ s generated at a radius of 6.2 m (just outside the AV), pointing radially inwards. Making some basic assumptions about these events, one can derive the expected survival rate of these γ particles as a function of radius [189, 190]. This leads to a correction factor of 1.17×10^{-6} being applied to the results of the radial shell simulation.

Similarly, simulations of background events in the external water and ropes are restricted from starting beyond certain maximum radii. The particular maximum radii chosen depends on the specific process, in order to ensure a negligible impact on the fraction of simulated events that actually deposit any energy into the scintillator. This leads to rate correction factors of 0.35 and 0.50 for external water ^{214}Bi and ^{208}Tl events, as well as 0.50 and 0.35 for hold-down and hold-up ropes [191].