



---

# Studies of Optical Scattering and First Measurement of the Solar Neutrino Oscillation Parameters in the SNO+ detector

---

Daniel Cookman  
Confirmation of Status Report  
Hilary Term, 2022  
Department of Physics  
University of Oxford

## **Abstract**

This report contains four main sections. A draft of the introduction to be included within my thesis is given. Following this, the SNO+ detector is briefly overviewed, as well as its scattering calibration system SMELLIE. The main body of the report contains work performed to improve the simulation quality of this calibration system: uncertainties in the angular emission distribution have been improved significantly through the use of more calibration data as well as a careful maximum likelihood estimate statistical approach. In addition, a new generator for simulating these SMELLIE events has been built, which dramatically reduced simulation times by three orders of magnitude. Finally, an outline of my thesis is included, describing the planned contents for each chapter, along with what work is left to be completed in each. A description of what data will be used as part of this thesis is included.

# 1 Thesis Introduction

Ever since their postulation by Wolfgang Pauli almost 100 years ago, the observational consequences of neutrinos have forced us to understand our Universe better. The apparent violation of energy and angular momentum conservation of radioactive  $\beta$ -decay led Pauli to propose the particle in the first place [], but the same pattern of surprising results has continued apace. Pauli suspected that the neutrino could never be detected, and then Cowan & Reines were able to detect (anti)neutrinos from a nuclear reactor before Pauli's death []. In the subsequent decades, two further flavours of neutrino were detected (muon and tau neutrinos) [], the three flavours coming from a variety of sources: not only the aforementioned radioisotopes and nuclear reactors, but also particle accelerators [], the atmosphere [], the Earth itself [], a distant supernova [], and even our Sun []. To this day, a major puzzle remains: we observe neutrinos oscillating between their flavours when they propagate through space, which mean they must have a mass. But in the Standard Model (SM) of particle physics, there is no way for neutrinos to obtain mass!

SNO+ is an experiment currently running whose flagship physics goal is the search for 'smoking-gun' evidence of a means by which neutrinos could get their mass: neutrinoless double-beta decay. The detector is also able to perform a wide variety of other neutrino physics, including measurement of solar neutrino oscillations — it is measuring these oscillations that forms a major part of this thesis. The neutrino physics community currently has a good phenomenological model for how neutrinos oscillate, known as the PMNS matrix. Most of the parameters within this matrix have been measured precisely by multiple experiments. However, it is quite possible that this oscillation model is wrong, and so further precise measurements of its parameters are highly valued. Solar neutrinos allow us to probe two of these parameters, known as  $\Delta m_{21}^2$  and  $\theta_{12}$ . SNO+ in particular is uniquely poised to be able to measure these parameters not only by solar neutrinos, but also anti-neutrinos generated from reactors. Combining these results together could give a world-leading limit. My work shown here will not only be the first analysis of data from the detector in pursuit of measuring solar neutrino oscillations, but also will provide a sensitivity study for how we expect this preliminary result to improve with more data in the future.

Of course, for SNO+ to be able to make any measurements with precision requires a well-calibrated detector. In particular, understanding how light interacts with the detector is of paramount importance: it is light that gets generated by the physics events we would like to study within our detector. One of the principal aspects by which that light can interact is by scattering off of the detector medium. When there is a high degree of scattering in the detector, light will bounce around a lot before being detected. In this scenario, our ability to reconstruct the position of the inciting physics event becomes limited, which is critical for precise analysis work. SMELLIE is a calibration system for SNO+ designed to measure and monitor this optical scattering. This forms the other major part of my thesis: taking data with SMELLIE, and then analysing that data to understand the detector's scattering properties. This measurement has never before been done in a large liquid scintillator experiment, such as SNO+.

I will start the thesis in Chapter 1 by giving a run-down of the theoretical concepts critical to understanding the neutrino physics being researched here. Following this, Chapter 2 is a practical look at how the SNO+ detector works: going from how physics events produce light in the detector to how those photons get detected, and then subsequently packaged into data that gets stored and analysed. It will also cover some of the basic in how this data gets calibrated and reconstructed for high-level analyses.

Chapter 3 will cover the theory necessary to understanding optical scattering in a fluid, and how we model this scattering in our simulations. Then, Chapter 4 takes a look at the specifics of the SMELLIE calibration system: its hardware, and how it gets commissioned for effective use. After this look at hardware, we jump to the software side of things and consider the simulation of SMELLIE events in Chapter 5. A number of major improvements were made here, notably making the simulation happen orders of magnitude faster, and also building a method to calibrate this simulation with far more data than used before.

Chapter 6 is split into two parts, both covering how SMELLIE can be used to calibrate the detector. In the first section, a careful comparison of data between when the detector was filled with water versus scintillator allows us to measure the extinction length of the scintillator as a function of

wavelength, and monitor this over time. The second section then covers how a more extensive analysis of SMELLIE data, and comparing to simulations can lead to one being able to make a measurement of scattering within the detector in both the water and scintillator phases.

Chapter 7 is also split into two parts covering my analysis of solar neutrino oscillations. I first build up the analysis approach with simulated data, and perform a sensitivity study of the detector to measuring the solar oscillation parameters. The second section then applies this to data taken during the scintillator phase of the experiment, and the first ever parameter estimates are obtained for SNO+. Finally we summarise all of the work described in this thesis in Chapter 8, the conclusion.

## 2 The SNO+ Detector and SMELLIE

SNO+ is a large-scale liquid scintillator neutrino detector based in SNOLAB, Canada. It is built from the infrastructure used in the pioneering SNO experiment: a 12m diameter spherical acrylic vessel (AV), that is submerged within a cavity of ultra-pure water. Surrounding this AV is a geodesic stainless steel support structure for  $\sim 9000$  PMTs, which detect the light generated by events within the detector. A diagram of the detector’s geometry can be seen in figure ???. The experiment is split into four phases: an initial filling of the inner-AV with ultra-pure water, followed by filling with the liquid scintillator linear alkylbenzene (LAB), primary fluor 2,5-diphenyloxazole (PPO), and wavelength-shifter bisMSB. Following this, the scintillator cocktail will be loaded with 0.3 % by mass of Te-diols, which contains the  $\beta\beta$ -emitting isotope  $^{130}\text{Te}$ . Filling of the liquid scintillator had to stop for half a year during the pandemic, and forms an additional fourth stable phase known as the partial-fill phase, as the detector was half-filled with the scintillator cocktail on top. Currently all the liquid scintillator has been added and data is being taken, with PPO-loading nearly complete.

One of the main optical calibration systems in SNO+ is known as the Embedded LED/Laser Light Injection Entity, ELLIE. It is composed of three modules, each of which play a different role in calibrating the detector. One of these is the Scattering Module, known as SMELLIE for short. SMELLIE fires laser light at visible wavelengths through one of fifteen optical fibres attached to the PSUP, as shown in figure ??. Laser heads of two different varieties are used: four PicoQuant (PQ) diode lasers of wavelengths 375 nm, 407 nm, 446 nm, and 495 nm, as well as a Super-Continuum (SuperK)<sup>1</sup> that can emit laser light over any chosen interval of a wide band of frequencies.

## 3 SMELLIE Beam Profiling

### 3.1 Previous attempts at SMELLIE event simulation

Critical to extraction of scattering information from SMELLIE data is an accurate Monte Carlo (MC) simulation of the SMELLIE system. By modelling the laser light emission into the detector correctly, we can simulate how SMELLIE light will be impacted by changing scattering lengths in the detector. Because of the complexity of the optics of the optical fibres used to direct the laser light into the detector, a given SMELLIE event is simulated as a point-like “flash” of visible photons emanating from the emission point of the fibre into the detector. This flash then requires a number of parameters to be correctly described. In particular, the fibre emission positions were recorded during the installation of the fibres. The wavelength and timing distributions were taken from measurements of the laser heads by their manufacturers, or by colleague Jeff Lidgard in the case of the SuperK wavelength distribution. The intensity is defined as the mean number of photons simulated per event; on an event-by-event basis we sample Poisson fluctuations about that mean intensity value. Determination of intensity must be done on a subrun-by-subrun basis. Unlike scintillation light, light from SMELLIE is not at all isotropic, and so we must specify some form of angular emission distribution. Determining and handling these angular emission distributions, also known as beam profiles, is the focus of this chapter.

Before we can determine the beam profiles, we must first decide how to specify them. Previous observations show that different fibres can have notably different angular profiles [], so we let each fibre’s beam profiles be unique. We assume for now that a given fibre’s beam profile is stable over time,

<sup>1</sup>Apologies to those who work at Super-Kamiokande; the laser was named this by the manufacturer.

and independent of the wavelength of light fired<sup>2</sup>. A straightforward, naïve approach to parameterising a beam profile would be as follows: specify some nominal fibre direction, corresponding to the direction light takes travelling from the fibre to the centre of the “beamspot” on the other side of the detector. Then, specify a 1D angular distribution, corresponding to the probability density of firing a photon at a given polar angle  $\alpha$  relative to the nominal direction. One might even assume this distribution is Gaussian in shape. The distribution in azimuthal direction,  $\phi$ , is assumed to be uniform.

This 1D beam profile approach is currently used by both TELLIE and AMELLIE for their simulations, and was used initially for SMELLIE. 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 ?? for an example. Not only was the distribution in  $\alpha$  not Gaussian, a distinct speckle-pattern can be observed within the beamspot that is not uniform in  $\phi$ . This fact led to colleague Esther Turner building a SMELLIE generator that could handle 2D beam profiles: angular emission distributions dependent on both  $\alpha$  and  $\phi$ . The distribution was stored as a mapping 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 ??.

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

1. Propose a test direction  $(\alpha, \phi)$ , by generating  $\phi$  uniformly in the interval  $[0, 2\pi]$ , and  $\alpha$  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  $(\alpha, \phi)$ , 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 certainly works, but has a key problem: efficiency. The 1D generator was able to generate a SMELLIE event (that is, to fully specify the starting parameters of all the photons emitted from a fibre) at a speed of  $\sim 1$  ms. However, the 2D generator specified here could take upwards of  $\sim 50$  s *per event* to generate. Because a typical SMELLIE analysis requires simulating many millions of events, the CPU time taken to perform this quickly became unfeasible. Fixing this generator speed problem was a high priority for the SMELLIE analysis.

### 3.2 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 ?? shows a histogram of number of attempts per event it took for a valid direction to be chosen for a representative SMELLIE simulation. Moreover, the calculations needing to be done for every proposed direction are relatively complex, notably trying to find the 3 nearest PMTs to some point on the PSUP.

A new 2D generator was built with these thoughts in mind. Firstly, the rejection method would no longer be used, given its inefficiency. We would also endeavour to try and “pre-calculate” as much

---

<sup>2</sup>These two assumptions certainly make the analysis in the chapter simpler, but they might not be necessarily be true! See ?? for some further discussion.

as possible before run-time. Starting with the existing PMT relative intensity maps, we plot these in the 2D direction-space  $(1 - \cos \alpha, \phi)$ : see Figure ?? . In a toy-MC simulation, 500,000 directions are then thrown uniformly in this 2D space per fibre. For each direction, the same method of obtaining an interpolated intensity value from the nearest PMTs to the corresponding point on the PSUP as from the original 2D generator was performed, the only difference being that these calculations were done well before any actual SMELLIE simulation. Figure ?? 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  $\phi$  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  $\phi$  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 further out. This allows us to both capture any rapid changes in intensity near the beamspot, where this matters greatly, as well as smooth out the very-low intensities seen at larger polar angles. One of these histograms can be seen in Figure ?? : 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  $\phi$ . 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 ?? .

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  $\phi$ , and then  $r$ . The CDF was then normalised to 1 so that it was well-defined. It is this CDF object that is then loaded in and sampled from during event generation. To do this, an “inverse-CDF” approach was used, which has the major benefit over rejection sampling of always producing a valid direction for every sample made. The algorithm works as follows:

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  $\phi$  of this selected bin: use linear interpolation of the random number to obtain a  $\phi$  value located between these two  $\phi$ -values.
4. Look at the 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  $\alpha$ .
5. The photon’s direction is defined by the  $(\alpha, \phi)$  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  $\sim 1$  ms per SMELLIE event, a speed improvement of nearly 50000. Event generation became 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 ?? 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 clearly that both generators are as accurate as one another.

### 3.3 Improving the beam profiles

Even with the new 2D profile generator, a problem remains: 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 of zero, providing little useful information about the beam shape. It was shown in [] that with the old 2D generator, the systematic uncertainty on the beam profiles was the dominant source of error in the main SMELLIE analysis. To help resolve this, 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 long wavelength, 495 nm, was chosen for each fibre. “Medium”-intensity corresponds to firing the relevant laser at a set intensity determined during an earlier commissioning process, for which the maximum occupancy of PMT hits at that intensity, i.e. the proportion of hits per event, corresponded to roughly 80%. This value was chosen as it allowed for high statistics in a relatively short run-time, but not so intense that the occupancy of any given PMT in the beamspot was 100%. Because Rayleigh scattering is strongly-dependent on wavelength, the long wavelength of light was chosen so that impacts from this scattering were small in the data.

SNO+ PMTs are unable to distinguish between one or more photoelectrons being generated (except for very large numbers, at which point the charge collected can provide some useful discriminating power). 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,  $\mu$ . Assuming the number of photoelectrons generated in a given event follows Poisson statistics, the probability of generating  $k$  photoelectrons is:

$$P(k|\mu) = \frac{\mu^k e^{-\mu}}{k!}. \quad (1)$$

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

$$P(\text{hit}|\mu) = P(k \geq 1|\mu) = 1 - P(k = 0|\mu) = 1 - e^{-\mu}, \quad (2)$$

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

$$\mu = \ln(1 - o). \quad (3)$$

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

Once the npe mapping from data was obtained, a correction was then made for the detector’s optics: even ignoring a fibre’s angular distribution, we still expect certain PMTs to be illuminated more than others because of e.g. reflections off of the AV, or the solid angle subtended by the PMT bucket opening. For each fibre, a simulation was made where the beam profile was set as uniform, 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 ??.

### 3.4 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  $\sim 490$  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 1. Because high-intensity runs require a different analysis approach (PMTs with high occupancies must use charge, not occupancy, to estimate npe), for this analysis we only considered subruns that used low or medium intensity set-points.

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 1: Water-phase runs used for new beam profiling.

For each subrun  $j$  of data per fibre, we look only at PMT hits for each PMT  $i$  that has been identified as “good” for that subrun,  $i \in G_j$ <sup>3</sup>. 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 of any reasonable beam direction. To isolate the hits arriving directly from the fibre without reflecting, scattering, or being noise, a time cut was also made. Because what matters is the time relative to emission from the fibre, and the expected time-of-flight from fibre to different PMTs varies, a quantity known as the time residual was used. Starting with the calibrated hit time of a given PMT  $t_{hit}$ , the expected time-of-flight  $t_{TOF}$  was subtracted off, estimated with the collaboration’s “Light Path Calculator”. Then, the emission time was also subtracted off,  $t_{emm}$ , estimated by looking at the second-earliest value of  $t_{hit} - t_{TOF}$  within the fibre’s beamspot, defined as the PMTs for which  $\alpha < 3^\circ$ . It was found that a “loose” time residual cut of  $t_{res} \in [-10, +12]$  ns was sufficient to remove the vast majority of non-direct light with little signal sacrifice.

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

$$\mu_{ij} = I_j k_i = I_j b_i f_i. \quad (4)$$

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

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

$$P(m_{ij} = m | \mu_{ij}) = L(\mu_{ij} | m_{ij} = m) = \binom{N_j}{m} p_{ij}^m (1 - p_{ij})^{N_j - m} = \binom{N_j}{m} (1 - e^{-\mu_{ij}})^m e^{-\mu_{ij}(N_j - m)}. \quad (5)$$

Here we have used equation 2, and noted that this probability distribution in  $m$  can be re-framed as a likelihood function for the parameter  $\mu_{ij}$ . Considering only a single subrun of data, the maximum likelihood estimate of the parameter  $\mu_{ij}$  can be shown to be:

$$\langle \mu_{ij} \rangle = -\ln \left( 1 - \frac{m_{ij}}{N_j} \right) = \ln(1 - o_{ij}) \quad (m_{ij} \neq N_j), \quad (6)$$

where  $o_{ij}$  is just the occupancy of PMT  $i$  in subrun  $j$ . This is just the multi-hit correction formula seen in equation 3, which makes sense.

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

$$L(\{I_j\}, k_i | \{m_{ij}\}) = \prod_j L(I_j, k_i | m_{ij}) = \prod_j \binom{N_j}{m_{ij}} (1 - e^{-I_j k_i})^{m_{ij}} e^{-I_j k_i (N_j - m_{ij})}. \quad (7)$$

<sup>3</sup>Strictly 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.

This leads to a log-likelihood distribution of

$$\mathcal{L}(\{I_j\}, k_i | \{m_{ij}\}) = \sum_j [\ln({}^{N_j}C_{m_{ij}}) + m_{ij} \ln(1 - e^{-I_j k_i}) - I_j k_i (N_j - m_{ij})]. \quad (8)$$

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

$$i \in I = \bigcap_j G_j. \quad (9)$$

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. For example, a simulation for run 114,023 contained **NEVENTS** number of 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 live, etc.).

For each isotropic MC run, both  $I_j^{MC}$  and  $k_i^{MC}$  were calculated via the method described above they would be. 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, we find that:

$$\mu_{ij} = I_j b_i f_i = c S_j b_i a k_i^{MC} = (ac b_i) S_j k_i^{MC}. \quad (10)$$

As a result of this, given the set  $\{S_j\}$  and  $k_i^{MC}$ , one can maximise the log-likelihood  $\mathcal{L}$  with respect to  $b'_i = ac b_i$  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 3.2.

Figure ?? 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 together, than if one looked at a single subrun alone.

Another benefit of using this log-likelihood approach is that the resulting distribution's shape can be used for uncertainty estimation. In almost all cases, Wilks Theorem [] allows us to produce  $1\sigma$  confidence intervals about the maximum likelihood estimate for  $b'_i$ ,  $\langle b'_i \rangle$ , because  $-2[\mathcal{L}(b'_i) - \mathcal{L}(\langle b'_i \rangle)] := X(b'_i)$  approximates a  $\chi^2$ -distribution. As a result, the error bounds on our parameter estimate are given by when  $X = 1$ . Figure ?? shows this parameter estimation in action. 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: we need not worry about this, as subruns where this occur have not been used. On the other end, however, there are some PMTs for certain fibres where after all subruns of data have been included, there remains no hits. In this scenario, one can show that the log-likelihood becomes linear in the beam profile parameter:

$$\mathcal{L}(b'_i | \{m_{ij} = 0\}) = b'_i k_i^{MC} \cdot \sum_j [I_j N_j]. \quad (11)$$



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

$$b'_{i,ulim} = \frac{k_i^{MC} \sum_j [I_j N_j]}{-\ln [1 - \text{erf}(1/\sqrt{2})]}, \quad (12)$$

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

### 3.5 Results

Figure ?? 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.

## 4 Thesis Outline

What follows is an outline of what will be contained in each of the main chapters of my thesis (excluding any front matter or the introduction). The data to be used for this thesis comes in two camps, one for the SMELLIE analysis, and the other for the solar analysis. For the former, the necessary data has already been taken for SMELLIE calibration in both the water and scintillator phases. If possible, further SMELLIE data will be taken as the optical properties of the scintillator change: when bisMSB gets added, for example. Of course, because of the data taken already, there is sufficient data for the analysis to be completed in both the water and scintillator phases. For the solar analysis, roughly 3 months of ‘good’ scintillator data has already been taken, and so once the analysis has been fully developed it can be run on that after any re-processing of that data has occurred. As more scintillator data gets taken in the coming months, it is hoped (but not necessary for the purposes of finishing my thesis) that this additional data can also be included.

### 4.1 The Theory of Neutrino Physics

The purpose of this chapter is to: convince the reader that neutrino physics is indeed an exciting field worthy of contemporary study (especially neutrinoless double-beta decay), give some context to the SNO+ experiment and my work, and go over the theoretical underpinnings of the subject necessary for understanding my research.

I shall start by giving a “whistle-stop” tour of the Standard Model (SM) of particle physics: the combination of Quantum Field Theory with non-Abelian gauge theories, and notably the Brout-Englert-Higgs Mechanism that ties everything together. I will cover the terms of the SM Lagrangian that contain neutrinos, and then describe how this gives rise to various properties of neutrinos. These include the three flavours of neutrino which couple to the weak force in both charged- and neutral-current forms. Experimental evidence supporting this model will be given.

Major assumption of SM has been that neutrinos are massless — I’ll explain why direct coupling to the Higgs in the “usual” way is not possible. But, neutrino oscillations change all that. I will briefly cover the most popular solution to the question of neutrino mass, the see-saw mechanism (which can be of multiple varieties). This will link in to a discussion about the difference Dirac and Majorana models for neutrinos, as well as models for baryo- and lepto-genesis. Then, a description of neutrinoless double-beta decay will be made, and how the Black Box Theorem helps us. I will then mention the differing experimental approaches to searching for neutrinoless double-beta decay.

After that, I will outline the history of measuring neutrino oscillations, starting with Homestake & John Bahcall, but then moving onto more recent notable experiments. With these observations in hand, I will describe the current phenomenological model we use for neutrino oscillations: the PMNS matrix with the MSW effect in the LMA regime. I will give more focus on the theory behind solar neutrino oscillations, given my research work on solar analysis. The Standard Solar Model will be introduced, along with what has yet to be measured fully: the solar metallicity problem. I will describe why only two oscillation parameters can reasonably be measured in a solar neutrino experiment, and how these measurements can be complementary to reactor anti-neutrino measurements.

Because I am already fairly confident in this theory, understanding the parts of this chapter will not take much more of my time. However, it will require a not-insignificant amount of finding and reading papers, and therefore is likely to take a few weeks to write up.

## 4.2 The SNO+ Detector

This is the main detector chapter. I will start with an overview of the high-level detector overview, covering its geometry and standard coordinate axis. I will endeavour to explain why certain design choices were made that enable the experiment. This will include a section on how the liquid scintillator works, along with Cerenkov light.

After the light from a physics event has been generated, it must then be detected. I will go into the details of the detector's TDAQ system: the PMT and concentrator design, how events trigger the detector, how information about raw PMT hit time and charge is then obtained via the electronics, and the building of an event from this information.

With this raw data, we must perform low-level calibrations to clean the data. These include electronics and timing calibrations, the latter via the Laserball and TELLIE subsystems. PMT hardware and software checks are also made for data quality purposes (possibly described in more detail in an appendix).

Calibration of the detector optics is needed to obtain an accurate model of our detector. This is done with a suite of optical calibration tools: SMELLIE, AMELLIE, and the Laserball. Calibration of our event reconstruction is achieved with a variety of radioactive sources, such as the AmBe and N16 sources.

This leads into approaches to event reconstruction themselves: a brief description of how our energy, position, and direction fitters work, given that their results are being used for my solar analysis. I will also briefly discuss RAT, the collaboration's simulation and data analysis software.

Finally for this chapter, I will describe the main experimental phases of the experiment, along with the physics plans for each (or what was achieved, as appropriate).

Because of my work in SMELLIE calibration, as well as numerous shifts for detector monitoring, and being on-call as a so-called "detector expert", I have gained a fair amount of experience with the low-level details of the experiment to be described in this chapter. Therefore, beyond e.g. chasing down some of the subtleties of how certain calibration systems work, this chapter should not take up too much of my time. Therefore, it is likely to take a few weeks to write it up.

## 4.3 Optical Scattering Theory

Before we discuss calibrating the detector's optical scattering, we must first understand what it is as a physical phenomenon, and why it is important to understand. I will start by showing the impact of scattering on physics analyses, e.g. the influence on the hit-time spectrum.

Then, I will move into the theoretical models for optical scattering in a fluid. We start with the Rayleigh scattering theory, which is devised for gases. The difference between Rayleigh and Mie scattering is discussed. This is expanded by the Einstein Smolokowsky density-fluctuation theory for materials. I will then look at how existing measurements of scattering compare to this theory, looking specifically at water and scintillator. Any limitations of the theory will be mentioned here, notably the impact of hydrogen bonding in water.

Finally, I will describe how scattering is modelled in RAT for both the water and scintillator, and also the extent to which theoretical uncertainties described above impact my analysis.

I will need to do some further background reading on some of the theory to be described here before writing up. Hopefully though this chapter can be written up in a few weeks.

## 4.4 The SMELLIE Calibration System

I will start this chapter by giving an overview for how the SMELLIE calibration system works: firing collimated laser light into the detector to observe scattering. Analysis is focused on measuring this scattering, ideally the double differential scattering cross-section as a function of wavelength and scattering angle. In addition, as the detector's optics changes through different phases of operation, SMELLIE can be used for monitoring purposes.

I will then give a description of the hardware used for SMELLIE. This includes the laser heads (and the difference between them), and how light travels from them into the detector via various pieces of apparatus. Alongside the path of light is the triggering system: I will go over how the lasers get triggered, and then how the triggers then get sent onto the detector-wide triggering system. If the new hardware was able to be installed by myself and colleagues in the coming year, I will also give a description of these new tools, and why they were valuable to include.

All of this hardware is controlled by software, and I will cover how this works at a high level. I will also cover how commissioning of the SMELLIE system works, and why it is necessary.

I have a very strong understanding of this calibration system, so much of the writing of this chapter should be straightforward. However, if installing the new SMELLIE hardware becomes possible that will likely take up to a month to complete. Commissioning of SMELLIE has already been achieved when the scintillator fill was completed, but whenever the optical phase changes notably (e.g. after the addition of bisMSB into the detector), and certainly after any hardware upgrades, this commissioning will need to happen again. Thanks to an updated commissioning procedure, if we need to do this again it should hopefully be a faster process. The writing of this chapter should take about a month, given the need to talk about commissioning.

## 4.5 Simulating SMELLIE Events

This is the content covered in the main body of this report. Some final work need to be added to this section - one fibre's beam profile still needs to be generated, for example. This should take at most another month of time. Some details will be further expanded in the thesis version of this chapter, such as how a trigger jitter in the detector's electronics during the water-phase required using a certain ' $t_2$  correction' method. Given that this report constitutes a first draft of this chapter, it will hopefully take less time to write this chapter in-place: possibly only a couple of weeks.

## 4.6 Scintillator-phase Scattering Analysis

### 4.6.1 Extinction-length Analysis

A significant discrepancy between MC and data was observed in the number of photons observed at different PPO concentrations by others in the collaboration. This seems to be because of a change in the scintillator's extinction length from expectations at short wavelengths. This can be measured in-situ by SMELLIE! I will go over the details of how one could theoretically make this measurement of the scintillator's extinction length as a function of wavelength. Here is where any assumptions made in the subsequent analysis will be noted, notably that the scattering length of the external water has remained stable over a period of years.

Core to this analysis is a set of time cuts for sets of certain regions of PMTs. Because of the trigger jitter problem noted above, determining the time residual of a hit is somewhat different between the water and scintillator phases. The selection of these regions is discussed here also. Then, a dry-run of the analysis will be performed on MC at a variety of extinction lengths, demonstrating the approaches' validity.

This approach will then be applied to actual SMELLIE data; I will show the conversion from raw hit information into extinction length values. Then, the set of data quality cuts made will be outlined. The possible origins of uncertainty will be explained, and how statistical uncertainty has been propagated onto the final values. Using MC as a guide, I will make a correction for the difference in time residual calculation between scintillator and water. The final results will be shown, and

comments made: how does it compare to the existing optics model, and other measurements that were made of the scintillator? What other conclusions can be made?

Much of this analysis has already been achieved. The MC dry-run, time residual difference correction, and final analysis of the data has yet to be performed. I expect this to take at most a month more of time for research, and a couple weeks for writing up.

#### 4.6.2 Internal Scintillator Analysis

This chapter is extending analyses started by previous students Esther Turner & Krishanu Majumdar. I will start by summarising briefly their analysis approaches, namely that of finding scattered light within a triangular region of the time residual vs.  $\alpha$  space of hits for a SMELLIE subrun. An example plot will be shown made from tracked MC, where I will go through the salient features of the plot, and why this leads to the existing analysis approach. I will also explain its challenges: dependence on lots of MC, and the major systematic of beam profiling. Scintillator analysis has a major additional challenge compared to their work on the water-phase, namely that scintillator re-emission is a major background to scattered light.

As a start, I will run a mildly-updated version of the existing analysis process on the water-phase data, now using the new SMELLIE generator and beam profile. From this, I will be able to obtain updated scattering length values for the detector water. I will obtain uncertainties in this result, especially from the beam profiles. Even though the water-phase is finished and analyses from it almost fully-complete, for SMELLIE it is a useful starting point before moving onto scintillator.

As proof of this fact, I will then try the exact same approach on scintillator, just to see how well it fares. The likely scenario is that it doesn't work nearly as well, due to all of the scintillator re-emission! Our best hope for discrimination between scattering and re-emission of photons is the difference in their angular distributions: re-emission is isotropic, whereas Rayleigh scattering very much isn't. We can try to observe this difference by attempting to reconstruct the paths of photons that are scattering candidates. Former project student Chloe Ransom developed a geometric method for this reconstruction, which I will take advantage of. I will obtain reconstructed scattering position and angle plots for MC, split by the possible optical processes. Hopefully, we will observe a separation between internal scattering and re-emission that can be taken advantage of, and hence lead us to a more effective method for scattering length calculation. With this new method in hand, I will apply it to the MC example I have already shown to derive its scattering length.

Given this new method, I will then apply it to a scintillator data set, and see what results are obtained in terms of scattering length. I will talk about where the major sources of uncertainty arise from, and how this fits into the extinction length measurements made of the scintillator. I will finish by discussing further improvements to the method that could still be made.

Much of this analysis has yet to be done. I have already run a form of the existing analysis on water-phase data, but am getting results I don't fully understand and suspect will need to debug. Applying the existing method to scintillator has not been performed yet. Some of the infrastructure for scattering photon path reconstruction has already been implemented, but not yet tested. Certainly the new method has not been developed yet or implemented. All of this research is liable to take possibly 5 months of time to complete to a satisfactory standard. Writing up this section will likely take a month to complete.

### 4.7 Solar Neutrino Oscillation Analysis in Scintillator Phase

#### 4.7.1 Sensitivity Study

I will start by reminding the reader of how a solar neutrino oscillation measurement works in general: one observes solar neutrino events in the detector at different energies, and the solar oscillation parameters will change the expected shape and magnitude of the events seen as a function of energy.

Next, I will give an overview of the work done by former colleague Javi Caravaca on this topic for SNO+: a background-free sensitivity study over 5 years of scintillator, keeping the solar flux fixed.

Javi also combined this result with similar results from the collaboration’s complementary reactor anti-neutrino oscillation analysis.

Given this context, I will go over the major improvements that are important for both obtaining a more realistic sensitivity study, but also a conversion into an actual analysis on data. I then discuss all of the major backgrounds relevant to this analysis, and their production mechanisms.

Once described, I will start explaining the low-level approach to the analysis: using production MC of all the relevant background and signal processes, with initial expectation rates for each process. I will make a note here of how the constraint on the  $^8B$  solar flux has a major impact on the final analysis, an indication of further discussion that will be had shortly. Then, I will explain the various cuts chosen to minimise background: energy, fiducial volume, BiPo and external classifiers, etc. From this, the final post-cut PDFs for each process are obtained, as a function of energy and position.

From this, I will explain how Asimov data sets are made, and how the next stage of the analysis will work: the 2D oscillation parameter space is scanned over, and at each point the MC PDFs are fit to the Asimov data by varying their normalisations. From each fit a maximum log likelihood is obtained, which can then be plotted for each point of the 2D parameter space. One then obtains 2D likelihood contours from this, which can be different based upon whether one uses a Frequentist or Bayesian approach. By marginalising or profiling this likelihood space, one leads to estimates of each of the oscillation parameters, with associated confidence/credible intervals as appropriate.

Given all of this setup, I will then show the results under different assumptions: varying the solar flux constraints being used, background levels, and the impact of varying cut points. The detector phase will also likely have an impact: namely how different PPO levels and the inclusion of Te-phase data might impact results. I will endeavour to see the impact of the energy scale systematic uncertainty on results.

Much of the software framework required for this analysis has already been written; that being said, more will need to be made. I have already confirmed how a basic background-free study is impacted by the solar flux constraint. Handling the backgrounds sensibly of course will take time: I expect this section will take up to 6 months of research time to complete, followed by a month of writing up.

#### 4.7.2 First Analysis with Scintillator Data

I will start this section by discussing the data set to be used for this analysis, noting the work that the Run Selection working group does in this regards. Following this, I will make some basic data quality checks, and then describe the data cleaning that is necessary in an analysis on actual data: things that are known not to be modelled in MC. Once those checks have been made, I describe the final production MC being used to produce PDFs for this analysis, and the final cuts chosen. If any blinding procedures have been decided to be added, I will describe them here also.

Unlike in MC, another subtlety of analysis on data is the calculation of livetime, especially given cuts such as the muon veto which remove detector livetime. I will then also explain the actual rates measured for certain background processes to be used as fit constraints, performing side-band studies where relevant.

With all of that out of the way, the machinery should be in place to perform the analysis on the actual data. I will show the results, and discuss to what degree they are sensible. If there is time, I will perform a preliminary combination of the solar and reactor anti-neutrino oscillation analyses, to see how they combine.

I will finish by summarising the work done in this analysis so far, and what work remains (in addition to the taking of more data!) in improving the analysis. I should be able to predict how better backgrounds and more live-time improve our results.

Most of the software infrastructure in this section is being used wholesale from the previous one, for obvious reasons. As a result, the challenges here will be all data-specific ones: applying the data-specific cuts, handling any data-MC discrepancies that may crop up, and analysing the results. I expect this to take about two months of my time, followed by a month of writing up.

## 4.8 Conclusions

This is the final chapter, and will be a short one. It will summarise briefly the work done in the thesis, now with the understanding that the reader will likely have read most of it! This is subtly different from the abstract of the thesis. Of course, until I have written the rest of the thesis, I cannot really write this section properly, but once the rest is complete both this and the abstract can be written within a day or so!