Statistical methods for Direct Dark Matter Searches with LZ



Joshua Robert Green

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2025

Acknowledgements

I would not be who and where I am without my entire family, my parents instilling my values whilst giving me space to be my own person, and my sister for being my best friend and confidant. I would like to acknowledge my uncle Phil, we lost you at the start of my joinery but I have always carried your character with me - maybe a tad too much at times. To my close friends, enduring and new, for putting up with my ramblings when I emerge from time to time. The LZ experiment is full of supportive and kind people who have been key to the completion of this work. To name a few who have been particularly supportive; Prof Kevin Lesko, Dr Amy Cottle, Dr Sam Eriksen, and Dr Billy Boxer. I would be remiss not to mention Dr Robert James for introducing me to the main topic of this work, FlameNEST, and realising my contributions to the statistics group. Although I will never be sure whether the former was confidence in my abilities or penance for my behaviour in South Dakota. You have all been good friends and mentors. Here at Oxford, there has been abundant support from staff, academics, and fellow peers. Dr Sparshita Dey started this journey with me and I honestly do not think either of us would have lasted without the other.

Thank you to my supervisor, Prof Kimberly Palladino , for the opportunity provided and your trust in me. In addition to my research, you have always supported me in pursuing community building and sustainability interests. Your efforts to clue me in to the high level decision making in academia and difficulties of the life of an academic will no doubt be valuable insight for my future.

Thank you to those I met on my travels for the unique experiences. I have had my first thanks giving, very long karaoke nights, traditional hot pot on Chinese new year, and learning scopa in Wollongong. As the world seems to become more divided, it is important to note that I have never failed to find kind and curious people wherever I have been. In particular, I would like to thank the SURF staff, technicians, and Lead locals for looking after us car-less Brits during my time there.

My Gemma, I could say all the things like you're my bedrock or a fountain of unending support, and it would be true. But in reality you are my partner in life, we share our highs and lows, and I am so very lucky to have found someone as patient, kind, and gentle as you.

Abstract

> ...we may search for rules to govern our behaviour with regard to them [hypotheses], in following which we insure that, in the long run of experience, we shall not be too often wrong.
>
> Jerzy Neyman 1933 page 291 in Ref.[1].

By mass, dark matter is the largest unsolved problem in astrophysics and cosmology, compromising 85% of the matter in the universe. Invisible to all current probes, except those that leverage gravity, one of the leading hypotheses is that it is made up of an additional particle, or sector of particles, which bridges these fields with particle physics which is also thought to be incomplete. Of a zoo of possible candidates, the LUX-ZEPLIN (LZ) detector searches for Weakly Interacting Massive Particles (WIMPs). The interactions of these particles have a current world-leading upper limit of 5.1×10^{-48} cm^2, set by LZ, which translates to less than 3 events in the LZ's 5.4 tonne fiducial volume a year.

For such rare processes, exceptional care must be taken in: the detector design and construction to minimise background radioactivity and noise, in the processing of data to exclude any remaining noise, determination of remaining backgrounds, and modelling of physics and detector processes. This is the culmination of hundreds of peoples' work and requires a robust statistical approach to carry out inference to quantify a result, be it evidence or exclusion. As part of the statistics team, the author was key to designing, implementing, and validating the likelihood model for the world-leading 2024 WIMP search of LZ using traditional Monte-Carlo (MC) tools to evaluate the physics and detector models. These tools are limited in the dimensionality and type of parameters they can use due to scaling of MC with dimensionality.

An alternative tool to MC, `flamedisx`, allows for the use of multiple dimensions without excessive computation. The implementation of the physics models for LZ caused performance issues that the author resolved. As a robust test of the tool, the author then implemented it for LZ's 2024 WIMP search model, leveraging new observables and shape varying parameters. The computation is of the order of 20 minutes per-fit, far exceeding the performance of MC, and is suitable for asymptotic inference. However, non-asymptotic inference necessary in rare event regimes would require excessive computing resources for the 2024 WIMP search or certain identified approximations. The non-asymptotic 2025 low mass WIMP search is computationally feasible and the asymptotic approximation may be applied for low energy ER searches that are more background dominated.

Contents

1 Dark Matter **1**

1.1 Evidence For Dark Matter . . . 1

1.1.1 Clusters . . . 1

1.1.2 Galactic Rotation Curves . . . 2

1.1.3 Cosmic Microwave Background and ΛCDM Cosmology . . . 3

1.1.4 Modern Observations . . . 6

1.1.5 Gravitational Lensing and Cluster Collisions . . . 8

1.2 Paradigms of Dark Matter . . . 9

1.2.1 Modified Gravity . . . 9

1.2.2 Primordial Black Holes . . . 10

1.2.3 Particle Dark Matter . . . 11

1.3 Particle Dark Matter Candidates . . . 11

1.3.1 Neutrino and Sterile Neutrinos . . . 14

1.3.2 QCD Axion and Axion-Like Particles . . . 14

1.3.3 Weakly Interacting Massive Particle Candidates . . . 15

1.3.4 Astronomical Probes to Dark Matter . . . 15

1.4 Searching for particle Dark Matter . . . 17

1.4.1 Collider Production . . . 17

1.4.2 Indirect Detection . . . 19

1.4.3 Direct Detection . . . 20

1.5 WIMP Direct Detection Basics . . . 23

1.5.1 Scattering Rate . . . 23

1.5.2 Sense of direction . . . 26

1.6 Conclusion . . . 27

2 The LUX-ZEPLIN Experiment **28**

2.1 The Detector and Operations . . . 28

2.1.1 The Time Projection Chamber . . . 29

2.1.2 Circulation System . . . 30

2.1.3 Vetoes: Skin and OD . . . 33

2.1.4 The Xenon Target . . . 33

2.1.5 Data Processing . . . 34
2.1.6 Calibrations . . . 35
2.2 Modelling Physics and Detector Response . . . 37
2.2.1 Microphysics and NEST . . . 37
2.2.2 `LZLAMA` and `BACCARAT` . . . 49
2.2.3 Calibrating the Model . . . 50
2.3 Backgrounds . . . 52
2.3.1 Activation of Xenon and the Electron Capture Model . . . 52
2.3.2 Xenon Long Lived Isotopes . . . 53
2.3.3 Neutrino Backgrounds . . . 54
2.3.4 Diffuse Contaminants . . . 55
2.3.5 Material Gamma and Neutron Backgrounds . . . 57
2.3.6 Accidental Backgrounds . . . 58
2.4 Further Analysis Considerations . . . 58
2.4.1 Analysis Cuts and Data Quality . . . 58
2.4.2 Bias Mitigation and Salting . . . 60

3 Statistical Inference **61**
3.1 Frequentist Hypothesis Tests and Confidence Intervals . . . 61
3.2 The Profile Likelihood Ratio and Power . . . 63
3.3 Combined and Extended Likelihood Function for Counting Experiments . . . 65
3.4 Practical Tests and Confidence Intervals . . . 66
3.4.1 Sensitivities, Brazil Bands, and Power Constraints . . . 66
3.4.2 Asymptotic Formulae . . . 67
3.4.3 Non-Asymptotic Monte Carlo Methods . . . 68

4 Explicit Likelihood For Physics Models **71**
4.1 Why go beyond Monte Carlo? . . . 72
4.2 Flamedisx method and `flameNEST` . . . 74
4.2.1 `flamedisx`: Explicit Likelihoods vs Monte Carlo . . . 74
4.2.2 New Extended Log-Likelihood Function . . . 77
4.2.3 `flameNEST`: using the NEST and `LZLAMA` Models . . . 78
4.2.4 Rate Estimation . . . 79
4.2.5 Summary . . . 80
4.3 FlameNEST model optimisation . . . 81
4.3.1 Quantifying Performance . . . 81
4.3.2 Performance of `flameNEST` . . . 82
4.3.3 The Solution: Reduce Degeneracy . . . 84
4.3.4 Quantifying Solution Performance . . . 85
4.4 Testing with Example Implementation . . . 87

4.5 Future Development and Implementation . . . 92
4.6 Summary . . . 94

5 Statistical Inference for LZ in WS24 **95**
5.1 LZFlameFit: Inference Package . . . 95
5.1.1 Modified `flamedisx` Implementation . . . 95
5.1.2 Goodness of Fit . . . 97
5.1.3 Fit Parameter Summary . . . 98
5.2 WIMP Search 2024 with MC Templates . . . 98
5.2.1 Combined Likelihood Spaces of WS24 . . . 99
5.2.1.1 Veto . . . 100
5.2.1.2 Mixed and Isolated Flow States . . . 101
5.2.1.3 Radon tagging . . . 102
5.2.1.4 The combined likelihood . . . 103
5.2.2 Modelling Signal and Backgrounds . . . 104
5.2.3 The ^{124}Xe Model and Template Morphing . . . 109
5.2.4 WS24 Template Likelihood Pull Tests . . . 113
5.2.5 Salted WS24 Results . . . 114
5.2.6 Un-salting WS24 and Including WS22 . . . 119
5.2.7 Cross-Checking Unsalted Result . . . 124
5.2.8 Constraint on Xe124 DEC Box Parameter . . . 126
5.2.9 Summary of Template Based WS2024 . . . 129

6 Towards Full `flameNEST` in LZ Inference **131**
6.1 Configuration . . . 132
6.1.1 Detector Response . . . 133
6.1.2 Physics Modelling . . . 134
6.1.3 Analysis Cuts . . . 137
6.2 Including Position Dependencies . . . 138
6.2.1 Spatially Uniform Deposits . . . 138
6.2.2 Spatially Varying Deposits . . . 140
6.3 Comparing `flameNEST` Simulations to Template Models . . . 148
6.4 Time Dependence . . . 152
6.4.1 Including Live Time Losses . . . 154
6.4.2 The Combined Likelihood for WS24 in `flameNEST` . . . 155
6.5 Multi-Dimensional Goodness of Fit . . . 156
6.6 Technical Considerations . . . 163
6.6.1 Re-Introducing Source Groups . . . 163
6.6.2 Rate Estimators . . . 164
6.6.3 Optimiser Considerations . . . 167

6.7 WS2024 Full `flameNEST` Pulls . . . 168
6.8 Improvements in Discrimination . . . 171
6.9 Performance and Applicability . . . 173
6.10 Summary of Full `flameNEST` in WS2024 . . . 175
6.11 Outlook . . . 176

7 LZ Projections **177**
7.1 Changing LZ Drift Field . . . 177
7.2 LZ Run Duration . . . 181
7.3 Mid-WS2024 ^{8}B CEvNS and WIMPs Prediction . . . 186
7.4 Summary . . . 188

8 Conclusion **189**

A FlameNEST Upgrades **190**
A.1 Detailed Calculation . . . 190
A.2 Non-Uniform Microphysics Case . . . 192

B LZFlameFit Package **193**

C Rates in Sensitivity Studies **194**

Bibliography **197**

List of Figures

1.1 Rotation curves of the NGC6503 galaxy taken from Ref.[7]. The solid line is the best fit radial velocity distribution, the dashed is stars and other luminous matter, the dotted the interstellar gas, and dot-dash is the inferred distribution of additional matter, or dark matter. 3

1.2 CMB power spectrum, the best fit to data agreeing with standard parameters, and several configurations of the dark matter fraction that conserve the total matter density. Generated using CMBAnalysis from Ref.[9] that uses Planck data from the 2018 release [8]. The additional configurations were created for this work using the CAMB software package [10]. The scale of the 75% dark matter universe is not shown. 5

1.3 Cosmological parameters that are best fit to the Planck CMB data [8]. The pie chart gives a visual representation of the proportions. Recall Ω_m=$\Omega_b + \Omega_\chi$. . . . 6

1.4 The matter distribution inferred by X-rays emission, as observed by the Chandra observatory, and that from weak-gravitational lensing. The X-rays are in yellow and weak-lensing red for the cluster merger MACS J0025.4-1222 taken from Ref.[15] (left). The distribution of X-ray emitting plasma with inferred distribution from weak-lensing for the cluster merger 1E0657-558 or "bullet cluster" taken from Ref.[16] (right). 8

1.5 Various constraints on the dark matter mass fractions of primordial black holes from different experiments, taken from Ref.[30] generated using the `PHBounds` package in Ref.[31] , 95% CL appears to be the standard. Dynamical limits are mostly accretion injecting energy into plasma impacting CMB observations. . . . 10

1.6 Diagram of particle dark matter candidates from the cosmic frontiers report of SNOWMASS 2021, taken from Ref.[38]. The y-axis is an indication of interaction strength with matter and the x-axis shows the candidate mass, both logarithmic. Red shades indicate wave-like particles and blue particle-like. The light to dark shaded regions show the gains from experiments proposed over the 20-year remit of plans of the group. 13

1.7 Plot of the modern day matter power spectrum with different theoretical dark matter candidates taken from Ref.[48]. The data points are from the combined study in Ref [13]. 16

1.8 Diagram of interactions between dark matter and the standard model. The circle represents a coupling between the dark matter and standard model species. . . . 17
1.9 Example of a quark-antiquark collision from protons that would produce a detectable dark matter signature (left). Limits from a search for said process in CMS compared to direct detection experiments, taken from Ref.[51]. 18
1.10 Limits on the annihilation cross-section for $\chi\bar{\chi} \rightarrow b\bar{b}$ taken from a 2017 overview of indirect detection in Ref.[59]. The grey bar indicates the range of cross-sections consistent with the WIMP thermal relic density. The magenta lines were digitised from a 2026 combined analysis of Fermi-LAT, HAWC, H.E.S.S, MAGIC, and VERITAS data on dwarf spheroidal galaxies, the two dashed lines are separate assumptions of the DM density distribution profile [60]. The confidence level is not stated but 95% confidence level seems to be the standard, with the exception of ICECUBE at 90% [61]. 20
1.11 Diagram of the collectable channels from a scatter and examples of technologies that utilise them. 21
1.12 Plot of current and future limits from direct detection spin-independent dark matter search experiments generated using code by Ciaran O'Hare at Ref[82]. Many limits are found in the SNOWMASS version for direct detection review[83], the author added the best-case scenario for CYGNUS-100k m^3 limits from their feasibility study in Ref.[84]. This omits some low-mass experiments discussed that do not project, or quote, limits in standard WIMP-nucleon cross-section, for example SENSEI. The "neutrino fog" shows where nuclear scattering of neutrinos becomes a significant background, discussed in the next section. 23
1.13 Plot of the nuclear recoil differential event rate for several WIMP masses and neutrinos from several sources in natural xenon. The light and dark regions indicate where the LZ 2024 WIMP search obtained $< 50\%$ and $\geq 50\%$ of events detected [90]. Labels indicate why the spectra drops off for WIMPs. 26

2.1 Detector diagram of LZ. The inset is a simple diagram of the detection of a scatter with key dimensions of the forward field region (FFR) between cathode and gate, reverse field region (RFR) between the screen and cathode, and the extraction region between anode and gate. The TPC geometry is such that it is equal in diameter and height to maximise light collection efficiency. 29
2.2 Diagram of the LXe and GXe circulation, adapted from Ref.[71]. 32
2.3 Simple diagram of differences between ERs. a long track assumed in Doke-Birks, and NRs, a high density deposit modelled using the Thomas Imel Box model (left). Diagram of the split between excitation, ionisation, and heat and production of photons and electrons adapted from code by Dr Vetri Velan (right). 38

2.4 The ER and NR simulations for flat energy spectra. Note that ERs require significantly less energy to produce above threshold scatters due to NRs being suppressed by Lindhard quenching. 45
2.5 Diagram of light/charge collection for S1s/S2s. The red arrows intend to indicate that the top-bottom asymmetry of an S1 can be indicative of its position, but this is not simulated in NEST in LZ at the time of this work. 46
2.6 The ER and NR bands for flat energy spectrum sources for the same data as Figure 2.4 but carrying out the above sampling steps. Note that the 3-coincidence requirement is applied but events below it do occur, these would be events in low-LCE regions that have upward fluctuations at the S1 detection steps. 49
2.7 Light collection efficiencies of S2 (left) and S1s (middle, right) in WS22, the S1 maps share a colorbar. Note that high drift times are near the bottom of the detector, where S1 LCE is greatest. 51
2.8 Decay schemes of the two dominant xenon activation products in LZ taken from Ref.[124]. The decay creates a SS produced by the electron capture to an excited state iodine, whose states and subsequent gamma decay are shown. 52
2.9 Diagram of inner shell vacancies caused by an electron capture or neutrino scatter taken from Ref. [124] and Ref. [125], respectively. 53
2.10 Sources of solar neutrinos backgrounds in LZ discussed in the text. The light and dark shaded regions indicate the NR energies with <50% and ≥50% in LZ's most recent science run tuned for WIMP masses greater than 10 GeV/c^2, taken from Ref.[90]. The ERs are only given an approximate indication as the spectrum is flat. Fluxes digitized from Figure 1 in Ref.[130]. Rates were produced in the LZ software DMCalc. 54
2.11 Modified figure of uranium and thorium series extracted from Ref.[132]. The decays relevant to a WIMP search are indicated in black boxes and the rare ^{218}Po decay has been added [133]. 56
2.12 Acceptance curves for WS22 plotted in cS1 (left) and $\log_10$(cS2[phd]) (middle), these encapsulate the data quality and trigger efficiencies. The total efficiency defined as the fraction of NR events surviving all cuts within the FV (right). These were obtained from the FlameNEST code from Ref[111], the uncertainties can be found in the WS22 paper [95]. 59

3.1 Plot of test statistic distribution using asymptotic formula in Ref.[140] and section 3.4.2 (left). Integral of the asymptotic distribution for different test values, μ, and two different cases where there is and is not a significant signal in data, indicated by the value of the best fit to data, $\hat{\mu}$. 63

3.2 An example of sensitivity using asymptotic formula in Ref.[140]. The distribution of observed test statistics under the null hypothesis is shown (left) and that distribution translated into p-values (right). As lift-off in the p-value curve caused by signal-like background can artificially lower the limit, it is important to use the distribution of upper limits. 66
3.3 The non-asymptotic calculation of a confidence interval, which requires the p-value and therefore the test statistic distribution. Noted are some redefinitions specific to the implementation in the dark matter and high energy physics community, particularly in the treatment of nuisance parameters in throwing toys. Outside the calculation of the confidence interval the only p-value of concern is p_0 for discovery. 69
3.4 Non-asymptotically calculating the sensitivities discussed in Section 3.4.1 using the test-statistic distributions from Figure 3.3. The confidence interval is explicitly separated, rather than calculating the distribution of predicted test statistics and evaluating their p-value, the p-value curve for each predicted data set is calculated and then the intercepts are found. This is important in upper limits as a background-only data-set as a lower limit may be caused by background faking the signal, artificially lowering the predicted upper limit. In those cases, the larger intercept is used as an upper limit. 70

4.1 Projections in $S2_{raw}$ and cS1, for a ^{8}B CEνNS source, which is near threshold, using WS22 configuration with no ROI cuts and 3-coincidence requirement. The high value S1 LCE map at the bottom of the detector enhances the number of low cS1 events passing the 3-coincidence requirement. The finite electron lifetime reduces the $S2_{raw}$ of events from the bottom of the detector. Exaggerating the electron lifetime, τ_e, and the S1 total light collection efficiency, g_1, to non-physical values reduces the impact. This shows the interplay of dimensionality and shape-varying parameters required to accurately model this source. 74
4.2 Calculating the probability for a range of signals following equation 4.1 using the method vs MC. The separation at low signal is solved using a continuity correction not applied in this toy model. The legend shows varying parameters of the model and reduced re-computation time of the flamedisx method due to the use of the saved computation graph compared to using MC. 76
4.3 Diagram of separable and inseparable variables where the axis into the page represents the total rate. Also shown is the batching of the grid for deployment on multiple threads. 80

4.4 Example of profiling the differential rates. Heap memory is that which is used dynamically over a program, and here it grows in each batch as the parallel sum over energies occurs. There is a pre-window heap usage, which is caused by tracing the function and can be the largest. 82
4.5 Diagram of the original `flamedisx` models (left) and implementation of the NEST models (right). Double edged solid arrows indicate a dimension of the quanta tensors and dashed arrows on the right indicate a parallelised sum over the energies contributing to an event and contraction over the ion dimension per energy. Blue, green and orange indicate 1,2 and 3 dimensional tensors, respectively. Batching is treated implicitly here. 83
4.6 Comparison of differential rates derived from MC templates and from `flameNEST` differential rate for a mono-energetic 20 keV NR (left) and 5 keV ER (right). The templates used 10^6 simulated events. 86
4.7 Variance of multiple calculations of the differential rate, gradient, and Hessians between calculations before and after the updates. All are compared to a single pre-update calculation. 87
4.8 Impact of varying chosen parameters on the total quanta (right axis, red) and charge yield (left axis, black) within $\pm 1\sigma$ of the default NEST uncertainties. For asymmetric uncertainties, the largest is chosen. The parameters α and β control the total quanta yield where ϵ controls the charge yield 88
4.9 Projection in each parameter of the rate estimator that is a linear interpolation over a 3D grid. Also shown is the gradient with respect to each parameter, by definition the second order gradient, and therefore any contribution to the Hessians, will be zero. 88
4.10 Pull plots of the test-case, each row shows a different parameter. The rate multiplier has a smaller variance as its sample value θ_0 was generated from the size of the data set and rate estimator at sampled values, i.e from the observed and not sampled value. That is from the observation not the prediction. The left column gives the varying MLEs with the value of the anchor points overlaid. The middle shows uncertainty on the MLE derived from the Hessian of the likelihood. The right column gives the pull, which should be a normal distribution. 89

4.11 Performing a fit with the test source using perturbed parameters. The simulated data with contours showing the default, best fit, and true shape for this sample (left). The pull plot with several definitions of the pull (right). For this, the true value of a parameter used to generate the data, θ_{true}, is sampled from a Gaussian distribution centred on the pre-fit value, θ_{exp}, and uncertainty, σ_{exp}. The fit gives the MLE, or post-fit, value, $\hat{\theta}$ and uncertainty, $\hat{\sigma}$. The red hollow points give the perturbation to the data, the stars how close the fit is to the true value, and black points the distance from the expected. The horizontal lines give how much this data set constrained that parameter. 90
4.12 Likelihood scan over the rate multiplier and one of the parameters α. The horizontal line gives the approximate intercept of the 68% confidence interval if the likelihood were Gaussian. This is not a profiled likelihood scan so these do not intercept with the uncertainty derived from the Hessian. In the scan of α there is a kink in the likelihood near the minimum, this indicates that the differential rate is varying in a way that is not captured by the interpolated total rate term. 91
4.13 Timing the calculation of the value, gradient, and Hessian of the likelihood from above test for the entire data set of 650 events with varying batch size. Increasing the batch size may cause some error in the computation shown on the right axis. 92
4.14 Using a modified sample k-mean clustered over many different α values. Rate estimation was then performed using a modified version of Equation 4.4, weighing calculated points by their volumes. Cubic volumes are used except for the non-linear space of S1 & S2 which used Voronoi tessellation. A sample size of 400 points was used and the function normalised to a single MC estimate at default NEST values. The approximation holds for a short range of α and β but breaks down when ϵ, which controls the electron yield and creates a different change in the shape of the model than what is captured in the sampling. 93

5.1 Example of the 2D GOF plots shown in this chapter. This space had an acceptable p-value but the position of deviations can inform us. Without knowing the model we can determine that the event at (a) is difficult to cover and the region (b) is showing some over-fluctuations under the ER band. 97
5.2 The WS22 and WS24 ROIs and 95% contours of various sources. The intensity of the contour diagrammatically indicates the rate of a source in that space, here the WIMP rate is much greater in the WS24 space due to the larger exposure. The beta bands are in separate colors as these backgrounds were uncorrelated. . 99

5.3 The WS24 ROI and vetoed spaces with 95% contours of various sources. The intensity of the contours indicate the ratios of rate between the two spaces. For an efficiency 90% on average 1 neutron in the ROI is accompanied by 10 in the veto, and for a dead time of 3% every 100 WIMPs 3 are observed in the veto. Only the edges of the WIMP contour are added to the veto space to show its negligible contribution. 100
5.4 Mixed and Isolated data sets showing the unique $^{14}CH_3T$ source and that the mixed and isolated radon daughters,^{214}Pb/^{212}Pb, are treated as distinct and uncorrelated sources. 101
5.5 Diagram of the radon tagged and untagged space, as presented in Table 5.2 15% of non-^{214}Pb sources are in the tagged space and approximately 70% of ^{214}Pb. This makes the ^{124}Xe DEC clearer in the untagged space, which is discussed in Section 5.2.3 and Section 5.2.5. 102
5.6 Plots of the various backgrounds with the WIMP signal. Atmospheric neutrinos have a negligible rate, see table 5.2. Both WIMP signals are not filled in the veto to indicate that they have a negligible contribution. 106
5.7 Log likelihood value, its gradients, and second order gradients with respect to the box parameter for different interpolation types on simulated data for 22 and 100 events (top). Pull plots sampling toys from a box parameter of 1.5 and the same number of events (bottom). The splines remove a bias of the optimiser to remain near anchor points, caused by a discontinuous likelihood, and retain nominal behaviour of pulls. 111
5.8 Pull test of the WS24 model, all parameters were sampled from the pre-fit Gaussian constraint terms except from the ^{124}Xe box parameter, WIMP signal, and Detector NRs. The box parameter was sampled uniformly within its bounds and Detector NRs from [0,5]. The WIMP was assigned 20 events, and all sampled counts are subject to Poisson fluctuations inside the likelihood model. These three all show mild negative biases in their pulls, due to boundary effects. Both the box parameter and Detector NRs pulls are negatively correlated with the WIMPs giving correlation coefficients of -0.08 and -0.19, respectively, imposing a small bias. No other parameters show a bias. 113
5.9 Pull plot of the parameters shown in Table 5.3 for which an expected value exists. Only the largest sources, ^{214}Pb isolated and mixed, Flat ER Veto, and radon tagging efficiency, linked to ^{214}Pb, are constrained by the fit itself. There is a significant over-fluctuation in the mixed state of ^{214}Pb and under-fluctuation of the Flat ER in the veto. 116

5.10 The 2D GOF of all spaces in WS24 from a fit using a WIMP mass of 10 TeV/c^2. All spaces pass the p-value requirement. There are strong over-fluctuations in the data below the ER band at high S1c in the untagged space and an apparent anomalous event at S1c $\sim$ 30 phd and $\log_{10}$(S2c [phd]) $\sim$ 2.9. 117
5.11 A profile scan of the likelihood (left) shows that the box parameter absorbs events otherwise attributed to WIMPs (top) and the data shows a preference for a model with approximately 1.7$\times$ greater ionisation density (bottom). The 2D GOF of a fit to data including a WIMP of mass 10 TeV/c^2 (right) with the box parameter fixed to a EC response (top) and floated (bottom). The test is not very sensitive to the box parameter. 118
5.12 Inference results for the combined WS22 and WS24 likelihood with the solid black lines showing the upper and lower limits in terms of WIMP counts vs mass. The median expected upper limit, given the conditional best fit nuisance parameters to data, is the dashed line, with the 1 and 2σ Brazil bands given by the green and yellow shaded regions,respectively. There is an upper and lower limit for WIMP masses $\geq$ 20 GeV/c^2 as there was rejection of the background hypothesis, shown as a function of the tested WIMP mass in red and the right axis, of up to 2σ. Adapted from the statistical analysis note Ref.[160]. 119
5.13 Scatter plot of the data post-unsalting, with the salt-events shown in red stars, for the entire WS24 data not including the veto. Several models are added to show their position relative to the salt. Note previously there are many salt events overlapping with the centre of the AC distribution. 120
5.14 The 2D GOF tests and plots for all spaces with the un-salted data set with a 10 TeV/c^2 WIMP model. Generated using code developed by Leah Wolf and the statistics team, including author. 122
5.15 Derived upper limit and Brazil bands for WS24 with (top) and without (bottom) WS22 included in the combined likelihood. The definitions are identical to Figure 5.12, except there is only an upper limit which is also shown as unconstrained in the gey dot-data. The dotted line is the cross-section required to obtain at least 3σ evidence 50% of the time, calculated the same as the Brazil bands using nuisance parameters from the conditional best fit at each mass. It shows that there is a -1σ under fluctuation at all masses for the WS24 data and that this is enhanced for masses below 100 GeV/c^2 by the inclusion of the WS22 data. This was observed in WS22 and is discussed in the text. These plots were taken from the statistical analysis technical note in Ref.[160]. The curves from other experiments were obtained from Ref.[161, 162, 163, 164, 165, 166]. 123

5.16 Pseudo salt study, injected a fixed number of 6 91GeV mass WIMPs and 1 Flat NR events into the unsalted observed data. The events are distributed randomly over the PDF and spaces appropriately without any Poisson fluctuations, so the number of salt events is fixed. Three configurations were chosen, the default WS24 model and dataset with all spaces, removing the AC constraint, and then removing the anomalous event as well. The left most extent of the histograms indicates the best fit to unsalted data. Caution should be exercised when interpreting a shift in the best fit counts of a background source as evidence for absorbed salt events. The best fit value for the beta component can increase by more than the total number of salt events, since adding events to the tail of the distribution can influence the likelihood function more strongly than the constraint on the total event count. The main result is that the recovered 91GeV mass WIMPs is unaffected by removing the AC constraint as the single event constrains the model such that WIMPs are still absorbed, removing it almost fully restores the recovered salt. 125

5.17 The regions for which the counting experiment was conducted (left). The frequency for which one event is observed in region 2 given a number of events in region 1. 126

5.18 Evaluating the test-statistic distributions assuming the default EC model and conditional best fit to observed data of nuisance parameters, $f(\tilde{t}(b = 1)|b = 1, \vec{\theta} = \hat{\hat{\vec{\theta}}}(b = 1.))$, for the box parameter discovery significance using 68k toys and ∼5k hours CPU time, 50-node hours with 100 tasks per node, with the observed test statistics of the two configurations in grey vertical lines (right). The colors are toys generated assuming a basic and split likelihood, following those in the legend of the plot on the left. The derived significance from the test-statistic distributions vs the expected from asymptotic (left). There is a constant shift in significance as the beta-band in many background toys may cause the maximum likelihood to prefer a box parameter $b < 1$, but the physical boundary forces this to 1. Therefore there are many toys for which the $-2ln\left(\frac{L(b=1.)}{L(b=\hat{b})}\right)$ would be greater than 0 but is forced to zero as $\hat{b}$ is forced to 1. Note that with and without the tagging and splitting of the WS, the distributions are identical but observations in the grey lines are distinct. Combining the distributions and doubling the test-statistics does not significantly increase accuracy, as one would need orders of magnitude more toys to do so. When the observed significance becomes greater than the end of the test statistic distribution, on the left plot, the derived significance plateaus, as seen on the right plot. 128

5.19 The median and $\pm 1\sigma$ band of the distribution of asymptotically derived significances under different box-parameter assumptions and conditional best fit to observed data nuisance parameters, $f(p(b = 1.)|b = b, \vec{\theta} = \hat{\hat{\vec{\theta}}}(b = 1.))$. The observed significances both over-fluctuate above the median, suggesting a reduced background. This would be strongly impacted by the ^{214}Pb rate and distribution, and may be due under-fluctuation in the isolated state in Figure 5.9 which would also explain the increased over-fluctuation of significance when applying the tag here. No corrections are applied for the shift shown in Figure 5.18. . . . 129

6.1 Field map before and after conversion into observed variables with the drift map. The field is assumed to be uniform azimuthally. 134

6.2 Light Collection Efficiency (LCE) maps evaluated for fixed points over the detector for S2 (left) and S1 (middle,right). The S1 LCE is in drift time as the position of an event impacts efficiency, with greatest efficiency at the bottom of the detector, whilst S2s are all generated in the thin extraction region at the top. 134

6.3 Default NEST models from Ref.[113] plotted with various fields, WS22 used these models and an average field of 192 V/cm. The Field for WS24, 96.5 V/cm, is plotted alongside the minimum and maximum drift fields from Figure 6.1. This causes a maximum light/charge yield variation of 0.2/1.1% in NRs and 0.6/4.3% in ERs. 136

6.4 The acceptance curves for WS24 ported to `flameNEST` by the author, the S2 based trigger efficiency is 100% well-before the lower S2 threshold so is omitted here. These are summarised in the background modelling technical note in Ref.[156] using the trigger and cut efficiencies tunings from technical notes [167, 168, 169]. The final NR efficiency within the FV is also shown and compared to that of the WS24 paper in Ref.[90], here the uncertainties from calibrations are omitted. 138

6.5 The approximated gradient of log sum differential rates for a 5 keV$_{ee}$ monoenergetic ER source using 50 events with respect to radius (left). The S2 and true radius dependent width of the Gaussian position smearing applied in `LZLAMA` to approximate reconstruction uncertainties for different S2 sizes in electrons (right). 140

6.6 Spatial distribution of a uniform source from `LZLAMA` with no LCEs applied. The radial (left) and vertical distribution (right) compare ER and NR to show no distinction. The radial distribution is modified correctly to mimic the wall, and the minimum and maximum extent of the FV cut is shown by the grey band. The 2D PDF shows the minimum and maximum extent of the FV cut azimuthally (right). 140

6.7 The discrete 2D Probability Mass Functions (PMF) of the ^{214}Pb decays extrapolating from ^{218}Po sites in corrected position, where z_{gate} is the maximum z position a SS can be detected in the TPC (top). Those positions after being processed through `LZLAMA` (bottom). Unfortunately, `LZLAMA` only supports (r^2_{true},z_{true}) inputs, so the 2D distribution on the bottom is more uniform than the top. Grey regions indicate the minimum and maximum FV boundaries, these are plotted in corrected co-ordinates with a linear approximation for drift time and assuming r_{obs}=r_{corr}. A red marker is where files were incomplete and the author had to copy the bins above, but the population is low in this region. . . . 142
6.8 High energy simulations of detector ERs between ¿1 MeV (left) and low energy simulations (right). Within the volume marked in red, the high energy position map is used and normalised to the low energy counts in this region. 143
6.9 Cutting population of multiple-scintillation single-ionisation events (b) and merged MSs (b) that appear in the simulations (left). Impact of the cut over the total volume (middle). Impact of the cut within fiducial volume (right). 144
6.10 Comparing the KDE estimation to veto sideband with a flat background from template based fits. 145
6.11 Combined simulations of the Detector NR showing change in spectra with an inner-volume cut and between Veto and WS (top). The impact of the veto tag and an inner radial cut indicates how the spectrum may change (top,left). The inner volume used for that check is indicated in the simulated data for the WS ROI (top,middle) and the simulation is also shown for the Veto sideband (top,right). An example of before (bottom,left) and after (bottom,right) the KDE smearing for a slice of the WS ROI over 450 < drift time < 550. 146
6.12 Position distribution of the xenon activation isotopes with and without the veto cut applied. 147
6.13 Comparing final background model from WS24 to `flameNEST` models configured by the author, with true position fixed to the centre of the detector. These are normalised PDFs, with the per-bin Poisson error in `flameNEST` propagated accordingly. The JSD metric does not rely on these uncertainties, it is just for inspection of any shape disagreement. The JSD is above the threshold chosen in text for Detector NRs and WIMPs have a visual disagreement near threshold. . 150
6.14 Projections of the WIMP and Detector NR models from Figure 6.13. 151
6.15 How the underlying spatial distribution of deposits changes the (cS1,$\log_1$0cS2) when assuming the distributions in Section 6.2.2 with bin counts, $N_{spatial}$, compared to the derived observed uniform distribution, $N_{uniform}$. 152

6.16 Plot showing the accumulated livetime of WS24 against calendar and days in. Shaded region shows the nature of data acquisition, for example when $^{14}CH_3T$ calibrations are taking place. Generated using code from Dr Aiham Al Musalhi, PDRA at University College London. 153
6.17 Activity of ^{127}Xe and ^{125}Xe plotted against days into WS24 derived by Olivia Valentino, PhD student at Imperial College London. 154
6.18 Plot of the averaged livetime fraction over WS24 (left). The expected model, i.e pre-fitting, before (middle) and after (right) applying the Bernoulli trial and scaling. There is little statistical power to discern a good-quality fit. 155
6.19 Pull plot of all the parameters in the likelihood model for WS24. Not that unlike in 5.9 these parameters are rate multipliers. These were derived using symmetrical uncertainties derived from the hessian of the likelihood function. . . 158
6.20 The 2D GOF as described in Section 5.1.2 but utilising templates for each space produced using `flameNEST` simulations. 159
6.21 The 1D radial GOF using the asymptotic χ^2_m distribution where m is the number of bins. 160
6.22 The 1D radial GOF using the asymptotic p-value calculation as in Figure 6.21. . 161
6.23 Pie plots of the veto sideband for this fit plotting in various key observables, contours of individual models are also shown (top). Diagram of location of scatters of the highlighted events in the various detectors, taken directly from the SR3 Neutron Efficiencies and Rates technical note in Ref.[173] (bottom). The radial position in the OD is a randomised value for visualisation purposes. . 162
6.24 Comparing calculations between grouped and ungrouped likelihoods for two grouped sources, Isolated ^{214}Pb and ^{136}Xe, with an ungrouped nearby source, ^{124}Xe. Two shared shape varying parameters are floated to check any errors in the differential rate calculation. The Isolated ^{214}Pb was used in case there were any complications in the presence of a spatial source. A normal and grouped likelihood was used then for 400 toys, data was simulated from the normal likelihood then the value, gradient and Hessians were calculated on both at a guess. The log-likelihood values are precise, but the gradients and Hessians (not shown) do vary within 10%. 164
6.25 Example of the total rate estimators, as all source spectra are normalised to a total true rate of 1., this represents a fractional acceptance of the source. The Poisson uncertainty from MC simulation, $\sqrt{N_{MC}}$ where N_{MC} is the number of trials, is shown for the anchor points. The polynomial 1D fits used for this work are shown, where the change in rate is indiscernible due to the uncertainty on MC it is assumed to be flat. 165

6.26 Checking the rate estimators for ^{124}Xe and two WIMPs, one low mass to represent a threshold dominated source and one at high mass. The left column is the difference of assuming separable parameters with linear cross-interpolation and estimating the rate with the source on a grid of points. The right column gives 2D contours of the models to interpret the rate. The bounds of the parameters is $\pm 3\sigma_{exp}$ and within this g_1 and g_1^{gas} are separable for both types of WIMPs, this may not hold for different thresholds or greater uncertainty. There is a dependence between g_1 and the box parameter, b, but this is small compared to the total variation in rate in Figure 6.25. 166
6.27 Fitting a simplified model with a handful of sources and saving the progress of the optimiser. Plotted are the normalised parameters, often to a guess or by their bounds, against the steps in the optimiser with the colour indicating the value of one of the optimiser's parameters. The optimiser wishes to minimise the trust radius, i.e stating it does not need to step far as it has reached a trusted optimum point, and this is seen to be initially minimised but then oscillating with no change in the parameters (left). The normalised gradient of the function with additional terms for constraints, the optimality, reaches a minimum then, alongside the parameters, never changes (middle). The only parameter that is changing after the first 10-20 steps is the barrier parameter, which quantifies how close parameters are to their physical boundaries (right). This may be due to the problem being tightly bound, especially in the shape varying parameters, so the tolerance for the barrier parameter should be adjusted. 167
6.28 Pull plots with and without using grouping, it can be seen that there is a bias in the shape varying parameters g_1^{gas} and g_1 as well as in the dominant ER rates of ^{214}Pb mixed, ^{214}Pb isolated, and ^{85}Kr sources. If these shape parameters are removed the bias in the rates disappear. The tolerances for these toy pulls were 10^{-2}, 10^{-4} and 10^{-2} for the normalised gradient, trust radius, and barrier parameter, respectively. These were informed by fits to observed data and fits to toys on smaller likelihood models, such as those used in Figure 6.27. As in Section 5.2.4, in toys the box parameter and gamma-weight were uniformly sampled within their physical bounds. The mean rate for Detector NRs were uniformly sampled between [0,9] counts in the veto sideband, which corresponds to approximately [0,1.] in the WS ROI. The WIMPs were sampled with a mean rate of 50 counts over the entire ROI. 169
6.29 Maximum likelihood estimators for g_1^{gas} and g_1, $\hat{\theta}$ without using grouping. The parameters are normalised to the constraint centre, θ_0, and width, σ. 170

6.30 The calculated ROC curve, explained in text, comparing discrimination vs signal acceptance curves for various sources using the full `flameNEST` calculation (solid line) to a MC derived 2D template PDF (dotted line). The signal source is always a 10 TeV/c^2 mass WIMP, except for the green line where it is a Solar νER source. The templates were binned in (cS1,log$_{10}$ cS2) using 10^7 MC simulations and the same binning as the WS24 analysis. The Detector NRs and 10 TeV/c^2 WIMPs have a similar spectra and distinct position distribution. The ER sources can be noisy at low values of discrimination, in the future this could be resolved by importance sampling. 172

7.1 Diagram of the flat beta ER background into a 10 TeV/c^2 WIMP model and light-blue shaded region for which an event is determined to be leakage (Left). The leakage is then broken into slices of S1c to see how it evolves and to evaluate the maximum (Right). 179

7.2 Results of `LZStats` inference carried out by Makayla Trask with guidance from the author using field configurations from Table 7.1 assuming 300 live days. The median number of WIMP events required to obtain 3 σ significance, or greater (Left). The median WIMP sensitivity (Right). One line represents 92V/cm and 192V/cm at the SR2 Gate and Anode configurations as they are virtually degenerate. 180

7.3 Change in NR mean yields between drift fields of 192 V/cm and 92 V/cm. The charge yield is shifted by a constant for changing fields. 181

7.4 Fitting the isolated cS1 (left) and isolated cS2 (middle) distributions for a preliminary AC model. The cS2 resolution function used is also shown (right). 183

7.5 The AC and ^{8}B CEνNS model for this sensitivity study. The ER and NR models for the WIMP search have been omitted. The 95% contour of the ^{8}B model contains approximately 20% of the accidentals rate. 183

7.6 Grey vertical lines indicate the test-statistic values for asymptotic 1,2,3 σ significances. The blue histogram shows the test statistic evaluated on background only toys, this should approach the asymptotic purple line. The pink histogram is the test statistic evaluated on the signal and background toys, and pink lines give the median and $\pm 1\sigma$ of the distribution. 184

7.7 Median WIMP sensitivity vs live days, the additional dotted line above the DOE requirement indicates a 30% variation in rate, see text (Left). Median discovery significance for a ^{8}B CEνNS search vs live days (Right). 185

7.8 Plot of the S1c (left) and S2c (right) acceptance curves and projections of three models to show the relative impacts. 185

7.9 Plot of the models considered for this mid-WS2024 study. 186

7.10 Results from the interim WS24 sensitivity study. The expected WIMP upper limits are derived using a combined likelihood with the WS22 models, not observation. The annotations and colors are identical to those in Figure 7.6 187
7.11 Predicted discovery significance of an ^{8}B CEνNS search vs live days accumulated, toys are generated for the quantiles but the significance of each toy is calculated asymptotically (Left). Sensitivity of a WS22 + set number of live days, evaluated non-asymptotically. 188

B.1 The LZFlameFit package, the creation of a likelihood flows from right to left and the use of classes are shown with dotted arrows. The borders indicate where the author made contributions beyond quality assurance, debugging and use. 193

Chapter 1

Dark Matter

As will be shown, dark matter remains one of the greatest mysteries in modern astrophysics and cosmology, filling many gaps in the behaviour of the universe. Whilst alternate paradigms for its origin remain under active study, namely primordial black holes and modified gravity, particle dark matter is a leading candidate. The benefit of particle dark matter, apart from its agreement with observations, is the ability to bring two fields of physics at separate scales, particle physics and cosmology, closer to completion in one step. Since the inception of particle dark matter, the type of particle that could explain it has expanded, with Weakly Interacting Massive Particles (WIMPs) remaining one of the earliest class of candidates and favoured still today.

This chapter aims to motivate particle dark matter, particularly WIMPs, whilst providing insights into the status of other candidates both in the particle and non-particle paradigms. First, the observational evidence for dark matter is presented, followed by a brief discussion of the possible paradigms. For particles, some discussion on the different nature of particle candidates, developing astrophysical probes into this nature, and different theoretical candidates is provided. Finally, a review of particle-like dark matter detection is provided, expanding on direct detection methods to provide context for the LUX-ZEPLIN experiment and this thesis.

1.1 Evidence For Dark Matter

In this section, we will discuss the astrosphysical evidence for dark matter in a semi-chronological manner. In this discussion it is assumed that dark matter is a collisional-less matter distribution, but modified gravity is discussed in the next section.

1.1.1 Clusters

The origin of the term dark matter is often attributed to Fritz Zwicky's first publication on the Coma cluster in 1933, though the term had been used earlier in mapping the distribution of matter within the Milky Way. In both cases it was viewed as dim stars, brown dwarfs, gas or

other forms of not-bright baryonic matter [2]. Fritz Zwicky in his 1933 and 1936 publications studied the Coma cluster and measured the velocities of constituent galaxies by using optical red shift. Then the virial theorem was used to determine the gravitational potential and therefore the mass of the cluster. The virial theorem states that for a stable system bound by conservative forces, the time average of the kinetic energy of its constituent bodies is equal to that of their potential energy. Using this, he was able to derive the mass of the cluster from the velocity dispersion of the galaxies within it. Doing so yielded a requirement for far greater mass in the cluster than was inferred from the luminosity in each galaxy, leading to the suggestion of dark matter in 1933 in Ref.[3]. Inferring the implications for the mass-to-light ratio of each galaxy and correcting for the inflated Hubble constant at the time yields a ratio of $\sim$8 in 1937 in Ref.[4], far greater than what was expected at the time[1] [2]. At the time the assumption was that this was dim baryonic matter or that the clusters were not stable and that this peculiar result was from passing galaxies and the notion of dark matter was not widely adopted[2].

1.1.2 Galactic Rotation Curves

The adoption of dark matter as a systemic misunderstanding of the dynamics or matter composition of galaxies began after the observation of rotation curves out to larger radii. The first observation of this kind was made by Ford and Rubin using optical redshift to measure the rotation curves of the M31 galaxy , observing a flatness in 1970 in Ref.[5]. This was supported by earlier radio studies of hot gas by Roberts observing flatness of the velocity distribution out to larger radii [6]. A suite of studies in the 70-80s confirmed this flat rotation curve phenomenon in a range of galaxies [2].

The rotation curve of a galaxy can be used to measure its mass through gravitation by the simple relation of

$$v_{clr} = \sqrt{\frac{GM(r)}{r}} \tag{1.1}$$

for circular orbits, where v_{clr} is the orbital velocity, G the gravitational constant, and r is the radius of the orbit. Through measurements within a galaxy of the distance and inferred orbital velocity of stars and gas, one can compare different assumptions for the distribution of mass $M(r)$. This is shown for galaxy NGC6503 in Figure 1.1, where instead of peaking and decaying as inferred from the distribution of stars and gas, the velocity distribution flattens. This implies a spherical distribution of matter quite different from the observed matter, called a halo, with $M(r) \propto r$ and therefore density $\propto 1/(r^2 + c)$ where c defines the profile. In contrast, luminous matter appears to have a near-constant density within the visible galactic disk that rapidly drops off outside of a core.

[1] Though not far off modern assumed ratio of 5 for a homogenous universe assumed today.

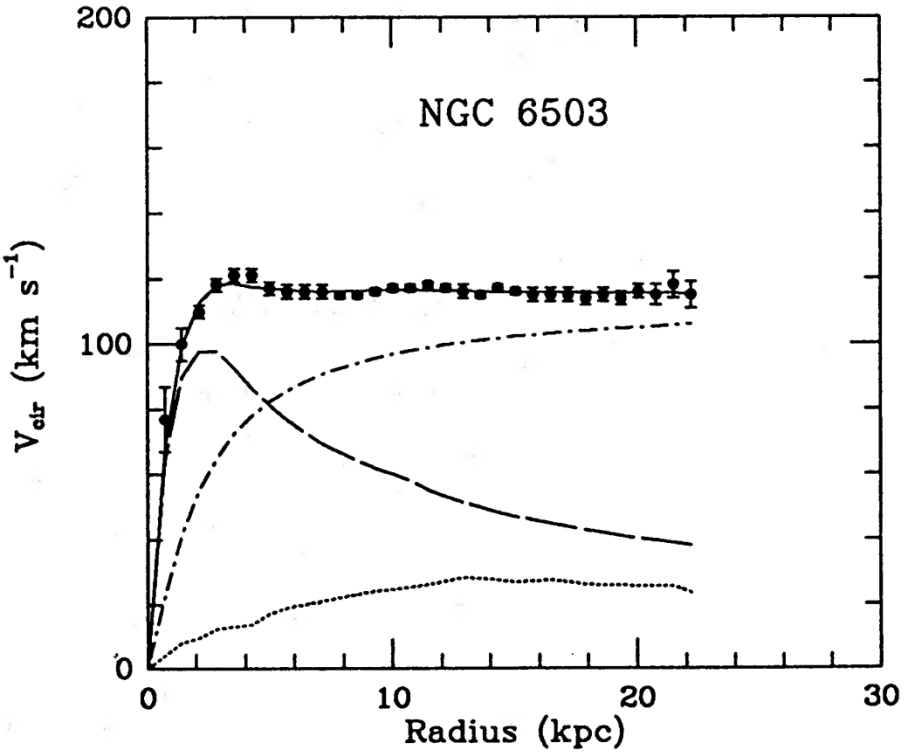


Figure 1.1: Rotation curves of the NGC6503 galaxy taken from Ref.[7]. The solid line is the best fit radial velocity distribution, the dashed is stars and other luminous matter, the dotted the interstellar gas, and dot-dash is the inferred distribution of additional matter, or dark matter.

1.1.3 Cosmic Microwave Background and ΛCDM Cosmology

The expansion of the universe can be modelled using the Friedmann-Robertson-Walker (FRW) metric, Einstein field equations, and assuming an isotropic and homogenous universe. The results for the scale factor of the universe, $a(t)$, are given by the Friedmann equations which may be expressed in terms of the dimensionless density parameters of different species,

$$H(t)^2 = \left(\frac{a'(t)}{a(t)}\right)^2 = H_0 \left(\Omega_r \left(\frac{a_0}{a(t)}\right)^4 + \Omega_m \left(\frac{a_0}{a(t)}\right)^3 + \Omega_k \left(\frac{a_0}{a(t)}\right)^2 + \Omega_\Lambda\right) \tag{1.2}$$

where $\Omega_{r/m/k/\Lambda}$ indicates the dimensionless density for radiation, matter, curvature, and dark energy, respectively, at time t_0 for which $a_0 = a(t_0)$, typically taken as today. The matter component further breaks down into baryonic and dark matter, $\Omega_m = \Omega_b + \Omega_\chi$. Each component's term, apart from curvature, is proportional to its physical density at that time, meaning the different scale factor dependencies show how it evolves. Intuitively, the matter density evolves with the volume of the universe and the same is true for radiation, but the red-shift as the wavelength expands must also be accounted for. By definition, dark energy must be a constant, giving the rate at which two objects recede from each other depending on their distance $a'(t) = \sqrt{H_0\Omega_\Lambda}a(t)$. Finally, the curvature component does not behave as an energy density and has a unique scaling.

One of the most robust measurements in cosmology is that of the Cosmic Microwave Background (CMB). The CMB was formed at recombination time, when electrons and light nuclei in the plasma of the baryonic matter of the early universe combined, allowing photons to move freely. Over the approximately 13.2 billion years that followed, the universe expanded red-shifting these energetic photons into microwaves. Contemporary observations by terrestrial and space-based microwave satellites map anisotropies in the CMB. For interpretation, these measurements are then mapped onto a spectral decomposition using spherical harmonics, shown in figure 1.2 for Planck data [8]. The harmonic multipole, ℓ, is inversely proportional to the scale of an anisotropy in the sky; e.g $\ell = 0$ represents a total average, and $\ell = 1$ would be a dipole, if half the CMB were hotter than the other. The amplitude D_ℓ^{TT} presents how prominent that scale anisotropy is in the sky.

Up to recombination the baryonic plasma would oscillate, as local over-densities attracted matter and then radiation pressure caused them to expand, called Baryon Acoustic Oscillations (BAO). The most prominent anisotropy is the first peak at $\ell \sim 228$, representing the largest possible compression given the causal horizon at the time and the speed at which a wave could propagate in the plasma. The flat spectrum observed at larger scales represents parts of the sky that were not causally connected at the time, so their perturbations could only be random noise. The subsequent peaks after the first represent more oscillations, so the second would result from the expansion of the first, the third the subsequent compression after the second.

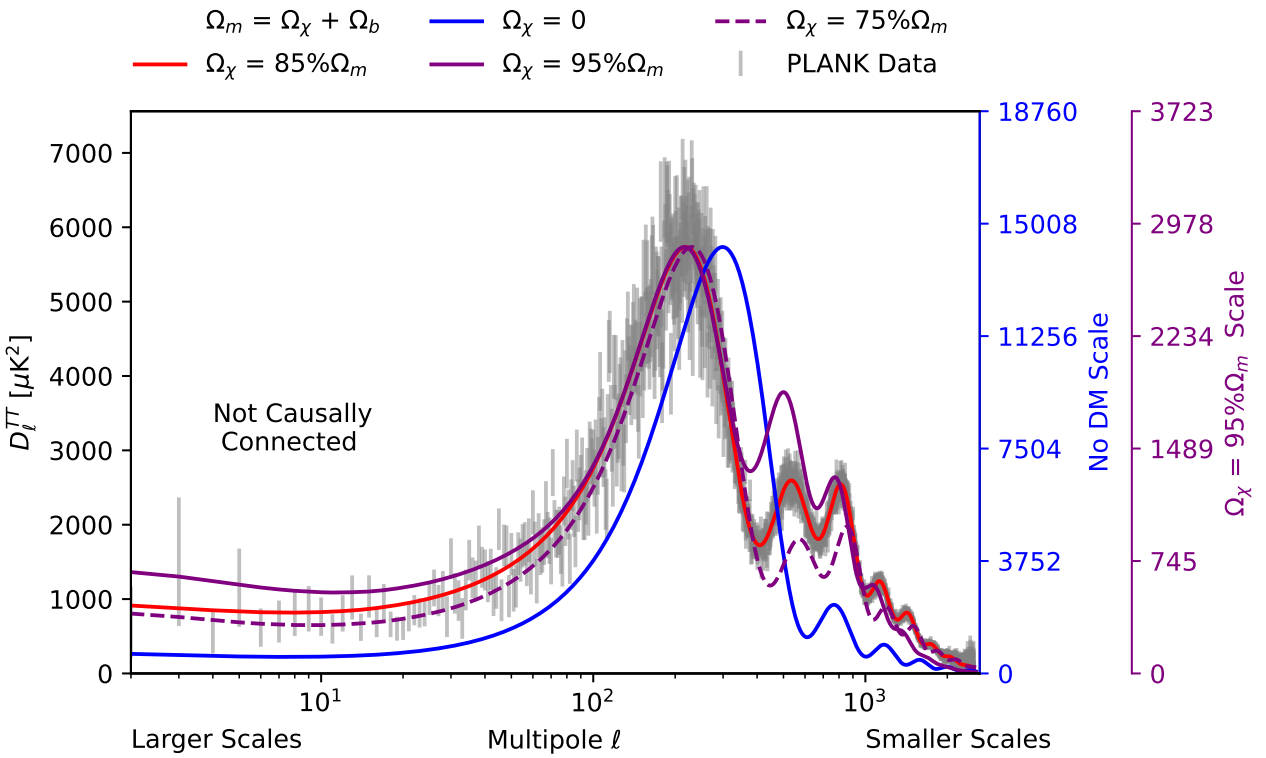


Figure 1.2: CMB power spectrum, the best fit to data agreeing with standard parameters, and several configurations of the dark matter fraction that conserve the total matter density. Generated using CMBAnalysis from Ref.[9] that uses Planck data from the 2018 release [8]. The additional configurations were created for this work using the CAMB software package [10]. The scale of the 75% dark matter universe is not shown.

The dynamics of the plasma can be used to parametrise the density of dark matter and baryonic matter. Firstly, dark matter is weakly interacting with itself and baryonic matter such that it experiences no pressure, from radiation or otherwise. This means that only baryonic matter experiences radiation pressure to drive oscillations. The baryonic matter is therefore the inertia of the system, so the greater the proportion of baryonic matter, the larger the peaks, as can be seen in the no dark matter scenario. After the first compression, gravity will pull in dark matter, which remains after baryonic matter expands from radiation pressure. This enhances subsequent peaks, as can be seen by comparing the no dark matter scenarios to those with DM. Varying the proportion of dark matter can give mixed results, as if the total amount of matter is conserved, there is a trade-off between inertia and magnitude of the gravitational potential.

Returning to the Friedmann Equation 1.2, the CMB constrains its parameters through a variety of features. The position of the first peak is sensitive to the total evolution of scale, which when other parameters are constrained is a unique probe into Ω_k. The ratio of the first and third peak is sensitive to the Ω_χ, which enhances subsequent compressions. The ratio of first and second, as well as subsequent peaks, informs Ω_b, as it suppresses them. The total matter

Ω_m can impact the shape as well, this causes matter to dominate radiation earlier and form larger scale perturbations, enhancing the lower ℓ peaks. This can be seen as a "tilt" clockwise in the CMB power spectrum. The best fit values to Planck CMB data provided in the 2018 data release are shown in Figure 1.3.

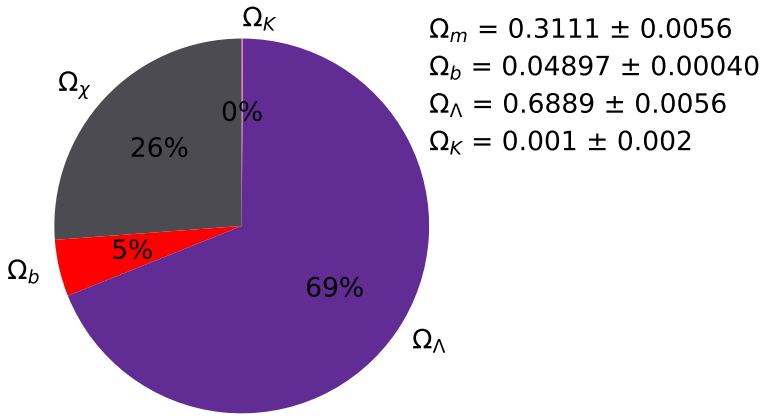


Figure 1.3: Cosmological parameters that are best fit to the Planck CMB data [8]. The pie chart gives a visual representation of the proportions. Recall Ω_m=$\Omega_b + \Omega_\chi$.

1.1.4 Modern Observations

In addition to the CMB and predictions from ΛCDM cosmology, there are three notable modern observations that help with their interpretation and motivating dark matter. These are; the abundance of light nuclei helps constrain Ω_b, observations of the expansion of the universe with redshift give Ω_Λ, and the modern structure of the universe supports ΛCDM.

Big Bang Nucleosynthesis

As the universe expands and cools, the energy of individual particles decreases so that certain processes cannot occur thermally, which is known as freeze-out[2]. In the early universe, the weak interactions,

$$p + \overline{\nu} \rightarrow n + e^+$$
$$n + \nu \rightarrow p + e^-,$$

were in thermal equilibrium. Given that the temperature of the universe was greater than the mass difference between the proton and neutron, this process would be in thermal equilibrium with a fixed proportion between the two. Once the process froze-out at a temperature[3] of 1 MeV the ratio of protons and neutrons was fixed, except from the known decay rate of neutrons through beta decay. The same processes of freeze-out occurs for the fusion of protons, neutrons,

[2]If red-hot poker has cooled, one could say the electronic transition to form visible light has frozen-out.
[3]the average energy of particles at this temperature.

and lighter elements. The cross-section of the above processes, and nuclear fusion, needs to be well-known to predict the relative abundances of these light elements. The production of the light nuclei is known as Big Bang Nucleosynthesis (BBN) and that term generally refers to this measurement. Further, stellar processes can alter the ratios of these elements, but Deuterium provides a strong lower bound as it is only consumed in stars. Constraints from modern day measurements of the deuterium-hydrogen ratio and ^{4}He provide Ω_b measurements consistent with the CMB analysis. Lithium is in tension with observations, but subject to more systematic uncertainties than the other light elements [11]. This helps to confirm the non-baryonic nature of the matter density of the universe.

Dark Energy Measurements

The distance to galaxies can be determined through measurements of the brightness of standard candles; initially Cepheid variables were used as their brightness oscillated with a frequency proportional to that of their total luminosity, allowing the use of its brightness to calculate its distance. These measurements coupled with red-shift allowed for the initial measurement of the rate of expansion and Hubble constant. Tracking this expansion further back into the history of the universe required the use of brighter standard candles, for which supernovae are used. A spectral analysis of their light reveals the type of supernovae observed, and from this the distance can be determined similar to the cepheid variables. These measurements, particularly expansion at earlier times from more distant galaxies, allows for the tracking of expansion throughout the history of the universe and direct determination of the dark energy density Ω_Λ [12].

Large Scale Structure

The CMB provides insight into the structure of the universe at recombination, but this would have evolved over time. There are many more observational probes, such as the formation of gases, the structure of distant galaxies, and the local structure that allow the reconstruction of the modern matter-power spectrum. These observations are consistent with predictions from the ΛCDM cosmology within approximately 1 σ [13]. A growing part of this field is reconstructing to smaller scales, which requires intense N-body or hydromechanics simulations to compare to [14].

1.1.5 Gravitational Lensing and Cluster Collisions

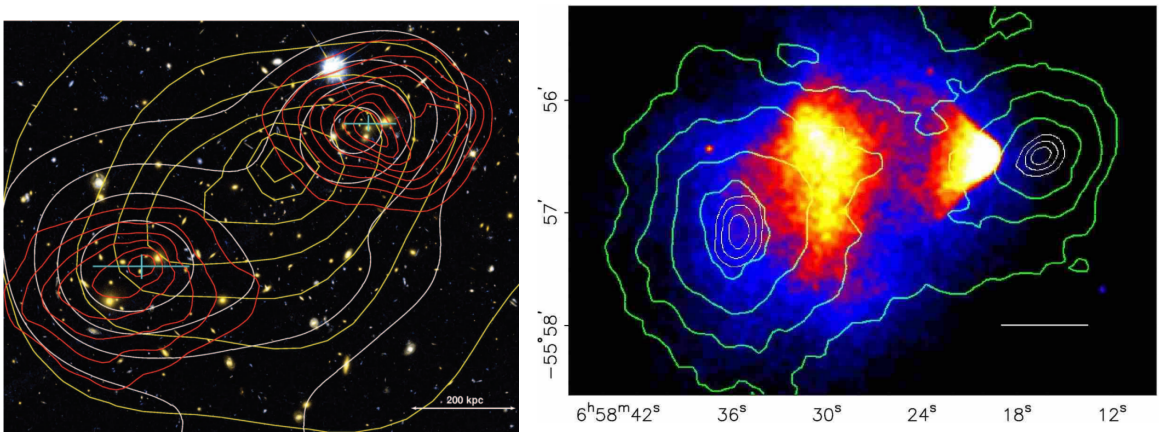


Figure 1.4: The matter distribution inferred by X-rays emission, as observed by the Chandra observatory, and that from weak-gravitational lensing. The X-rays are in yellow and weak-lensing red for the cluster merger MACS J0025.4-1222 taken from Ref.[15] (left). The distribution of X-ray emitting plasma with inferred distribution from weak-lensing for the cluster merger 1E0657-558 or "bullet cluster" taken from Ref.[16] (right).

Gravitational lensing was first observed in 1979 in Ref.[17], through a particularly strong effect of a double image of a quasar. Since then, gravitational lensing has been used as a measurement to measure the total mass of a body or a collection of bodies in the sky. When a large mass such as a black hole or a cluster of galaxies is between the observer and a distant object, the light travelling from the distant object is distorted. Strong gravitational lensing can create multiple images of an object, such as Einstein rings. The weak lensing effect may also be observed through small distortions, usually requiring many distant bodies to map. Using the strength of gravitational lensing over the intermediate object, the distribution of matter can be determined. This can be used as an independent measurement of the total matter of an object and as a measure of its distribution [16][18]. Of particular interest are collisions of galaxy clusters, where the luminous matter distribution can be observed through the behaviour of colliding interstellar gas, which emits X-rays and weak-lensing effects can map the total matter distribution. In Figure 1.4, two such studies are provided clearly showing the distribution of matter displaced from the visible. The right image is the bullet cluster, found to give strong evidence for a displacement between the centers of mass inferred by weak lensing and that of the luminous mass in Ref.[16]. Lensing provides another probe into the nature of dark matter via microlensing searches for compact, often sub-galactic, objects. This has been used to study black hole candidates and the size and mass of substructures in galactic dark-matter distributions[19][20].

1.2 Paradigms of Dark Matter

There are three primary paradigms to explain these phenomena; here, their status is reviewed with the exception of particle dark matter, which is the subject of Section 1.3.

1.2.1 Modified Gravity

As discussed above there are numerous astrophysical signatures implying a distribution of non-luminous mass, but it may also be possible that we misunderstand the dynamics of the systems or gravity itself. Modified Gravity or Modified Newtonian Dynamics (MOND) theories seek to explain these observations with minimal modifications to Newtonian dynamics. First invoked to describe observations of rotation curves, these modifications typically rely on gravity acting differently in the low-acceleration limit at $a \sim 10^{-10} ms^{-2}$[21]. Wide binary star systems can be used to study this low acceleration regime, with several notable recent results from the Gaia space satellite survey. To date, there are four major studies, with two rejecting Newtonian dynamics in favour of MOND and two not. The rejection in Ref.[22] is as great as 5σ, however Ref.[23], which includes free-parameters for potential systematics like tertiary stars, claims to reject MOND in favour of Newtonian dynamics at 19σ. A review of these is available in Ref.[24], claiming the latter are biased by omission of observational error and excluding a high acceleration sample. With a new round of Gaia data in 2026 these results will be re-analysed, possibly resolving this tension[25].

Regardless of these observations, MOND still faces many issues as an explanation for dark matter. Although MOND, through modified general relativity, can explain cosmological observations, it is harder to reconcile cosmological models with those that explain modern astrophysical observations than it is in ΛCDM/FRW cosmology [26] [27]. Similarly, whilst MOND can explain in isolation phenomena like the bullet-cluster, it struggles in combination with many results like the CMB third peak and cluster masses inferred from X-ray emitting gas without invoking either more exotic modifications or a combination with some other form of dark matter [28][29].

1.2.2 Primordial Black Holes

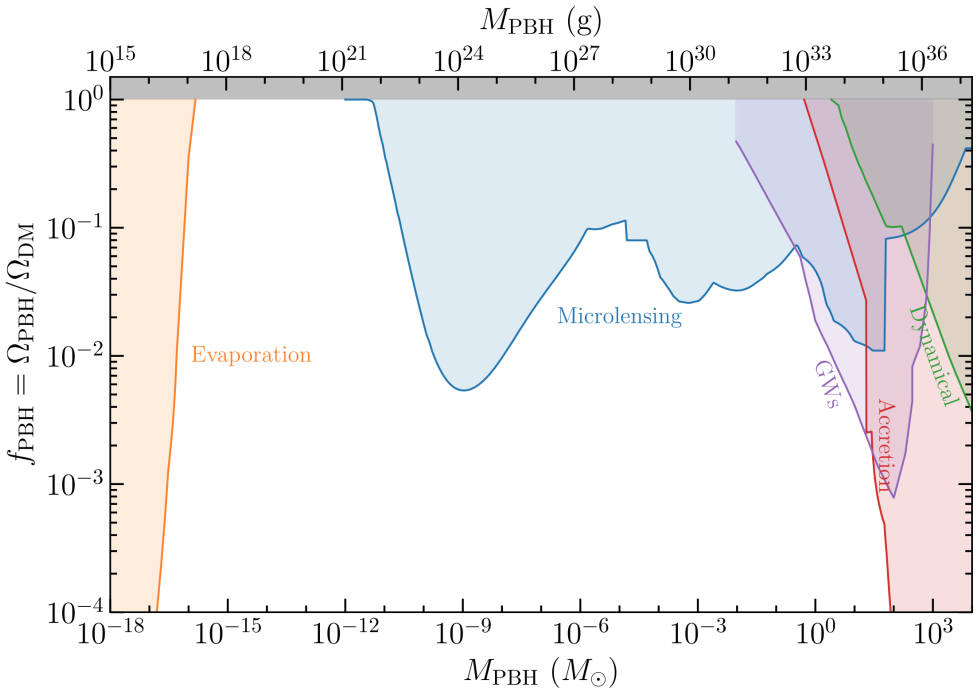


Figure 1.5: Various constraints on the dark matter mass fractions of primordial black holes from different experiments, taken from Ref.[30] generated using the `PHBounds` package in Ref.[31] , 95% CL appears to be the standard. Dynamical limits are mostly accretion injecting energy into plasma impacting CMB observations.

Massive Compact Halo Objects (MACHO) were an early explanation for dark matter observations, assuming that the additional non-luminous matter was dim baryonic objects. Baryonic matter in the form of brown dwarfs, faint stars and stellar remnants have long since been ruled out by micro-lensing observations and struggle when introducing un-observed chemical abundances in the case of the latter [32]. As discussed, the fraction of baryonic matter is also limited by BBN and BAO, requiring some exotic cosmology for MACHOs to be a sensible candidate. However, black holes could remain as a possible explanation for dark matter. Whilst stellar black holes would be ruled out by the same micro-lensing observations, so-called primordial black holes (PBH) formed by density perturbations in the early universe, pre-BBN and CMB formation, could account for some if not all dark matter. These candidates could meet the requirements for dark matter, namely dark and collision-less, but have a limit on what fraction they can explain depending on their mass. For instance, if they are too small and they will evaporate too quickly to explain dark matter or not be observed in gravitational waves and gamma-ray bursts, too large and they are limited by micro-lensing observations[30]. There are dynamical limits as well, if they are too large they cannot exist or would have injected too much energy into early universe baryonic matter via accretion and change the CMB [33]. Further, photon emission in evaporation and gravitational waves from formation of binaries and mergers

can be used to search for PBHs [30][34]. Limits from some of these searches are shown in Figure 1.5.

1.2.3 Particle Dark Matter

For a particle, or sector of particles, to explain dark matter it must meet the following criteria;

- **Cold**: Dark matter must be non-relativistic otherwise it would not form the gravitational potential wells necessary to describe the CMB and large-scale structure of the observations.
- **Dark → Neutral**: To be non-luminous, fit the observed CMB, galactic rotation curves, and galaxy collisions, dark matter must have a charge-to-mass ratio less than 10^{-4} for $m_\chi \geq 1\ GeV/c^2$[35].
- **Non-Baryonic**: To not interfere with BBN the dark matter candidate must not be a baryon and have a weak baryon interaction cross-section.
- **Collisionless**: In order to form diffuse halos, CMB, lensing, and cluster collision observations, dark matter must have a weak self-interaction cross-section.

In the next section, this work discusses why it is reasonable to postulate that such particles may exist, the wide zoo of candidates, and the current status of searching for them.

1.3 Particle Dark Matter Candidates

No fundamental particle exists to explain dark matter in the standard model of particle physics. Despite its success in explaining observations so far, there are several aspects of the standard model which particle physicists consider to be incomplete. A non-exhaustive set of reasons for suspecting beyond the standard model (BSM) physics are:

- **Quantum Gravity** There remains no quantum description of gravity.
- **Observed Matter/Anti-Matter Asymmetry** Our universe is not an even mix of baryons and anti-baryons, requiring some processes in the early universe to be asymmetric between the two. This requires baryon number and charge-parity (CP) symmetry violation. Currently, the only confirmed source of CP asymmetry in the standard model is in the electroweak interaction of quarks and is insufficient to produce the observed asymmetry[36].
- **Neutrino Mass** Neutrinos must have mass in order to undergo flavour oscillations as they travel through space. In the standard model the Higgs field provides a self-coupling in fermions and bosons, generating their masses. Currently, no such coupling exists for

neutrinos as it requires their parity opposite counterparts to exist, which are yet to be observed.

- **Naturalness** This refers to several problems in which, in order to explain observations, the standard model requires some serendipitously small constants or fine tuned cancellation. Namely, the lack of antimatter asymmetry in the strong interaction of quarks, the estimated vacuum energy compared to dark energy, and the hierarchy problem, whereupon the Higgs receives large corrections to its mass, are all naturalness problems [37].

Some of the above issues may be resolved, in combination or isolation, by the addition of a new particle or entire sectors of particles. A diagram of the zoo of potential candidates for cold dark matter is shown in Figure 1.6 and how projected experiments may cover the space over 20 years. For particle dark matter at lighter masses, wave-like behaviour dominates with substantial implications for the structure of dark matter halos. These lighter candidates are produced differently from the typical freeze-out mechanism for leptons, baryons, and particle dark matter, through a process known as misalignment [38]. In order to obtain a breadth of the field the sterile neutrino ν_sDM, the QCD axion, and the classical WIMP will be discussed.

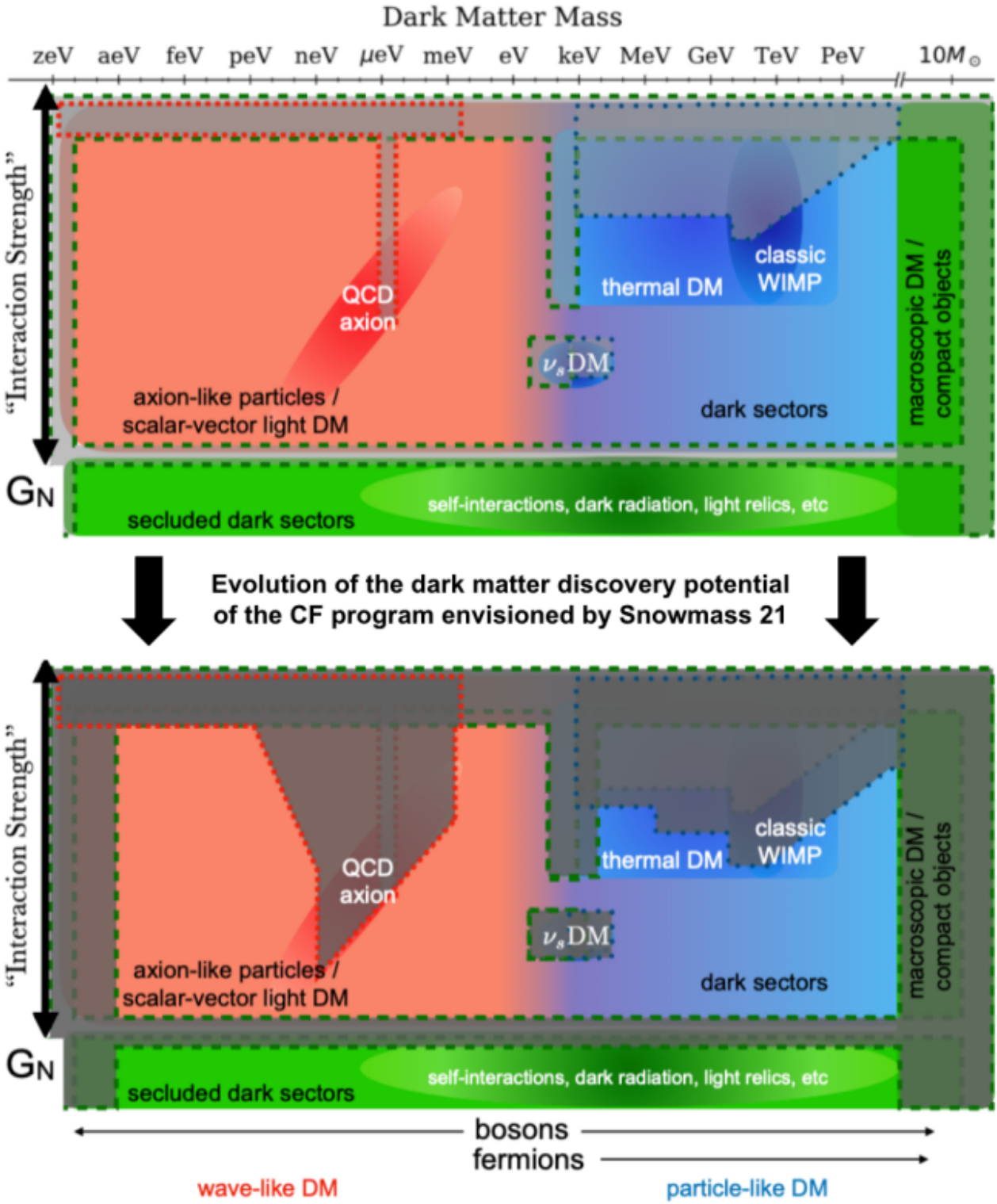


Figure 1.6: Diagram of particle dark matter candidates from the cosmic frontiers report of SNOWMASS 2021, taken from Ref.[38]. The y-axis is an indication of interaction strength with matter and the x-axis shows the candidate mass, both logarithmic. Red shades indicate wave-like particles and blue particle-like. The light to dark shaded regions show the gains from experiments proposed over the 20-year remit of plans of the group.

1.3.1 Neutrino and Sterile Neutrinos

Initially, neutrinos may appear to be an ideal candidate; they are weakly interacting, have mass, and are known to exist. Further, there is evidence of CP violation in neutrino interactions, with CP conservation excluded by 3σ under different possible scenarios, but its value may be small [39]. However, the observed neutrinos cannot explain dark matter as the limits on their mass are $m_{\nu_e} < 0.2$ eV, meaning that they could not create a sufficiently large component of the matter density of the universe. Also, since they are produced at relativistic speeds they would not be able to form the large scale structure and CMB observations [40]. However, their mass is not yet explained by the standard model and some of the methods to generate it require either the presence of heavy right-handed neutrinos which present as a sterile neutrino. This would only be able to interact with matter through mixing with the standard neutrinos, and therefore be suppressed below weak interaction to be feebly interacting, as shown in Figure 1.6. These models also allow some of the baryon asymmetry to be accounted for through processes known as leptogenesis [41]. Attempts to search for the sterile neutrino mainly involve searching for neutrino disappearance when they oscillate into the sterile state in short baseline experiments, limits from the CMB structure, and experiments probing the nature of neutrinos like neutrinoless double-beta decay.

1.3.2 QCD Axion and Axion-Like Particles

Axions are an interesting class of particle candidates, as they were theorised around the same time as the acceptance of the dark matter phenomena in 1977 in Ref.[42] for a completely different problem. This problem was the absence of CP violation in the strong force, despite its absence requiring a vanishing constant for the allowed four-gluon self-interaction term in Quantum Chromo-Dynamics (QCD), which is considered unnatural. It was proposed that a new mechanism, the Peccei-Quinn mechanism, could remove this unnatural term but predicted a new massive boson, the axion. Axion particle production is not a freeze-out or in mechanism, instead it is formed when QCD undergoes a phase transition in the early universe, which allows it to be light whilst being produced at non-relativistic speeds to form cold dark matter, unlike standard neutrinos [43]. As their mass is given by QCD interactions, it is proportional to their coupling strength. In addition to limits from cosmology, this restricts the physically allowed region of parameter space, as shown in Figure 1.6. However, a different field of Axion-Like Particles (ALPs) was conceived that do not acquire mass from QCD effects and therefore are free of these constraints [44]. Coupling via virtual quarks allows a photon in the presence of a strong magnetic field to convert to an axion through the Primakoff effect, and its inverse is also possible. The latter allows for direct detection via microwave cavities that catch photons from halo axion conversion, reviewed in the SNOWMASS 2021 whitepaper in Ref.[44]. Both processes allow for production measurements by using strong magnetic fields to convert photons in a laser to axions and back after some obstruction blocking the other photons [45]. Finally,

the axioelectric effect allows for detection in dark matter experiments sensitive to electron scattering; the sensitivity of xenon based experiments is reviewed in Ref.[46]. To date, there has been no evidence of the QCD axion or ALPs, but as shown in Figure 1.6, much of the phase space remains unexplored.

1.3.3 Weakly Interacting Massive Particle Candidates

Weakly interacting massive particles (WIMPs) are a subset of thermal dark matter, that is, particles that are produced by the aforementioned freeze-out mechanism. As with BBN, but far earlier, this mechanism requires a thermal equilibrium of a process that mixes the two species, here instead of protons and neutrons it is WIMPs and standard model particles via annihilation,

$$\chi + \overline{\chi} \rightarrow S + \overline{S}$$

where S indicates the standard model particle. Requiring that the dark matter species freezes out when it is non-relativistic, as argued in the CMB and structure observations, it can be found that the resulting relic density is

$$\Omega_\chi h^2 \approx \frac{3 \times 10^{-27} cm^3/s}{\langle \sigma_{\chi\chi} v \rangle}, \tag{1.3}$$

where $\langle \sigma_{\chi\chi} v \rangle$ is the velocity-averaged annihilation cross-section, at the time of freeze-out. The rate of annihilation is directly proportional to this value. It is worth noting that this only works if annihilation to standard model particles, and therefore a coupling, exists to allow thermal equilibrium to exist. If this coupling is at the electroweak scale the Fermi constant can be used to estimate the scale of the cross-section at $\sigma_{\chi\chi} \sim G_F^2 m_\chi^2$, which would yield a WIMP mass of $m_\chi \sim GeV/c^2$ for the correct relic density. At the conception of WIMPs, and today, heavy weak-scale particles were postulated in many BSM models such as supersymmetry, which solves a number of the proposed problems predicting a set of heavier partners to current particles. The lightest of these heavy particles would be stable and form ideal WIMP candidates, such as neutralinos [47]. This is just one example, and WIMPs cover a broad range of BSM candidates. Given the theoretical motivation for these particles and the good fit to ΛCDM dark matter, this remarkable coincidence was called the "WIMP miracle".

1.3.4 Astronomical Probes to Dark Matter

In Section 1.4, techniques to detect WIMP particle dark matter, it is worth noting that qualities of dark matter, wave vs particle -like or the strength of their self-interaction, may be probed through astronomical observations.

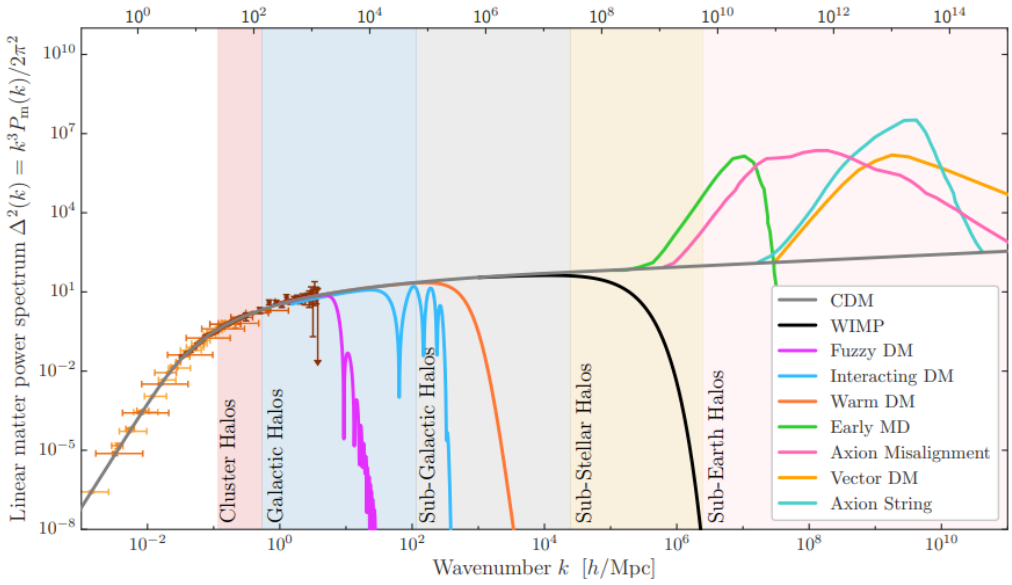


Figure 1.7: Plot of the modern day matter power spectrum with different theoretical dark matter candidates taken from Ref.[48]. The data points are from the combined study in Ref [13].

Studying the modern day matter power-spectrum through a series of probes and comparing to N-body simulations is a powerful tool in investigating dark matter, as shown in Figure 1.7. The current observations are limited to galactic halo scales and as improved probes are devised and tested, more of the nature of dark matter will be revealed [48]. One point in this field is that the classical cold non-interacting particle-like dark matter ΛCDM model can fail local observations of galaxies in three senses:

- **Cusp/Core** Observations show galaxies with greater density cores and lower density cusps than predicted.
- **Missing satellites** There are fewer dwarf-galaxies observed than predicted.
- **Too big to fail** There are too few observed galaxies given the local dark matter density. These should have formed dense enough regions to trigger baryonic matter into star formation, yet they are not observed.

These are discussed in detail in Ref.[49], it is worth noting that within current observation and simulations it is possible to tune a ΛCDM model that fits these observations. These problems may be resolved by wave-like dark matter and self-interacting particle-like dark matter, both of which assist in producing the correct observations. However, it is worth noting that studies show that wave-like dark matter requires more complex baryonic or dark matter effects to explain the behaviour [50].

A

1.4 Searching for particle Dark Matter

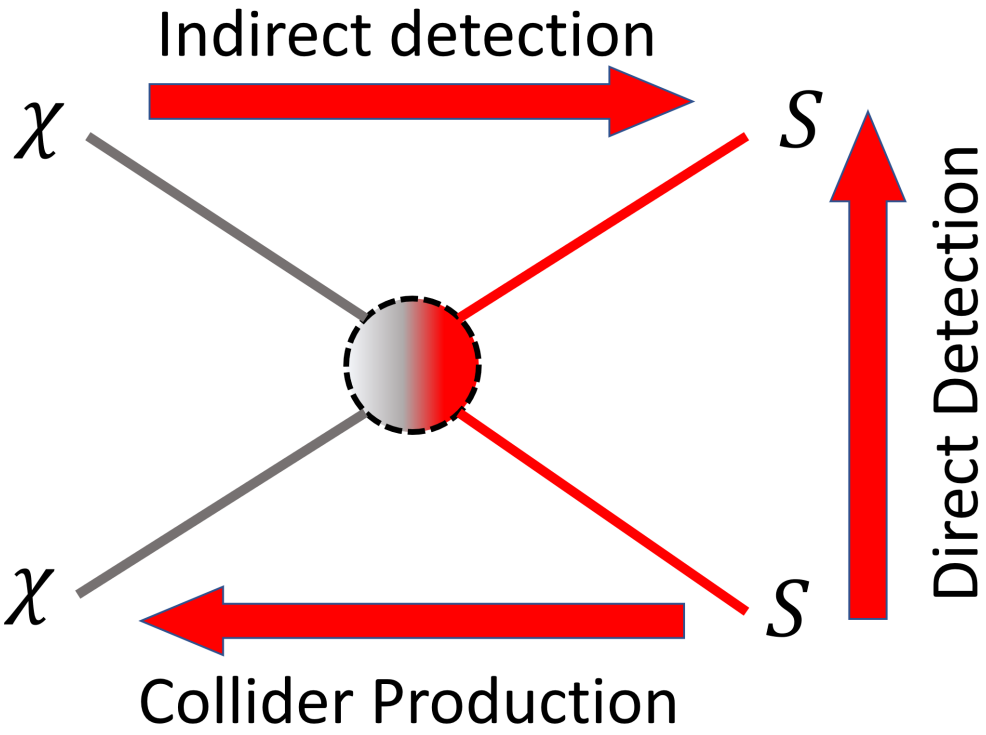


Figure 1.8: Diagram of interactions between dark matter and the standard model. The circle represents a coupling between the dark matter and standard model species.

There are three main methods to search for particle dark matter which are,

- **Collider Production** The production of dark matter particles through the annihilation or decay of standard model ones.
- **Indirect Detection** The measurement of standard model particles created in the annihilation or decay of ambient dark matter.
- **Direct Detection** The measurement of ambient dark matter interacting with standard model particles.

The term ambient here refers to dark matter in the halo of our galaxy or elsewhere. These three processes are shown in Figure 1.8, the existence of a coupling for one implies the others. Our primary focus is WIMPs, but in the following discussion we will comment on other particle candidates.

1.4.1 Collider Production

Collider experiments such as ATLAS and CMS at the LHC have a range of utilities, however their primary function is to detect the decay products of the particle produced resonantly in

the proton-proton collision. Using conservation of momentum in the transverse plane, the mass of these products can be reconstructed. If dark matter production in a collider experiment was precisely as labelled in Figure 1.8, two SM model particles annihilate to DM, then there would be no signature to detect. Therefore dark matter must be searched for in processes like those shown in Figure 1.9, where one or more standard model particles, like the scattered quarks, can be reconstructed and a large amount of missing momentum being carried away by the undetected dark matter.

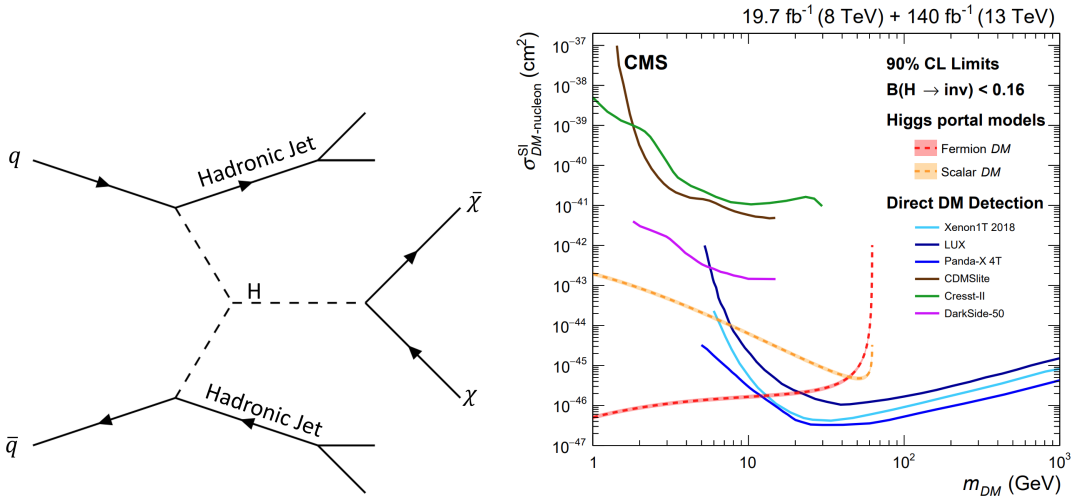


Figure 1.9: Example of a quark-antiquark collision from protons that would produce a detectable dark matter signature (left). Limits from a search for said process in CMS compared to direct detection experiments, taken from Ref.[51].

Whilst early searches used relativistic Effective Field Theory (EFT) for comparison to direct or indirect detection, it was quickly found that this was not applicable as the mass of any detectable direct detection mediator would be similar to energies probed at the LHC and instead simplified models are required, breaking the EFT assumptions. For a tree level process like the one shown, there are three parameters that impact the interaction rates; the coupling strength of the mediator with the scattering standard model particles, its coupling with the dark matter particles, and its mass [52]. Shown are the limits derived from a search using CMS compared to direct detection limits using the Higgs portal model, assuming a mediator of Higgs mass with Higgs coupling to be standard model values and placing limits based on varying its coupling to potential dark matter particles [51]. There is currently no evidence of WIMP dark matter in colliders, but they do provide important, if model-dependent, constraints particularly in the low-mass regime.

There are other methods of searching for dark matter at colliders that could be interpreted as indirect; one of the simplest is that if dark matter is available as a final state for a known particle to decay into, it will decrease its lifetime. The contribution to this lifetime can be determined from the energy spectra of decays into each other possible daughters, and so limits

are set on any unknown contribution to the particle's lifetime. These are known as decay-width studies and have been used to limit sterile neutrinos using electroweak gauge bosons [53]. Other examples are dark matter candidates contributing to measured standard quantities, like the muon magnetic moment in the g-2 experiment[54]. Although there are several anomalies of this type currently present in the standard model, many of them have large theoretical systematics, are not consistent between experiments, or are not consistently explained by a BSM extension; a review is available in Ref.[55].

1.4.2 Indirect Detection

Experiments searching for standard model particles from the annihilation or decay of ambient dark matter work by characterising the Cosmic Rays (CR) that reach earth. These can be a variety of species, from neutrinos to gamma rays, and this section will discuss some of the most prominent. For all, the expected rate has a strong dependence on the density of dark matter and the source being observed. These types of searches are strongly dependent on models of the halo and typically search within the local region, i.e within the Milky Way, nearby dwarf galaxies, or clusters. Observations by the Fermi gamma-ray telescope indicated an excess of gamma rays in the Galactic Center (GC) which could have been explained by annihilating WIMPs of mass $m_\chi \sim \mathcal{O}(50)\ GeV/c^2$ [56]. Although recent analysis found that this excess extends into the galactic plane, in Ref.[57], and is in tension with expected local dark matter distribution, another study in 2025 in Ref.[58] indicates that this may yet be consistent with dark matter when considering updated local density measurements. Experiments such as Alpha Magnetic Spectrometer (AMS) on the international space station search for excesses in anti-protons and heavier baryons and IceCube can detect high energy neutrinos, both of which could be produced from the decay of dark matter annihilation or decay produced [59].

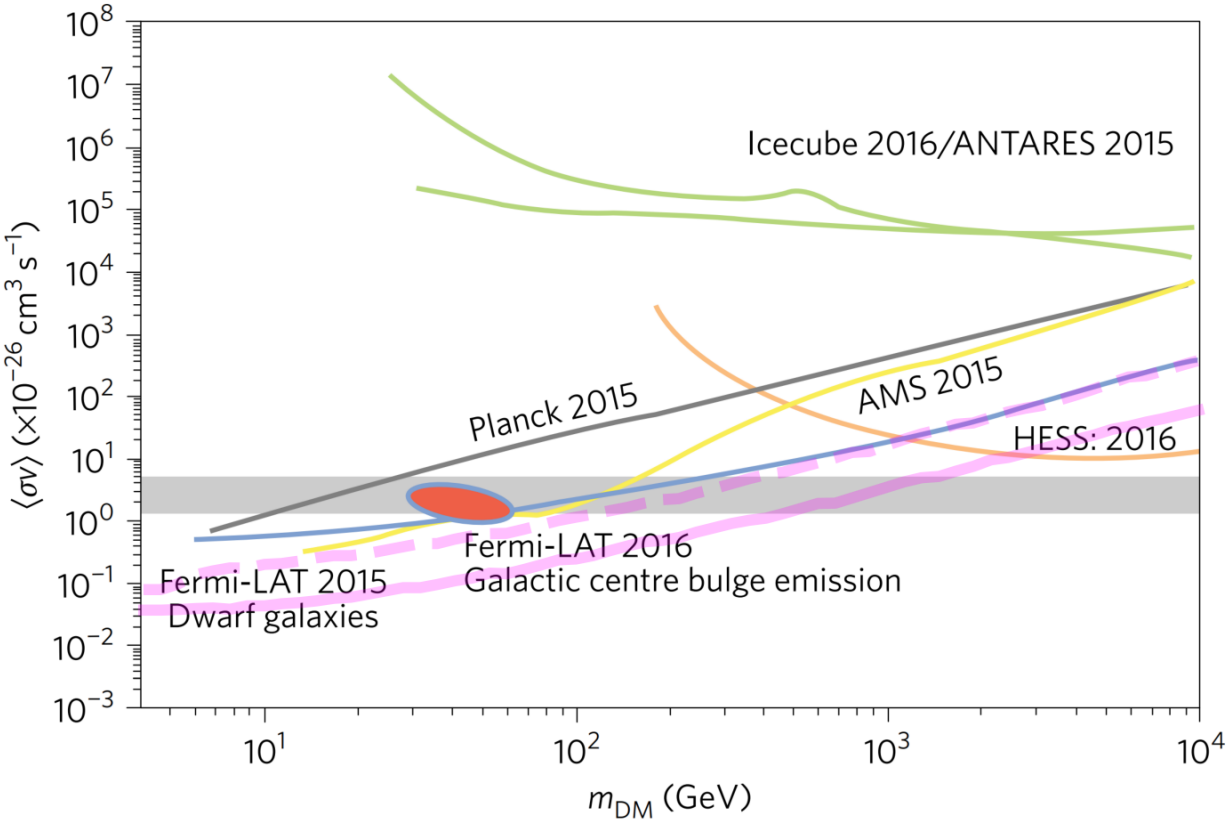


Figure 1.10: Limits on the annihilation cross-section for $\chi\bar{\chi} \rightarrow b\bar{b}$ taken from a 2017 overview of indirect detection in Ref.[59]. The grey bar indicates the range of cross-sections consistent with the WIMP thermal relic density. The magenta lines were digitised from a 2026 combined analysis of Fermi-LAT, HAWC, H.E.S.S, MAGIC, and VERITAS data on dwarf spheroidal galaxies, the two dashed lines are separate assumptions of the DM density distribution profile [60]. The confidence level is not stated but 95% confidence level seems to be the standard, with the exception of ICECUBE at 90% [61].

The comparison of these experiments depends strongly on the standard model products of the WIMP annihilation, shown in Figure 1.10 are the limits if the products are b-quarks. Even omitting the recent analysis, the GC bulge is excluded by dwarf galaxy studies. An interesting note is that the presented cross-section is directly applicable to that in Equation 1.3, so the requirements for it to describe the thermal relic density can be directly imposed. This shows that WIMPs of mass $m_\chi \lesssim 100\ GeV/c^2$ cannot constitute the entire relic density according to limits from Fermi dwarf galaxy observations.

1.4.3 Direct Detection

Broadly speaking, direct detection experiments search for energy imparted onto standard model particles within terrestrial labs. Different candidates invoke different experimental methods, like microwave cavity searches for axions, as discussed, or search for shifts in fundamental constants visible to experiments such as atomic clocks caused by the interaction of wave-like dark matter

[62]. Here, the discussion will be focussed on particle-like dark matter, particularly WIMPs. Since the favoured mass of WIMPs is $m_\chi \sim \mathcal{O}(GeV/c^2 - TeV/c^2)$, maximum momentum transfer is expected for nuclei or nucleons, and suppressed for electrons. Often these experiments are still sensitive to electron scattering, albeit with larger backgrounds [63]. Dedicated experiments have begun to search for electron scattering using solid-state detectors.

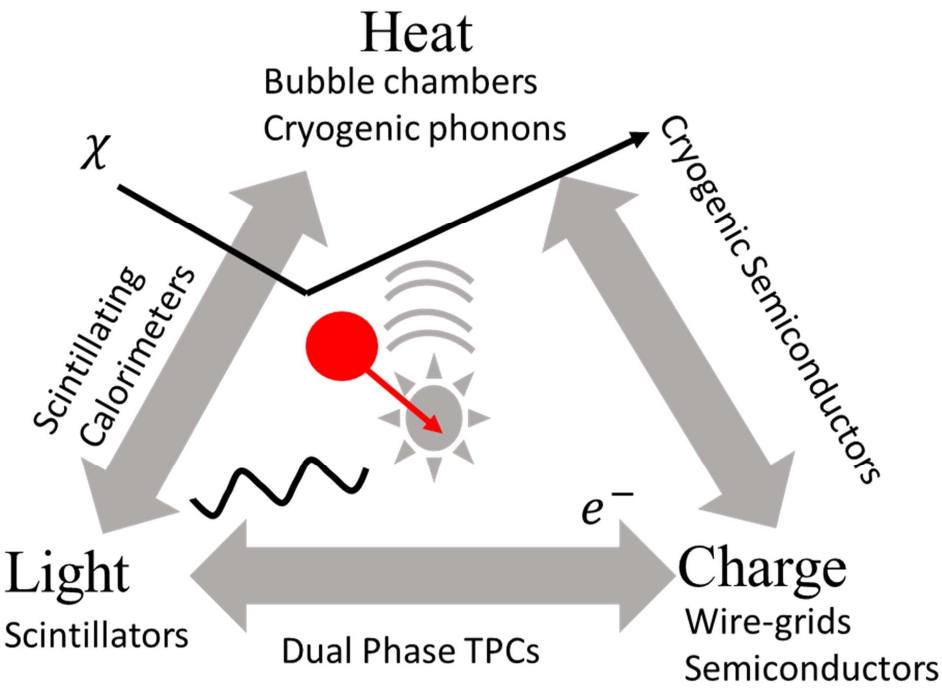


Figure 1.11: Diagram of the collectable channels from a scatter and examples of technologies that utilise them.

The identification of dark matter scatters can occur through three channels; scintillation, ionisation, and heating of the target. Here, a few examples of the technologies utilising different channels, and present limits, are discussed. It is typically advantageous to combine channels, not only increasing signal detection efficiency but rejecting backgrounds as different recoiling particles distribute energy differently across each channel. However, in some cases, focussing on a single channel can be beneficial when read-out for different channels would interfere with each other or are not feasible.

To measure the ionisation channel one needs to extract charge, and therefore a target with high charge mobility. For example, charge collection grids throughout the detector are used in CYGNUS for track reconstruction, whilst the central electrode in NEWS-G gives a very

high field for proportional scintillation and low energy thresholds [64] [65]. Another common technique is to use semi-conductor systems, like the charge-coupled devices in DAMIC and SENSEI[66][67]. The light channel may be used exclusively through scintillating crystals or liquids, such as at COSINE and ANAIS or DEAP-3600[68] [69]. These can still utilise charge, as depending on the scintillation properties, the time-profile of recombination can be distinguished from primary scintillation [70]. Comparing scintillation and ionisation yields can distinguish electronic and nuclear recoils, the former typically having a greater amount ionisation than scintillation relative to the latter. Using dual-phase Time Projection Chambers (TPCs) to read out both light and charge is a common technique, as it allows for 3D position reconstruction within the target; the charge extraction point can give the x-y plane and depth from the delay between the prompt light and charge signal. Current experiments that use these are the xenon experiments, LZ, XENONnT, and PANDA-X, as well as the argon-based Darkside-20k [71][72][73][74].

For the heat channel, to resolve a single event it is often required to produce a detectable phonon or a localised phase-transition. For example, the PICO experiment detects gas bubbles from the nucleation of a superheated fluid by the deposited energy of a nuclear recoil [75]. These methods are insensitive to electronic recoil backgrounds or signals, as these types of scatter do not deposit much of their energy as heat, as discussed in Section 2.2.1. It is rare to find experiments using solely the phonon channel, as experiments often use semiconductors which allow for charge collection, as in superCDMS, or scintillating crystals that enable light collection, as in CRESST [76][77]. Heat and light have been combined in experiments like HeRALD which uses super-fluid helium to detect phonons and scintillation light [78]. Scintillating Bubble Chambers (SBCs) have also been demonstrated and show excellent potential in background discrimination [79]. A combination of all three channels has been proposed with a super-heated liquid xenon bubble chamber [80].

Current and projected direct detection limits on spin-independent dark matter-nucleon cross-section are shown in Figure 1.12. Not shown are the sub-GeV experiments such as Dark Sphere, a successor to NEWS-G, and HeRALD/TESSERACT [81]. Here we can see that the collection of heat from nuclear recoils for proposed SBC experiments significantly lowers threshold and background, as it competes with multi-ton-year experiments. The most competitive experiments above 10 GeV/c^2 are 10-tonne scale noble element dual-phase TPCs. At sub-GeV masses there is competition between ionisation-only searches in the noble TPC experiments and the kg-scale SuperCDMS experiments. Xenon and argon-based TPCs are scalable and exhibit excellent background discrimination, so are considered some of the best in the field.

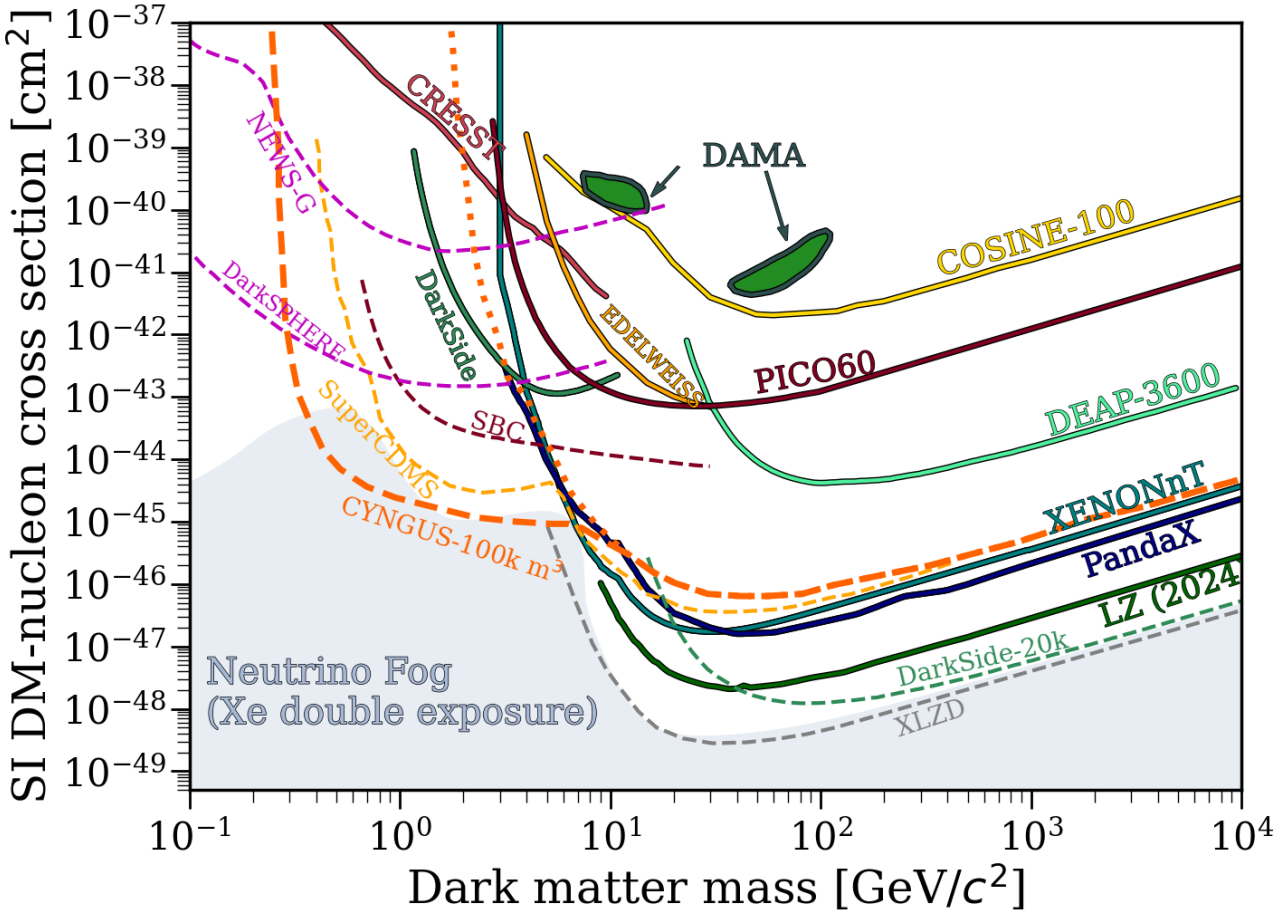


Figure 1.12: Plot of current and future limits from direct detection spin-independent dark matter search experiments generated using code by Ciaran O'Hare at Ref[82]. Many limits are found in the SNOWMASS version for direct detection review[83], the author added the best-case scenario for CYGNUS-100k m^3 limits from their feasibility study in Ref.[84]. This omits some low-mass experiments discussed that do not project, or quote, limits in standard WIMP-nucleon cross-section, for example SENSEI. The "neutrino fog" shows where nuclear scattering of neutrinos becomes a significant background, discussed in the next section.

1.5 WIMP Direct Detection Basics

The technologies available to search for WIMPs through direct detection have been covered in the previous section, and here the basic inputs to calculating the differential event rate are presented, as well as comments on how future technologies may develop past the neutrino fog.

1.5.1 Scattering Rate

The calculation of the differential event rate allows for its convolution with an observable-dependent detector response. For this thesis only energy is considered as a relevant observable, but direction of the recoiling nucleus could be useful, discussed in the next sub-section. The

event rate of scatters inside a terrestrial detector may be expressed in terms of

$$\frac{dR}{dE_R} = \frac{\rho_0 M}{m_N m_\chi} \int_{v_{min}}^{\infty} v f(\vec{v} - \vec{v}_E) \frac{d\sigma}{dE_R} dv \tag{1.4}$$

where ρ_0 is the local dark matter density, M the mass of the target material, m_χ and m_N the WIMP and the target nucleus mass. The velocity $\vec{v} - \vec{v}_E$ determines the lab-frame velocity of the WIMP and f is the dark matter halo distribution. The differential scattering cross-section may be calculated utilising kinematics, namely

$$E_R = \frac{\mu^2 (v = |\vec{v} - \vec{v}_E|)^2}{m_N} (1 - cos\theta) \tag{1.5}$$

where θ is the scattering angle and μ is the reduced mass of the WIMP-target nucleus system. Using chain rule this yields,

$$\frac{d\sigma}{dE_R} = \frac{dcos\theta}{dE_R} \frac{d\sigma}{dcos\theta} = \frac{m_N}{\mu^2 v^2} \frac{1}{2\pi} \frac{d\sigma}{d\Omega}. \tag{1.6}$$

The last term being the differential cross-section obtained from scattering theory,

$$\frac{d\sigma}{d\Omega} = |T(k, k')|^2 \tag{1.7}$$

where k and k' are the ingoing and outgoing WIMP momentum, respectively, and T the transition probability. The velocity distribution of the dark matter halo, $f(\vec{v})$, is an astrophysical input. A white paper on direct detection, see Ref.[85], recommends the standard halo model, which is a Maxwell-Boltzmann distribution with the mean lab-frame velocity of the standard halo model $\sim 350\ kms^{-1}$. Assuming a WIMP of mass $m_\chi = 1\ TeV/c^2$, the average velocity corresponds to a de Broglie wavelength of $\lambda = h/p$ of similar size to the nucleus. Therefore, as individual nucleons are not resolvable, the scattering of a WIMP can be sufficiently described as non-relativistic plane-waves subject to a near-point like potential, with T being the transition probability found by

$$T(k, k') = \int \varphi_i(r) \varphi_f(r) d^3r = \int \mathcal{N} e^{ik.r} \varphi_f(r) d^3r \tag{1.8}$$

where $\varphi_{i/f}$ are the initial/final wave functions. Since this is non-relativistic the time independent Schrödinger equation and for a weak potential first-order perturbation theory, in this context called the Born-approximation, may be used to calculate this transition probability. Assuming the final wave function is a plane wave, it is

$$T(k, k') = \int \mathcal{N}^2 e^{iq.r} V(r) d^3r \tag{1.9}$$

where we have reduced the dependence to that of the momentum transferred, $q = k - k'$ and $q = \sqrt{2 m_N E_R}$. The potential, $V(r)$, may be arbitrarily complex, but for a spin-independent

WIMP we expect only coupling to the nucleons via a point-like interaction, therefore $V(r) \propto \rho(r)$, where ρ is the nucleon density[4]. The final event rate is

$$\frac{d\sigma}{dE_R} = \frac{m_N}{\mu^2 v^2}\sigma_n^{SI} A^2 |F(q)^2| \tag{1.10}$$

where σ_n^{SI} is the Spin-Independent (SI) WIMP-nucleon cross-section, A is the number of nucleons, capturing the coupling strength of the WIMP to each nucleon. The Helm form factor, $F(q)$, is derived by assuming that the nucleon density distribution is well modelled by a solid sphere smeared with a Gaussian to produce a skin effect. This gives a form of

$$|F(q)|^2 = \left(\frac{3j_1(R_0 q)}{R_0 q}\right)^2 e^{-q^2 s^2} \tag{1.11}$$

where j_1 is the first spherical Bessel function. The parameters R_0 and s are determined from electron-nuclear scattering as charge should follow the mass distribution of the nucleus [86][87]. At higher momentum transfers, WIMPs can resolve the substructure of the nucleus, causing interference and reducing the rate. Essentially, the scatter loses coherence at higher momentum transfers. For spin-independent scattering, the momentum transferred cannot induce a transition in nuclear states, and the WIMP is typically assumed to have no accessible excited states, so the total scattering is suppressed. In cases where scattering is more complex, nuclear excitations can occur, such as the spin-dependent case in Ref.[88], or the scattering dark matter transitions into a heavier state in Ref.[89].

[4]Here, point like means that the mass of the mediator is great enough that it does not contribute to the dynamics.

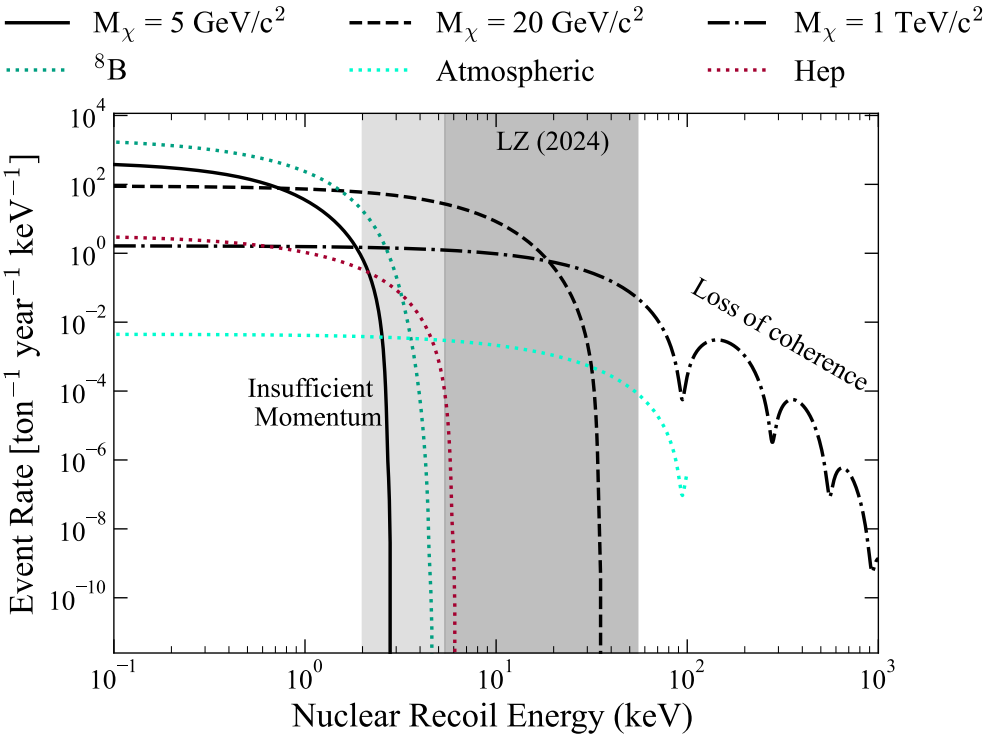


Figure 1.13: Plot of the nuclear recoil differential event rate for several WIMP masses and neutrinos from several sources in natural xenon. The light and dark regions indicate where the LZ 2024 WIMP search obtained $< 50\%$ and $\geq 50\%$ of events detected [90]. Labels indicate why the spectra drops off for WIMPs.

In Figure 1.13, increasing the mass extends the differential event rate out and decreases its total area as the flux drops. At high masses, and therefore high momentum, the form factor begins to suppress the rate. There neutrinos from several sources are plotted, the ^{8}B decay, Hydrogen-electron-proton (Hep) process in the Sun, and those produced in decays of products from cosmic ray scatters in the atmosphere. These neutrinos produce nuclear recoils identical to WIMPs, discussed in detail in Section 2.3.3. Naively, the only method to distinguish them would be through their spectra, but reconstruction of the trajectory of the recoiling nuclei would allow for further discrimination. These form the neutrino fog, a point where direct detection experiments can no longer scale sensitivity proportional to exposure. It is calculated for a range of targets in Ref.[91] and is unique for different target nuclei.

1.5.2 Sense of direction

The velocity distribution in Equation 1.4 is shifted by $\vec{v}_E$ as the orbital and rotational motion of the Earth affects the relative velocity between the halo WIMPs and the target. This shifts the spectrum to higher energies when the Earth moves parallel to the direction of the Sun through the halo. This oscillation is weak in amplitude and can only produce an oscillation in total event rate of the order $\sim$5%, requiring $\mathcal{O}(10^4)$ events to detect [92]. Despite its low amplitude, it is useful if one is looking for a trace against a dominant background over time,

as consistently observed by the DAMA-LIBRA experiment, which claims discovery of a dark-matter like trace [93]. This result has been excluded by direct detection experiments, and to 3.5 σ by a combined analysis of two experiments using the same target nuclei and methodology, COSINE and ANAIS [68].

If the direction of the recoiling nuclei can be reconstructed, one could utilise the direction of the solar system relative to the WIMP wind, which is in the direction of the Cygnus cluster, to reject backgrounds from the detector or neutrinos. Experiments are being designed for this purpose but require low density in order to create detectable tracks and, therefore, high volumes to match the fiducial masses of current generation direct detection experiments. For example, the CYGNUS experiment in the 10^3 m^3 configuration would at best match the SI WIMP-nucleon cross-section sensitivity of current xenon experiments at scales of $\mathcal{O}(1)$ m^3 [64]. There has been interest in measuring correlated effects in the scintillation of denser targets, such as xenon, with no observations yet [94]. This would likely require a single-phase spherical experiment, which would introduce other challenges for signal-background discrimination.

1.6 Conclusion

Dark matter is a fundamental phenomenon in our universe, it may be explained by three surviving paradigms; modified gravity, primordial black holes, and particle cold dark matter. Each is developing rapidly, with cold dark matter one of the leading candidates to explain cosmology and develop the standard model further. There are many candidates and methods of search, but leading candidates are WIMPs with direct detection. For masses greater than 10GeV/c^2 the best tools to search for WIMPs are noble element TPCs.

Chapter 2

The LUX-ZEPLIN Experiment

As seen in the previous chapter, LUX-ZEPLIN (LZ) is the most sensitive WIMP dark matter detector in the world for masses $> 10\text{GeV/c}^2$ and uses a xenon dual-phase Time Projection Chamber (TPC). Several competitor experiments exist, with plans for several of these to merge in the future to form a detector of the same principle with an order of magnitude larger target mass and improved background mitigation. This chapter will discuss the construction, operation, and data extraction of the TPC and supporting systems in LZ. The underlying physics of scattering and detection, as well as details of their modelling within LZ, are discussed next. Relevant backgrounds and additional considerations for WIMP search analyses are explained, as this is the primary focus of this thesis. The following discussion will focus on two WIMP search science runs: WS22 which refers to the first science run in Ref.[95], and WS24 which was the second WIMP search in Ref.[90].

2.1 The Detector and Operations

There are three detectors in LZ, the primary one is the TPC containing a target of 7 tonnes of Liquid Xenon (LXe), encapsulating it is an instrumented skin of 2 tonnes of LXe. Both are contained in the Inner Cryostat Vessel (ICV), the TPC has a Gaseous Xenon (GXe) portion at the top and the skin does not extend above the liquid surface. The ICV sits inside the vacuum of the Outer-Cryostat Vessel (OCV), providing insulation. The OCV is then surrounded by the Outer Detector (OD), consisting of acrylic tanks filled with liquid scintillator that are inside a larger water tank. A portion of the water tank is instrumented to form the OD, the rest acting as blind shielding. The OD Photo-Multiplier Tubes (PMTs) are placed on a support system with a UV reflective polymer curtain to increase its light efficiency. This section covers the basic principles of these detectors and their operations.

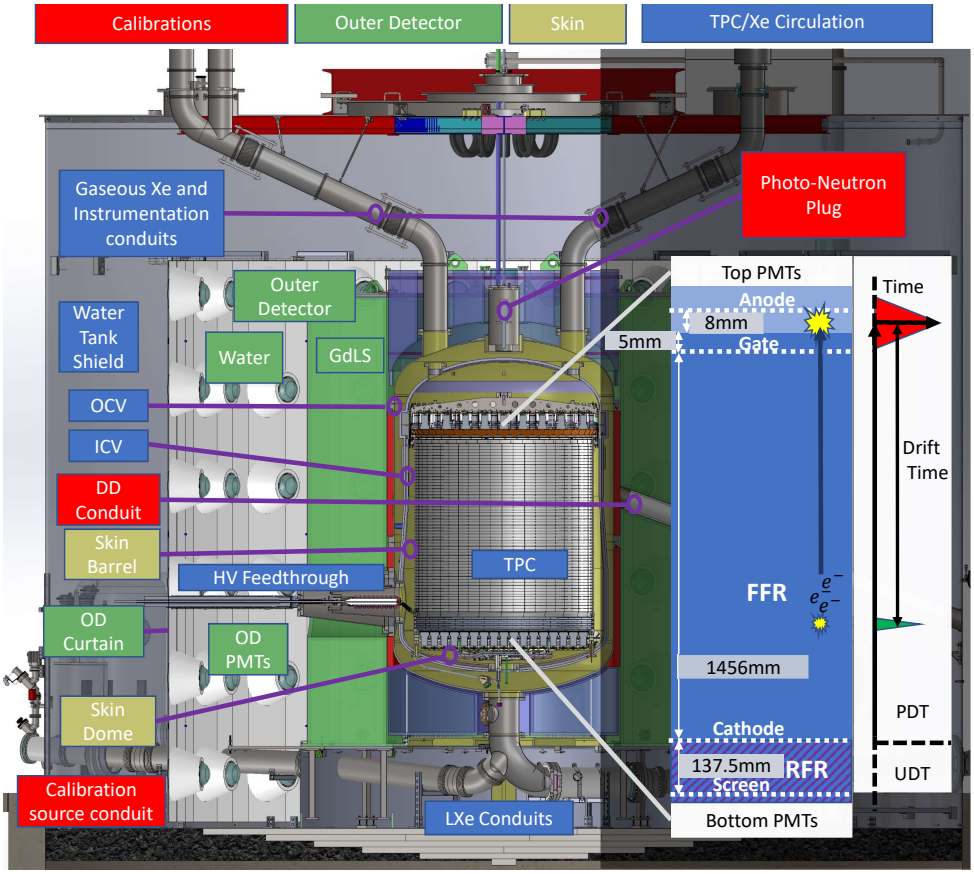


Figure 2.1: Detector diagram of LZ. The inset is a simple diagram of the detection of a scatter with key dimensions of the forward field region (FFR) between cathode and gate, reverse field region (RFR) between the screen and cathode, and the extraction region between anode and gate. The TPC geometry is such that it is equal in diameter and height to maximise light collection efficiency.

The detector is based in a cavern 4850 feet underground at the Sanford Underground Research Facility (SURF) in Lead, South Dakota. The large rock overburden shields the detector from cosmic radiation. The local bars, namely Greenfield Pub, Dakota Shivers, and Jail House Tap, maintain the spirits of the scientists through the cold harsh winters.

2.1.1 The Time Projection Chamber

TPCs collect the scintillation and ionisation channels by observing the prompt scintillation light (S1) and a delayed ionisation signal from the extracted electrons (S2). In LZ this is achieved by using two separate physical phases of the target, in this case; a bulk liquid with a drift field $\mathcal{O}$(100 V/cm) to drift electrons away from the interaction site and a thin region that transitions from liquid to gas that is subject to a field $\mathcal{O}$(5-10 kV/cm). The former is called the forward field region (FFR), and the latter is called the extraction region. Once electrons from ionisation drift from their interaction site in the FFR to the extraction region, the high field both overcomes the potential barrier between the two phases and allows each electron to

undergo proportional scintillation in the gas phase as the electrons drift to the anode. The S2 signal is defined as the observed light from this proportional scintillation which occurs after the prompt S1 scintillation. Both signals are collected by PMT arrays at the top and bottom of the TPC. The time between the signals, called drift time, gives the depth in the TPC and the site of the S2 can be reconstructed to give the observed horizontal position (x_{obs}, y_{obs}).

The FFR and extraction region were designed for fields of 300-600 V/cm and 10.2 kV/cm, respectively. These are established by three woven wire grids at different biases: a cathode at the bottom of the TPC at 50-100 kV, a gate below the liquid surface at -7 kV, and an anode between the liquid surface and the top PMTs at 7 kV. To produce a sufficient drift field, the cathode must have a significantly higher bias than the anode; therefore, a field screening grid is placed between it and the bottom PMTs to prevent the high field from interfering with the PMTs. It also requires a dedicated high-voltage (HV) feed through the skin. The field cage ensures uniformity in the FFR; this is achieved by sequential copper rings from the cathode to the gate, separated by two parallel chains of 1 GOhm resistors that drop the potential difference. The region below the cathode, called the reverse field region (RFR), allows for the detection of S1s in the LXe but S2 electrons are lost. The dimensions of these regions can be found in Figure 2.1. Sparking or electron emission from grid defects can both contribute significant background photons and electrons. As such, the HV components are designed to keep the maximum field strength of any point below a threshold, and the gate grid subject to the highest local field in the liquid was pacified with an acid bath to create a smooth surface [71].

The top and bottom PMT arrays contain 253 and 241 3-inch Hamamatsu R11410-22 PMTs, respectively [96]. To maximise S1 light collection efficiency, the FFR is of equal height and diameter and the TPC walls are constructed with polytetrafluoroethylene (PTFE) which has a reflectivity greater than 97% for xenon scintillation light [97]. It has been hypothesised that PTFE contributes to spurious photon and electron noise through fluorescence and photoionisation, but fluorescence to xenon scintillation light wavelengths have not been directly observed [98][99]. For LZ, the bottom PMTs collect most of the S1 due to total internal reflection at the liquid-gas interface and because the liquid refractive index better matches the quartz faces of the PMTs than that of the gas.

2.1.2 Circulation System

The circulation system removes impurities from the xenon and controls its temperature, pressure, and flow. This is achieved through two systems: an inline radon reduction system (IRRS) that removes radon from the GXe phase, and a hot zirconium getter that removes non-inert impurities from both the LXe and GXe phases[71]. Beta decays from radon daughter nuclei are a dominant source of electronic recoil (ER) background in the LZ WIMP search; the radon may come from uranium and thorium decays in the bulk of materials, but its dominant source

is direct surface contamination during manufacture and construction. The braided cabling for the instrumentation at the top of the detector is thermally connected to the outside of the cryostat, is in the gas phase, and their insulation can readily hold contaminants. Therefore, the IRRS only acts on the GXe phase because radon gas emanation is greater in a gas and from warmer material. It is not used in the LXe circulation as the IRRS itself emanates radon, and so it is only used where emanation is greatest in the TPC. Its function is analogous to gas chromatography, with synthetic charcoal as a stationary phase that holds radon long enough for it to decay [100][71].

The getter can remove >99% of non-noble gases such as O_2, N_2, and CH_4 from xenon in a single pass; at its high operating temperature the bound contaminants diffuse into the bulk to rejuvenate the surface [101]. In addition to being sources of radioactivity, these impurities are often more electronegative than xenon and therefore absorb electrons before they can form an S2 signal. Therefore, the purity of LXe can be measured directly using the electron lifetime value, which would ideally be greater than the maximum drift time. It has varied over the running of LZ, but is typically 5-9$\times$ the maximum drift time for the science run conditions of LZ [95].

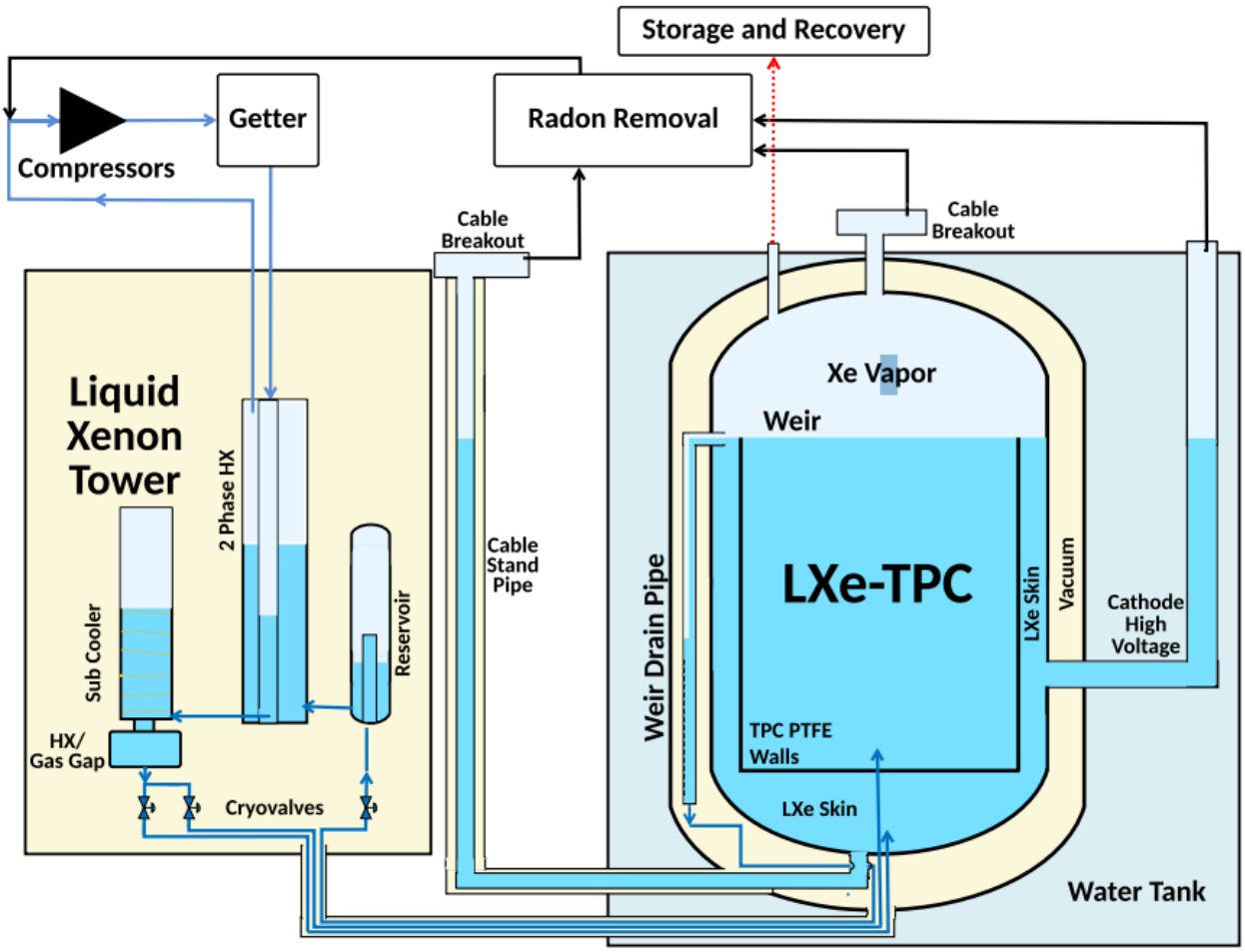


Figure 2.2: Diagram of the LXe and GXe circulation, adapted from Ref.[71].

The LXe is delivered to the ICV through two lines: one feeds directly into the skin via an inlet at the bottom of the ICV, the other splits into one central inlet and six surrounding inlets, distributed hexagonally, totalling seven TPC inlets. Each inlet has its own valve and heater for, in conjunction with the subcooler and cryogenic systems, control of the LXe temperature and flow. There are two main LXe flow states discussed in this work, measured in reference to the surface temperature as the GXe conditions are kept constant:

- **Mixing** When the temperature entering the LXe is greater than that of the liquid surface. This drives turbulent vertical currents to mix all the xenon.
- **Low-Mixing** Also referred to as 'the isolated state' in this work, this is when the input temperature is lower than the liquid surface temperature. This reduces vertical flow and produces a horizontally circular flow in the bottom half of the detector.

The mixing state is used so that the gaseous injection calibration sources may mix, but so does radon contamination from the TPC walls. The isolated state reduces the radial flow from the walls and skin to the centre of the TPC for some of the TPC regions. This yields a measured

decrease in the rate of radon, and its daughter, decays in the centre of the TPC. This was shown for WS22 in Ref.[102].

2.1.3 Vetoes: Skin and OD

The veto detectors consist of the instrumented skin and OD. Their purpose is to identify particles that scatter multiple times, but potentially only once in the TPC, as this cannot be a WIMP. The vetoes help to reduce background from internal radiation, such as long-lived radio-isotopes in detector components, and external gamma rays from the cavern rock. The decision to veto an event takes place at the data analysis stage, using the calibrated response of both veto detectors.

The skin detector is 4 cm thick consisting entirely of LXe and is instrumented with 3 different types of PMTs: 93 1–inch Hamamatsu R8520-406 PMTs, facing down attached to the outside of the TPC just below the liquid level; 20 2–inch Hamamatsu R8778 PMTs, viewing upwards from the bottom of the TPC and 18 2–inch R8778 PMTs, directly under the TPC's bottom PMT array with 12(6) facing out(in)-wards. Note that the top skin PMTs are below the TPC liquid surface level and unlike the bottom there is no instrumented dome above the top TPC PMTs. With no gas phase and a predominantly radial field, the skin can only observe scintillation light, so the ICV and all PMT struts are constructed from PTFE to optimise light collection efficiency. The skin is particularly helpful for detecting gamma rays, which would otherwise be absorbed by material between the TPC and OD, but may also detect scatters from neutrons. The field is large and more non-uniform in the short distance between the ICV and the field cage, requiring careful design of the field cage to prevent breakdown.

The OD consists of several tanks of a gadolinium-loaded liquid scintillator surrounding the OCV, with a primary goal of detecting neutrons and achieving >95% efficiency. To take advantage of the high energy released upon capture in gadolinium, it is doped such that the gadolinium neutron capture dominates over that of hydrogen in the scintillator [71]. It contains 120 8-inch Hamamatsu R5912 PMTs in a support structure in the water tank [96].

2.1.4 The Xenon Target

LZ and other experiments use liquid xenon for the following advantageous properties;

- **Inert** Its inert nature makes purification easier, reducing contaminants and allowing the use of calibration sources mixed into the target.
- **Density** Xenon, particularly in the liquid phase, shields the central volume from radiation from radio-isotopes both external and internal to the detector.

- **Kinematics** Maximum momentum transfer occurs for particles of similar mass, making the xenon nuclei ideal for WIMP masses $\mathcal{O}(\ 1 - 10^4)$ GeV/c^2, which are well motivated.

- **Radio Pure** With the exception of ^{124}Xe and ^{136}Xe, the former having the second largest half-life, natural xenon has no long-lived radio isotopes that contribute as background to a WIMP search.

- **Nuclear Structure** The odd number neutrons in ^{129}Xe and ^{131}Xe allow for searches for spin dependent interactions.

- **Neutrinoless Double Beta Decay** The even number of both protons and neutrons in ^{136}Xe and ^{134}Xe make them candidate isotopes for neutrinoless double beta decay searches [103].

- **Nuclear Structure** Alongside A^2 enhancement of the SI nuclear scattering cross-section, xenon nuclear properties allow for enhanced interaction rates compared to other target isotopes when considering non-relativistic effective field theory for WIMPs ≥ 100 GeV/c^2, particularly in neutron couplings [104].

- **Ionisation and Scintillation** Light and charge produced by particle scatters in xenon are significantly different for electron and nuclear scattering, this allows for discrimination.

There are several disadvantages to xenon compared to other target elements. For example, compared to argon, the timing of the scintillation light cannot be easily used to discriminate between ERs and NRs [70].

2.1.5 Data Processing

A particle may scatter multiple times, but for typical WIMP searches, a single-scatter (SS) requirement is imposed because of its low cross-section of the interaction. A notable exception in LZ is the search for ultra-heavy dark matter that scatter multiple times (MS) in the TPC, which was studied for WS22 in Ref.[105]. The ratio of the S2 and S1 signals provides discrimination between electronic recoils (ER) and nuclear recoils (NR), as ERs typically produce a higher number of electrons released per unit of energy; the source of this discrepancy is contentious and will be discussed in detail in the next section [106]. Due to its extremely low background and favourable kinematics, xenon TPCs focus on searching for WIMPs in the NR channel. Despite this, the ER channel can still provide competitive searches for a range of particles, such as ALPs, as seen in hints of a low energy ER excess that were later falsified in XENON1T in Ref.[107], and competitive limits set by LZ analysis of the first WIMP data Ref.[63].

The data acquisition in LZ operates on a trigger, predominantly based on prominent S2s but

other triggers exist for monitoring, such as the random trigger. A trigger prompts the collection of waveforms from all PMTs and detectors within 2 ms before and after it. [1] This data is later processed through the LZap software package to combine waveforms of an event into pulses and classify them, i.e are they S1 or S2, classify event types, such as SS, and derive Reduced Quantities (RQ). Examples of an RQ is the total light in a pulse or the drift time of a SS event. Xenon scintillation properties give S1s a typical width of $\mathcal{O}$(100 ns) with a fast rise time and exponential fall-time. The S2 signals typically have widths of $\mathcal{O}$(1-10 μs), determined mainly by overlapping Gaussian profiles of proportional scintillation in the gas phase, with additional structure from diffusion of the event's electron cloud. A SS is defined by an event with a single S2 and no other S1 between it and the prominent S1. A few examples of RQs are the total light collected of each pulse, pulse shape information, S2 reconstructed position, drift time, etc. The use of LZap reduces the quantity of data to allow for more interactive and easier processing by analysers using the Analysis LZ PACkAge (`ALPACA`) software package to apply cuts and derive higher-level information from RQs. For example, when imposing the vetoes by requiring no pulse in the veto coincident with the S1, the analyser will optimise the area thresholds and time windows of when a pulse in the skin or the OD warrants a veto to maximise efficiency, true vetoing of a gamma or neutron, and to minimise dead-time, false vetoing of a random TPC event.

2.1.6 Calibrations

Aside from the optical calibration systems, which use LEDs in all the detectors to measure PMT responses, there are four main calibration systems in LZ:

- **Gaseous Source Injection** Gaseous sources can be directly injected into the xenon circulation system, these fully mix with the xenon. There is some control in distribution through inlets, e.g can select the skin or TPC.
- **Calibration Source Deployment tubes (CSDs)** Three tubes that are equally spaced around the ICV, inside the OCV, they lower sources by a wire to precise positions measured with a laser.
- **Large source conduit** For large photo-neutron sources that require heavy shielding, a large conduit through the water tank to the OCV at the top of LZ is available.
- **DD Neutron Generator and Conduits** Two horizontal conduits, one angled at 20°, exist which can be evacuated of water to allow neutrons from a customised Adelphi DD-108 deuterium-deuterium generator (DD) to enter the TPC. These are located at the top of the TPC.

[1] The volume of data is further reduced by using pulse only digitization (PoDs) to read data from channels only over a set threshold.

Specific calibrations mentioned in this thesis are listed in Table 2.1. For the context of this thesis, it is worth noting here that the primary neutron source for calibrating the TPC NR response is the DD and not only localised to the edges of the detector, but also at the top.

Table 2.1: Sources discussed in this thesis. For defining the energy distribution of each particle in the energy column: γ are peaks, β are the end-points of the spectrum, and the DD is mono-energetic. Adapted from [108] table 1 unless cited otherwise.
$*$ - Though mono-energetic beam, the energy spectrum of events in the TPC is a smooth spectrum because of unresolved scatters along the neutron path.
† - Value signifies the end point of the α-n process to form neutrons, α do not escape the CSD.

Source	Particle	Energy [keV]	Target Response
Gaseous Source Injection			
^{3}H	β	18.6	WIMP search region ER.
^{14}C	γ	156	Above WIMP search region ER.
^{131m}Xe	γ	164	Similar ^{14}C but mono-energetic.
^{83m}Kr	β,β	32.1,9.4	Mono-energetic and both scatters within an event window
^{222}Rn	α	5.29 MeV	Less field dependent[90].
CSD			
AmLi	(α,n)	1.5 MeV†	NR response
AmBe	(α,n)	11 MeV†	and Vetoes
DD Generator			
Direct Mode	n	2.45 MeV $*$	NR response
Photo-neutron Conduit			
^{88}YBe	(γ,n)	152-n$*$	Threshold NRs, γ background
^{88}YMg	(γ)	n/a	Control γ background for ^{88}YMg

The author spent three months on-site during calibration campaigns between data taking for the LZ WIMP search in 2022, assisting with the maintenance of LZ, operations of the DD source and of the CSD systems. Throughout their PhD the author assisted in remote operations of the CSDs and detector monitoring.

2.2 Modelling Physics and Detector Response

In their work in Chapters 4 and Section 6, the author required a detailed understanding of the physics models and simulation tools in LZ. They are reviewed here, first covering the physics principles they are based on, then some technical implementation details, and finally their calibration.

For modelling the response of the LZ detector to a scatter there are four major packages used for simulations: the Noble Element Simulation Technique (NEST[2]), LZ light analysis Monte Carlo (MC) application (`LZLAMA`), Basically A Component Centric Analogue Response to AnyThing (`BACCARAT`), and the Detector Electronics Response (`DER`).

- **NEST** A software used to simulate the response of a noble element target to an energy deposit, it may also model some detector response like S1 coincidence and light collection.
- `BACCARAT` A custom GEANT4 interface that allows for simulation of particle transport and scattering throughout detector geometry. This is mainly used for sources that require particle transport, i.e. those that produce neutrons or gammas. It may also simulate the transport of photons produced by a scatter, the number produced is modelled by NEST.
- `DER` This is a parametric electronics response that creates fully simulated waveforms and events with the correct timing properties from information about photons hitting PMTs, typically from `BACCARAT`. It aims to fully replicate the raw data, which can then be processed through LZap.
- `LZLAMA` A LZ software designed to accurately replicate a required subset of RQs, tuned to data. It uses NEST, but includes processing of `BACCARAT` deposits and some specific detector responses like deposit and pulse clustering.

Details of `BACCARAT` and the `DER` are given in Ref.[109], NEST is an ever-evolving package found in Ref.[110]. `LZLAMA` is an internal package yet to be published, details can be found in Ref.[111] and Ref.[112]. The use of `BACCARAT` and the `DER` to produce full waveform simulations is often referred to as the slow chain, as it requires more resources than `BACCARAT` and `LZLAMA`, which is referred to as the fast chain.

2.2.1 Microphysics and NEST

Throughout both the fast and slow chain simulations the package NEST is used to model the microphysics. Here, microphysics refers to the production of electrons and photons by a particle scattering off the medium, referred to as an energy deposit. The subsequent detection

[2]NEST refers not only to the package, but the set of physics and detector models governing a noble element's response, so the italics have been dropped.

of the produced photons and electrons is the detector response, which NEST may also model. The electron and photon production is unique for specific types of scatters, namely electronic recoils (ERs) and nuclear recoils (NRs). These two types of recoils have different distributions of photons and electrons produced in the initial scatter per unit energy, most notably having distinct ratios of photons and electrons. The different ERs are beta decays, gammas, and electron captures and can be modelled by modifying the energy-dependent charge yield of the beta model. These models have been tuned to a range of detectors in different configurations, making them predictive within those conditions and physical regions of interest. NEST should not be naively extrapolated to regions outside of its tuning as, despite the fact that its models attempt to mimic the form of first-principle derivations, it is highly empirical.

The current NEST models for ER and NR are reviewed in Ref.[113] and Ref.[111]. In the sensitivity studies in Chapter 7 the author required a functional understanding of the behaviour of NEST. For the development and configuration of FlameNEST in chapter 4 and section 6, the author needed to modify the implementation of the NEST models. Therefore, here we describe the relevant model construction and parameters following a similar structure to that of Dr Robert James in Ref. [111].

Photon and Electron Production: First Principle ER and NR

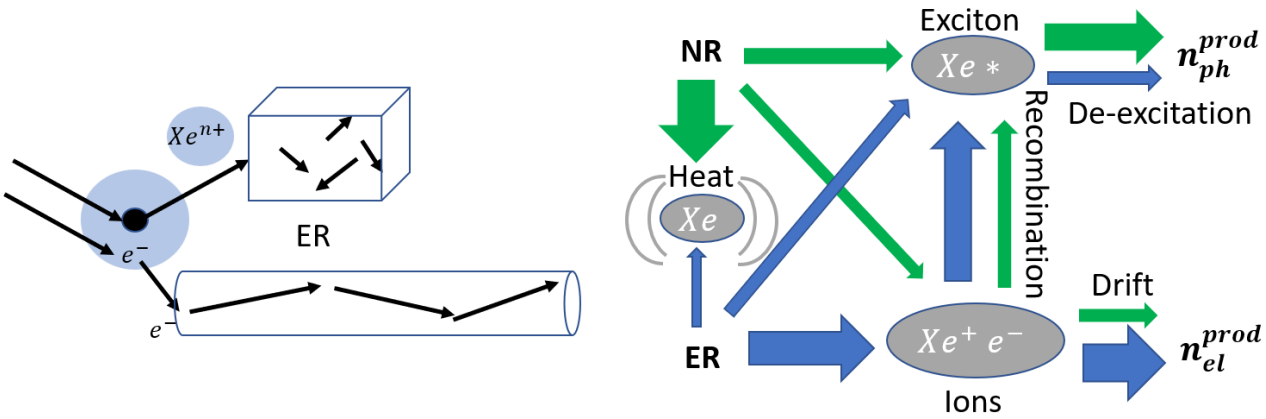


Figure 2.3: Simple diagram of differences between ERs. a long track assumed in Doke-Birks, and NRs, a high density deposit modelled using the Thomas Imel Box model (left). Diagram of the split between excitation, ionisation, and heat and production of photons and electrons adapted from code by Dr Vetri Velan (right).

The physics of scattering is an interplay between the density, shape, and primordial ionisation or excitation of the track. In the following discussion, the track of a scatter is the spatial distribution of ions and electrons left after the xenon ion, in NR, or electron, in ER, or X-ray produced from a particle scatter interacts with the medium. For a beta ER and an NR, the scattered particles are an electron and an ionised xenon atom, respectively. Both are

charged species and as they recoil they will either excite or ionise nearby xenon atoms to form a distribution of excitons, ions, and electrons. In the case of ionisation, there is a chance that the electrons from ionisation will recombine, which decreases with the applied electric field. At high densities, Penning quenching increases the electron-photon, or S2-S1, ratio by converting excitons to ions and electrons [114]. The interaction of a charged species, S, with xenon can be written as follows,

Excitation & Penning Quenching
$S^+ + Xe \rightarrow Xe^* + S^+$
$Xe^* + Xe \rightarrow Xe_2^{*,\nu}$
$Xe_2^{*,\nu} + Xe \rightarrow Xe_2^* + Xe$
$Xe_2^* \rightarrow 2Xe + \gamma$
$Xe^* + Xe^* \rightarrow Xe^+ + Xe + e-$

Ionisation & Recombination
$S^+ + Xe \rightarrow Xe^+ + S^+ + e^-$
$Xe^+ + 2Xe \rightarrow Xe_2^+ + Xe$
$e^- + Xe_2^+ \rightarrow Xe^{**} + Xe \rightarrow Xe^* + Xe + heat$
$Xe^* + 2Xe \rightarrow R_2^* + R + heat$
$Xe_2^* \rightarrow 2Xe + \gamma$

adapted from Ref.[106]. The Xe^* and Xe_2^* denote an excited xenon atom and dimer, the latter dominates as the interaction with other xenon atoms happens over shorter time periods than scintillation. Notice the additional xenon atoms in the formation and relaxation of the dimers, $Xe_2^{*/**/+}$, required to conserve momentum and energy. There is a rotational excitation in the dimer, $Xe_2^{*,\nu}$, which is lost in the form of heat with another xenon atom. Due to elastic collisions with the medium, NRs lose a significant portion of their energy to heating in a process known as Lindhard quenching. This also causes the scattering to follow a branching structure instead of a straight track. Taking this into account, the observed number of photons and electrons per unit energy has led the inferred ratio of excitons to ions, $\frac{n_{ex}}{n_{ions}}$, to be 0.2 in ER and 1 in NR. Differences in this ratio, in practice measured as S1/S2, for ER and NR interactions allows the experiment to discriminate between the two. Recombination reduces the number of surviving electrons and increases excitons, effectively increasing the S1/S2 ratio. In ERs, recombination appears to be significant; this is inferred since the photon yield is reduced by 2/3rd using high drift fields, as electrons are removed prior to recombination. Compared to ERs, NRs show a weaker dependence on the field, implying that recombination plays a lesser role [106].

The origin of the difference in the inferred exciton ratios for ER and NR, and the role of recombination, is still contentious. Two possibilities discussed in Ref [106] for NRs are as follows:

- In collisions between the recoiling xenon ion and the medium atoms, the atomic orbitals may be altered to increase the probability of excitation over ionisation.
- Regions of very high ionisation density from the dense track structure force electrons to recombine, increasing the exciton ratio whilst having a weak field dependence as these electrons would be trapped.

Regardless of the process, it was found in Ref.[115] that NRs are well-modelled by assuming an exciton ratio of 1, Lindhard quenching, and the Thomas-Imel box model (TIB) for recombination up to an energy of $\mathcal{O}$(100 keV).

The TIB model is for dense tracks and assumes a uniform distribution of 1-to-1 ions and electrons in a cubic geometry, as shown in Figure 2.3, applicable for NRs. First derived in Ref.[116], it gives the fraction of electrons that recombine to be

$$r = 1 - \frac{ln(1+\zeta)}{\zeta}, \qquad \zeta = \frac{N_0 \alpha}{4L^2 \mu_- \epsilon_L}, \tag{2.1}$$

where N_0 is the density of the ions and electrons, μ_- is the mobility of the electron, L is the dimension of the cube, and ϵ_L is the applied field. If the number of ions and electrons is fixed, substituting $L = (N_0/\rho)^{2/3}$ shows that $\zeta \propto \rho^{2/3}$ where ρ is the density of the deposit. Dense tracks of any nature may follow TIB for recombination, including ERs below 15 keV [117][113].

For energies above 15 keV, the ERs appear to follow Doke-Birk's law [117]. Birk's law for inorganic crystal scintillators explains recombination in terms of linear energy transfer (LET), $\frac{dE}{dx}$, assuming ions and electrons distributed in a cylinder as indicated in figure 2.3[118]. Derived from Birk's law in Ref.[119], the Doke-Birk's law accounts for the larger thermalisation length in liquid noble targets, which adds a large LET limit of η_0 as even at high ionisation densities electrons can escape. With similar notation as in Ref.[117], but following the original more closely, it is

$$r = \frac{AB\frac{dE}{dx}}{1 + B\frac{dE}{dx}} + \eta_0, \;\; A + \eta_0 = 1,$$

where A and B are constants that are given in a fit to observed data and η_0 was found by assuming that all electrons recombine in the large LET limit. Of course, at that point the TIB regime occurs. Given r and the initial exciton ratio, one can write the number of electrons produced, n_{el}^{prod}, per unit energy, known as charge yield, as

$$Q_y = \frac{n_{el}^{prod}}{E} = \frac{1}{WL(E)} \frac{1-r}{1 + \frac{n_{ex}}{n_{ion}}} \tag{2.2}$$

where $L(E)$ is the Lindhard factor and W is the work function of the xenon, the energy required to produce one quanta, ion or exciton. Their functional forms can be found in NEST in Ref.[113], [117], [120].

Photon and Electron Production: NEST

As ERs and NRs have unique processes in a scatter, NEST treats them separately when modelling photon and electron production. NEST relies on the core assumption that the total quanta in the microphysics processes are conserved such that

$$n_q^{prod} = n_{el}^{prod} + n_{ph}^{prod} = n_{ion}^{prod} + n_{ex}^{prod}.$$

where n_j^{prod} indicates the produced number of a species, j, which could be q,el,ph,ion, and ex indicated the total quanta, electrons, photons, ions, and excitons, respectively. Therefore, in calculating the total quanta and electrons produced, the photons produced may be obtained. The two simulation chains follow

ER	**NR**
$\overline{n_q} = \frac{E}{W}$	$\overline{n_q} = \alpha E^{\beta}$
$n'_q \sim Norm(\overline{n_q}, F_q \overline{n_q})$	$n'_{ion} \sim Norm(\overline{n_q}\alpha, \sqrt{F_{ion}\overline{n_q}\alpha})$
$n_q^{prod} = n'_q \Theta(n'_q)$	$n_{ion}^{prod} = n'_{ion}\Theta(n'_{ion})$
$n_{ion}^{prod} \sim Binom(n_q^{prod}, \alpha)$	$n'_{ex} \sim Norm(\overline{n_q}\alpha R, \sqrt{F_q \overline{n_q}\alpha})$
$n_{ex}^{prod} = n_q^{prod} - n_{ion}^{prod}$	$n_{ex}^{prod} = n'_{ex}\Theta(n'_{ex})$
	$n_q^{prod} = n_{ex}^{prod} + n_{ion}^{prod}$

where $\overline{n_j}$, n_j, and n_j^{prod} indicated the mean, intermediate, and final values for the j^{th} species. Primes are used to indicate an intermediate value. Both ERs and NRs derive the mean number of total quanta, $\overline{n_q}$, from the energy of the deposit, with NRs having a different functional form due to Lindhard quenching. For both, α represents the ratio of mean ions to mean total quanta, but NRs require an additional ratio to obtain the mean excitons given by

$$\alpha = \frac{1}{1 + \frac{\overline{n_{ex}}}{\overline{n_{ion}}}} = \frac{\overline{n_{ion}}}{\overline{n_q}} \qquad\qquad R = \frac{1}{\alpha} - 1 = \frac{\overline{n_{ex}}}{\overline{n_{ion}}}$$

The functional form of the exciton ratio and mean yields differ as well. For NRs the exciton ratio depends on the total quanta derived from the charged and light yield, while for ERs it is a simple function of energy.

The ERs behave intuitively, fluctuating the total quanta around the expected and then deriving the ions and subsequent excitons produced by conserving the total quanta. The NRs are slightly more complex, as both ions and excitons are derived from the mean quanta, and then a final total quanta is derived from them. This means that ERs have a strict anti-correlation between excitons and ions, but there is a freedom for this to be weaker in NRs as seen in previous results[113].

Following exciton and ion production, the recombination process is then modelled through a normal distribution using the recombination fraction and the number of ions produced. The recombination fraction is derived using Equation 2.2,

$$\frac{\overline{n_{el}}}{E} = \frac{\overline{n_q}}{E}(1 - r)\alpha$$

where the above definition of α has been substituted in. The recombination fraction is then

$$r = 1 - \frac{\overline{n_{el}}}{\overline{n_q}}\frac{1}{\alpha}$$

recalling that r is the ratio of electrons that recombine. This represents a probability of recombination, so the mean and width of surviving electrons can be given by that of a binomial process

$$\mu = n_{ion}^{prod}(1-r)$$
$$\sigma_r^2 = n_{ion}^{prod}(1-r)r.$$

In order to accommodate the positive skew on the observed number of electrons observed in previous experiments, motivated to be from microphysics and not some detector effect in Section V of Ref.[121], a skew normal with width and mean corrections is applied. This involves two new parameters that are distinct for ERs and NRs: a width parameter, ξ, and skew parameter, ω. These parameters depend on the mean yields, for ERs the drift field as well, and a distinct set of free parameters not discussed here. The width is adjusted by

$$\sigma_r^2 = n_{ion}^{prod}(1-r)r + (\omega n_{ion}^{prod})^2.$$

The addition of skewness incurs two correction factors to the mean and with $\Delta\mu$ and $\Delta\sigma$ given in Ref. [121] and Ref.[111]. The final model for recombination is

$$n_{el}^{prod} = \max(\text{SkewNormal}(\mu - \Delta\mu, \frac{\sigma_r}{\Delta\sigma_r}, \xi), 0). \tag{2.3}$$

As the number of electrons cannot exceed the number of ions, the following condition is imposed:

$$n_{el}^{prod} \rightarrow \begin{cases} n_{el}^{prod} \text{ for } n_{el}^{prod} > n_{ion}^{prod} \\ n_{ion}^{prod} \text{ for } n_{el}^{prod} \leq n_{ion}^{prod} \end{cases} \tag{2.4}$$

Finally, the photons are derived from conserving total quanta but ensuring they are always greater than the number of excitons as recombination can only add photons.

$$n_{ph}^{prod} = \begin{cases} n_q^{prod} - n_{el}^{prod} \text{ for } n_q^{prod} - n_{el}^{prod} > n_{ex}^{prod} \\ n_{ex}^{prod} \text{ otherwise} \end{cases} \tag{2.5}$$

Photon and Electron Production: Mean Yields and Functions

The above omitted the functional form of the mean yields, and other model parameters. These are again unique for ERs and NRs as well. This section aims to describe the key steps without going into detail.

ER

$$\overline{n_q} = \frac{E}{W}$$

$$\overline{n_{el}} = Q_y^{ER}(E, \mathcal{E}, \vec{m}) \times E$$

$$Q_y^{ER}(E, \mathcal{E}, \vec{m}) = m_1 + \frac{m_2 - m_1}{\left[1 + \left(\frac{E}{m_3}\right)^{m_4}\right]^{m_9}} + m_5 - \frac{m_5}{\left[1 + \left(\frac{E}{m_7}\right)^{m_8}\right]^{m_{10}}}$$

$$\frac{\overline{n_{ex}}}{\overline{n_i}} = \alpha_0 erf(E/20)$$

NR

$$\overline{n_q}' = \alpha E^{\beta}$$

$$\overline{n_{el}} = Q_y^{NR}(E, \mathcal{E}) E$$

$$\overline{n_{ph}} = (\overline{n_q}' - \overline{n_{el}}) \left[1 - \frac{1}{1 + (E/\theta)^{\iota}}\right]$$

$$\overline{n_q} = \overline{n_{ph}} + \overline{n_{el}}$$

$$\overline{n_i} = \frac{4}{\varsigma(\mathcal{E}, \rho)} (exp[\frac{\varsigma(\mathcal{E}, \rho) * \overline{n_{el}}}{4}] - 1)$$

$$\overline{n_{ex}} = \overline{n_q} - \overline{n_i}$$

where the drift field, $\mathcal{E}$, has been introduced. The dual sigmoid nature of the ER charge yield allows for an interpolation between a TIB response and a Doke-Birks one. The exciton ratio here is a constant representing the underlying ion-exciton ratio in the track, and field-dependence is managed by other functions, predominantly the charge yield. The ER responses are controlled by, m_i, for betas these are m_i^{β} in Table 1 of supplementary material of the NEST review in Ref. [113] and are not explicitly used in this work, and so are omitted here. Gammas will have a unique LET and require a unique response that suppresses the charge yield, enhancing recombination. This is achieved by using the same simulation as ERs but using a unique set of m_i^{γ} parameters provided in Table 2 of the supplementary material Ref. [113].

Once again, the NR model has been given more freedom at the cost of increased complexity. The strict anti-correlation between mean yield photons and electrons has been broken with a sigmoid in the calculation of mean photons, the third step in Equation 2.2.1. The author notes that the mean number of ions for NRs is directly derived from the TIB model in Equation 2.1, where the ions replace the ion density and a new scaling function that retains the xenon density and drift field dependence, given by

$$\zeta_{TIB} = \overline{n_i} \left[\varsigma(\mathcal{E}, \rho) = \gamma\, \mathcal{E}^{\delta} \left(\frac{\rho}{\rho_0}\right)^{\upsilon}\right]. \tag{2.6}$$

The NR charge yield also depends on this scale function to provide field dependence, it is given by

$$Q'_y(E, \mathcal{E}, \rho) = \frac{1}{\varsigma(\mathcal{E}, \rho)} \frac{1}{(E + \epsilon)^p} \left[1 - \frac{1}{1 + \left(\frac{E}{\zeta}\right)^{\eta}}\right]. \tag{2.7}$$

the second term is the functional form at characteristically low energies, and the third allows for sub-keV flattening. The parameters of the NR model functions, and their role in the response, are given in Table 2.2. These are used in Chapter 4.

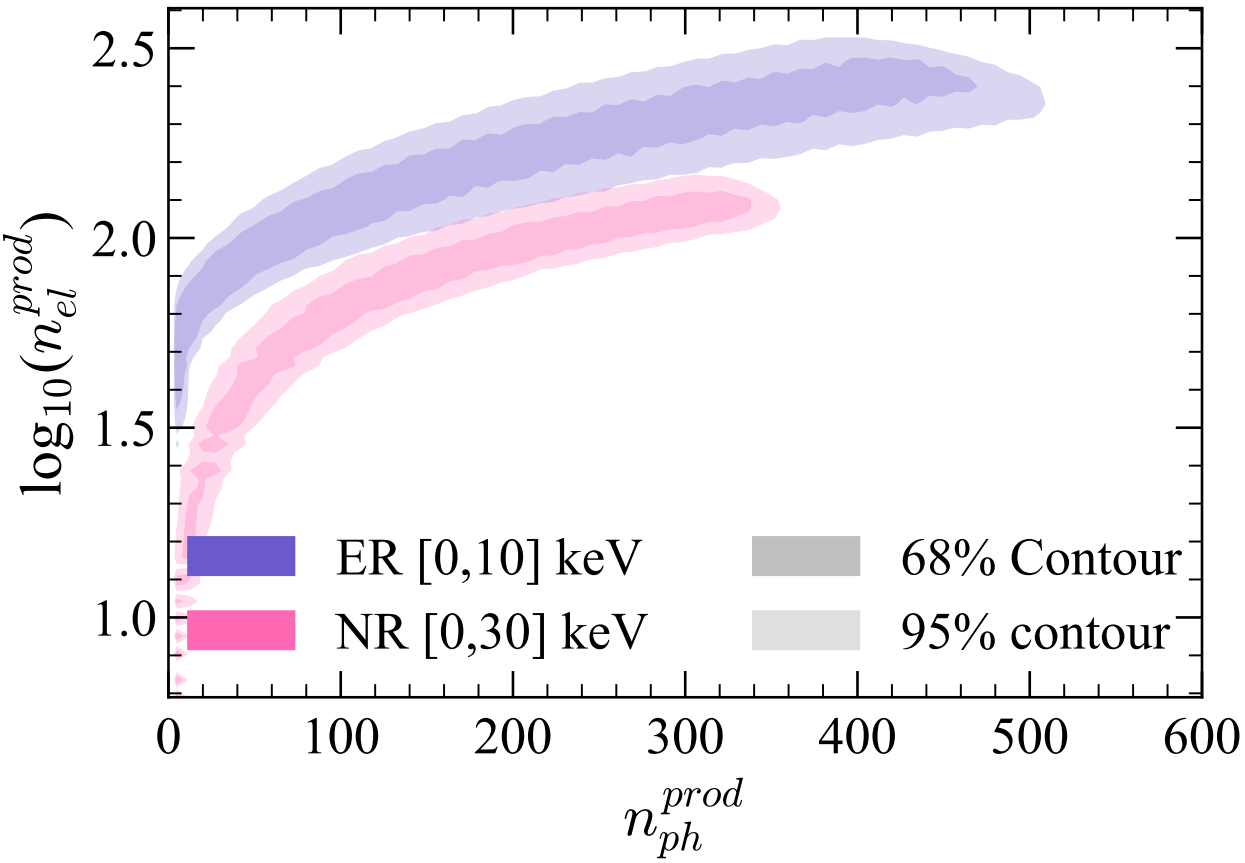


Figure 2.4: The ER and NR simulations for flat energy spectra. Note that ERs require significantly less energy to produce above threshold scatters due to NRs being suppressed by Lindhard quenching.

Table 2.2: NR charge and light yield parameters, taken from Table 3 in supplementary material of Ref.[113] with adapted text.

Parameter	Value	Role
a	$11^{+2.0}_{-0.5}$	Total quanta normalisation.
b	1.1 ± 0.05	Total quanta energy-scaling exponent.
γ	0.0480 ± 0.0021	Field scaling normalisation.
δ	-0.0533 ± 0.0068	Field-dependence exponent.
υ	0.3	Density-dependence exponent.
ϵ	$12.6^{+3.4}_{-2.9}$ keV	Point at which p drives energy dependence.
p	0.5	Controls energy dependence E> $\mathcal{O}(\epsilon)$.
ζ	0.3 ± 0.1 keV	Sub-keV behaviour scale.
η	2 ± 1	Control sub-keV shape.
θ	0.3 ± 0.05 keV	Sub-keV Light yield behaviour change scale.
ι	2 ± 0.5	Controls sub-keV light yield shape.

Electron and Photon Detection in NEST

After the photons and electrons are produced, they need to be detected. This section describes and compares the independent S1 and S2 simulation procedures in NEST. It is worth noting that NEST has the ability to interface with other programmes for this step, but that was not used in this work.

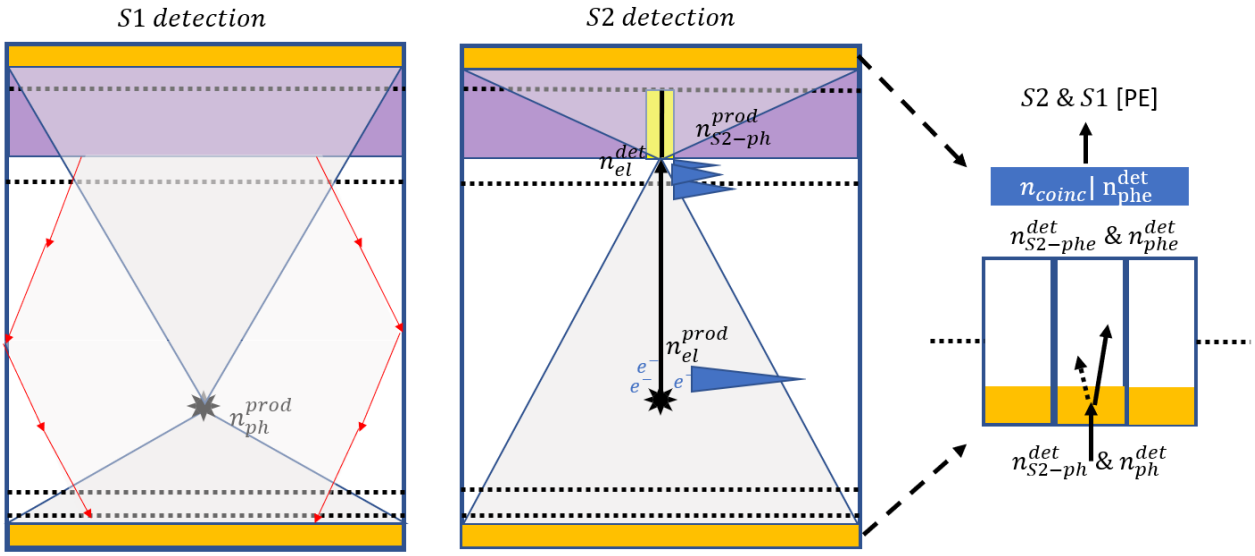


Figure 2.5: Diagram of light/charge collection for S1s/S2s. The red arrows intend to indicate that the top-bottom asymmetry of an S1 can be indicative of its position, but this is not simulated in NEST in LZ at the time of this work.

S1 detection begins with a Binomial modelling of the probability of the PMT arrays detecting the photons produced,

$$n_{ph}^{det} \sim Binom(n_{ph}^{prod}, p_{det} = g_1 f_{s1}(x,y,z) p_{coin}(n_{ph}^{prod})).$$

The probability of detection has three components; the total detection efficiency, g_1, the impact of varying light collection efficiency (LCE) over the volume of the detector, which must be a function of the 3D position of the interaction site $f_{s1}(x, y, z)$, and the probability that, for a given number of photons, at least one photon is detected in 3 separate PMTs, $p_{coin}(n_{ph}^{prod})$, from a parametric model in NEST. The latter ensures the coincidence requirement is achieved. Due to the path of the photons, this should depend on the interaction site, on the number of coincident PMTs, and positions of the PMTs that detected a signal. Due to the total internal reflection at the liquid-gas boundary, as indicated in figure 2.5, the strongest factor is the vertical position of the interaction. There is on-going work to implement a separate model in NEST or `LZLAMA` that simulates coincidence as a variable, potentially with per-channel information or at least S1 top-bottom asymmetry, but this has yet to be applied in analysis.

After detection in PMTs the production and detection of photoelectrons, phe, is modelled by a constant double-photoelectron probability, p_{DPE}, and a detection efficiency, p_{det}. that is a function of the single photoelectron detection efficiency, ϵ_{SPE}, the number of PMTs, N_{PMT}, and photoelectrons produced.

$$n_{phe}^{prod} \sim n_{ph}^{det} + Binom(n_{ph}^{det}, p_{DPE})$$
$$n_{phe}^{det} \sim Binom(n_{phe}^{prod}, p_{det}(\epsilon_{SPE}, N_{PMT}, n_{phe}^{prod}))$$

The S2 chain has similar components for photon detection, but these first need to be generated through modelling the drift and extraction of electrons with a Binomial,

$$n_{el}^{det} \sim Binom(n_{el}^{prod}, p_{det} = \eta_{ext} \exp(-\frac{\text{drift time}}{\tau_e})) \tag{2.8}$$

This is a function of an extraction efficiency, $\eta_{extract}$, which NEST derives from the extraction field and gas properties, the drift time, and the electron lifetime, τ_e, modelling the loss of electrons to electronegative impurities. The simulation must then model the generation of S2 photons, $S2 - ph$, for the detected electrons. This is modelled through a gain and smearing term, again derived from a NEST function of the gas-phase conditions.

$$n_{S2-ph}^{prod} \sim Norm(n_{el}^{det}\mu_{el}|\sigma_{el}\sqrt{n_{el}^{det}}) \tag{2.9}$$

After production, the S2 photon detection process is similar to that of S1 photons except that their LCE, $f_{S2}(x, y)$, does not depend on depth, as all photons are produced in the small extraction region near the PMTs. The S2 photon gain in the TPC, g_1^{gas}, will be substantially different from that of S1s. The typical single-electron size for LZ conditions is $\mathcal{O}(10)$s of photons detected, so coincidence and SPE effects are ignored. This means after photon detection only the DPE effect is considered to produce the S2 photoelectrons detected,

$$n_{S2-ph}^{det} \sim Binom(n_{S2-ph}^{prod}|p_{det} = g_1^{gas} f_{S2}(x, y)) \tag{2.10}$$
$$n_{S2-phe}^{det} \sim Binom(n_{S2-ph}^{prod}|p_{DPE}) + n_{S2-ph}^{prod}. \tag{2.11}$$

At this point the procedure for modelling the final signals are the same for S1 and S2 and measured in photoelectrons detected (PE), though they can be converted to photons detected (phd) by a constant of $\frac{1}{1+P_{DPE}}$. The model for the final signal is

$$S1/S2 \sim Norm(n'^{det}_{phe}|\mu_{spe}n'^{det}_{phe}, \sigma = \sqrt{\sigma^2_{spe}n'^{det}_{phe} + \lambda'^2 n'^{det^2}_{phe}}) \tag{2.12}$$

where n'^{det}_{phe} indicates either S1 or S2 photons detected. The mean single-photon electron size, μ, and width, σ, determine the average PMT response and can be adjusted independently for each channel. There are also independent linear noise terms, λ', for each channel that can arise from detector/analysis effects.

For an analysis which only uses (S1,S2) observables, it is useful to introduce a set of corrected variables (cS1,cS2) which unfold any position dependence. These are given by

$$cS1 = S1\frac{f_{S1}(0,0,dt_c)}{f_{S1}(x_{obs},y_{obs},dt_c)} \tag{2.13}$$

$$cS2 = S2\frac{f_{S2}(0,0)}{f_{S2}(x_{obs},y_{obs})}\exp(\frac{\text{drift time}}{\tau_e}) \tag{2.14}$$

where dt_c is the drift time of the centre of the detector. This normalises signals to a central point, removing any position dependence. Here, the observed position (x_{obs}, y_{obs}, drift time) is explicitly used, as LZ utilises this when correcting the observables but uses the true position of the deposit when simulating the S1/S2. The cS2 is often reported in $\log_{10}$, as shown in Figure 2.15. With these corrected variables, it is possible to reconstruct the energy of an event using g_1 and $g_2 = g_1^{gas}\mu_{el}\eta_{ext}$.

$$E_{ER} = W(\frac{cS1}{g_1} + \frac{cS2}{g_2}) \tag{2.15}$$

$$E_{NR} = [(\frac{cS1}{g_1} + \frac{cS2}{g_2})/\alpha]^{\frac{1}{\beta}}.$$

Typically, only the ER definition is used as these two observables are monotonic and it does not require microphysics parameters. The parameters discussed in this section will be referred to as detector parameters from here on. The work function, W, and NR equivalents, α and β are the microphysics parameters from the previous section.

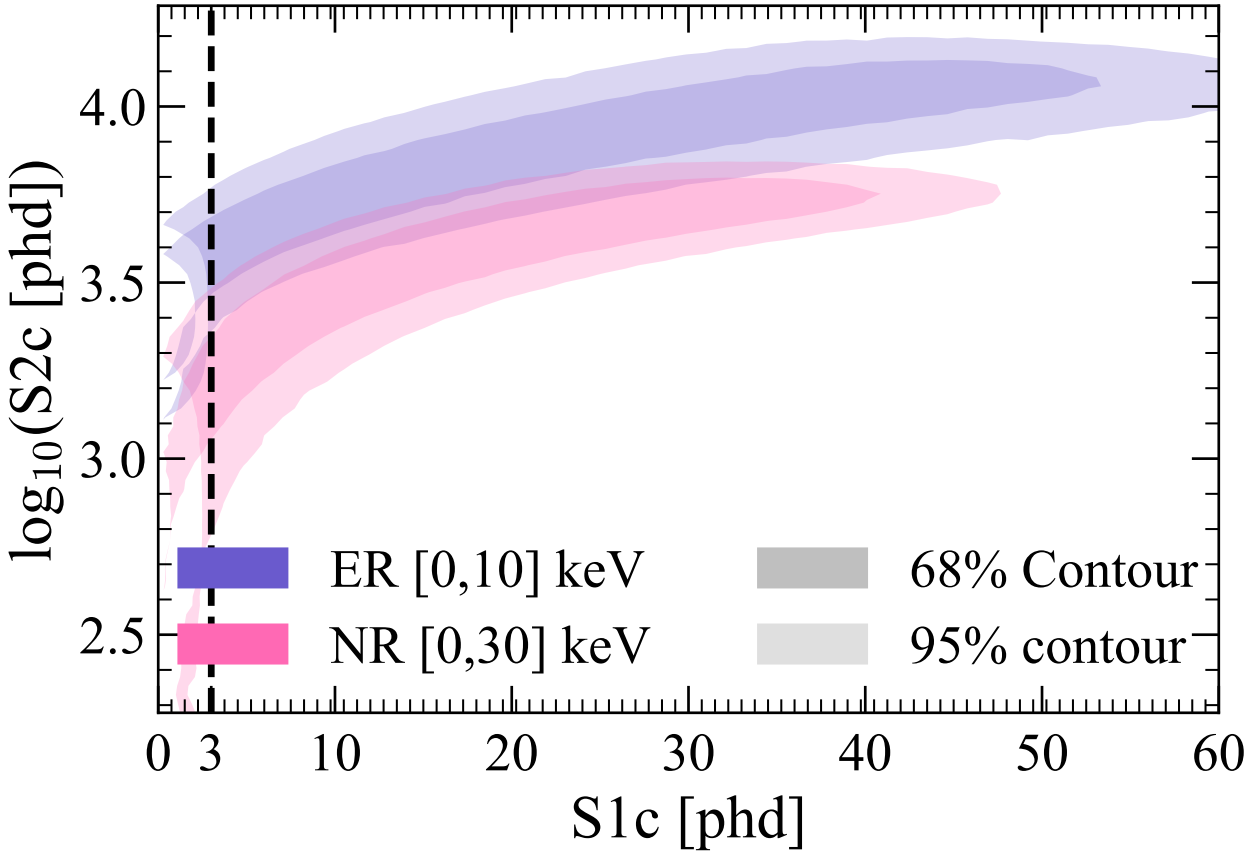


Figure 2.6: The ER and NR bands for flat energy spectrum sources for the same data as Figure 2.4 but carrying out the above sampling steps. Note that the 3-coincidence requirement is applied but events below it do occur, these would be events in low-LCE regions that have upward fluctuations at the S1 detection steps.

2.2.2 `LZLAMA` and `BACCARAT`

The `BACCARAT` package contains a model of the entire LZ geometry built using GEANT4, it can use physics lists modified for low energies and standard or custom generators to track a range of particles from defined sources through the detector. The software records each particle interaction and information like the position, energy deposited, particle information, and volume in which the interaction occurred, to name a few.

The `LZLAMA` package contains custom generators for simple models or can take `BACCARAT` particle track information as inputs. For example, to model ^{214}Pb one only needs the position of the decays and the known energy spectrum, but for a neutron generated from spontaneous fission all deposits along its track are required to reconstruct the final energy and whether or not it is classified as a SS, MS or vetoed. It adds several features, from integrating a dedicated electric field model to adding clustering of deposits that may be merged into single events due to the resolution of the detector. The latter feature involves modelling pulse timing, which can be useful for many studies but as seen in Section 6 has little impact on the WIMP search models.

The electric field model used in LZLAMA is imported from a custom package called LZ Electric Fields (LZEF).

A primary feature of LZLAMA is the modelling of position reconstruction in the TPC; this utilises the drift map derived from the LZEF model independently to determine the position at which the S2 would be extracted. Due to the finite number of PMTs on the top array there is a finite position resolution, so a Gaussian smearing is applied. This smearing may be a function of the S2 size as a larger S2 should be easier to reconstruct. Further discussion on this smearing and its impact for WS24 are discussed in Section 6.2.1. In LZLAMA the true position of the deposit is used to simulate the size of the raw S1 or S2 signal and the observed, post-smearing, position is used to produce the corrected signals. If the LCEs are derived from the observed position in the data, it would be better to use that consistently, or to use true position and simulation-derived LCEs. While not being completely correct, simulations do show consistency with data, and it is therefore a small effect that may require more careful consideration in larger detectors.

For deposits from BACCARAT, LZLAMA is able to identify and model MSs from multiple energy deposits within the TPC, as well as the response of the OD and Skin to deposits in their volume, all of which is essential for modelling backgrounds from radiogenic neutrons and gammas. The skin model uses NEST alongside a dedicated S1 LCE and field map. As there is no extraction region, there is no S2 but the field is considerably larger in the short distance between the field cage and ICV walls and needs to be accounted for in S1 modelling. The OD also collects only scintillation light but has a dedicated response model based on a modified Doke-Birks response [109]. The timing of the deposits is stored by BACCARAT and can be used later to determine the impact of different veto configurations in a given analysis.

2.2.3 Calibrating the Model

The author describes some of the technical details of calibrating the models here to reference later. The work described here was carried out by the calibrations team[3] and their predecessors.

The NEST models depend on LCEs, so these must be derived first. In WS22 a ^{83m}Kr source was used to measure S1 and S2 LCE; however, as the charge yield for beta ERs has a strong dependence on the field, this source would convolve LCEs and drift field. Therefore, in WS24 an α source, ^{222}Rn, whose charge and light yield is weakly dependent on the drift field, was used. Additionally, this source has a longer half-life, allowing for calibrations to cover longer periods. For both maps a finite binning is used throughout the observed position, in each bin

[3]Calibrations team: Greg Rischbieter (University of Michigan), Chami Amarasinghe (University of California, Santa Barbara), Erin Barillier (University of Michigan), Yongheng Xu (find out), Amardeep Chawla (Royal Holloway University, London), Ellie Bishop (University of Edinburgh), Mark Murdy (University of Massachusetts), Asher Kaboth (Royal Holloway University, London)

selections in observables such as S1c/S1c identify the calibration source. The maps are then formed by deriving the mean S1/S2 for each bin, then renormalising this to the central point. The WS22 map is shown in Figure 2.7. For S2s, the drift time is used to remove the impact of the electron lifetime, which could produce a non-Gaussian distribution within each (x_{obs},y_{obs}) bin.

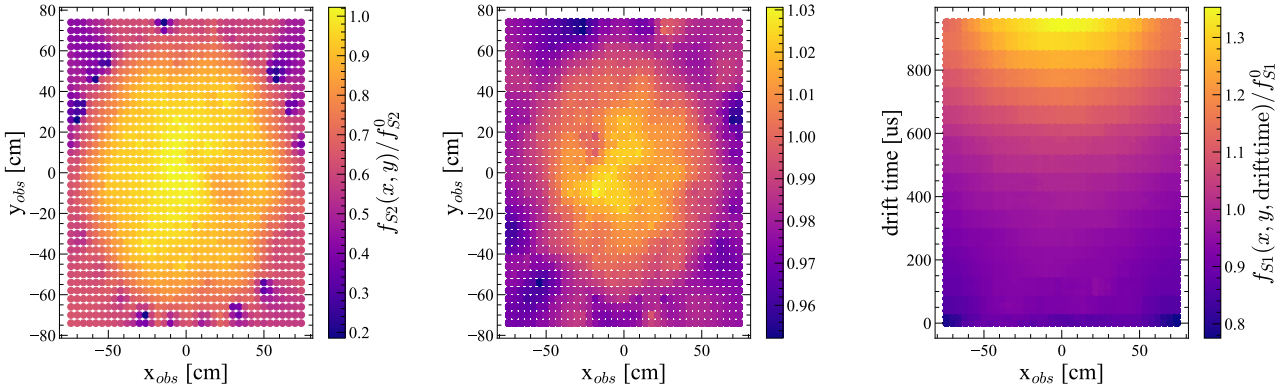


Figure 2.7: Light collection efficiencies of S2 (left) and S1s (middle, right) in WS22, the S1 maps share a colorbar. Note that high drift times are near the bottom of the detector, where S1 LCE is greatest.

The microphysics and detector parameters can be tuned with in situ backgrounds and calibration sources. There are many ways to tune these parameters, in WS24 the primary constraints were formed by minimising χ^2, derived from the means and widths of S1-S2 bands in NEST, compared to calibration data from $^{14}CH_3T$ ERs and DD NRs. Additional constraints on the detector parameters g_1 and g_1^{gas} from using ^{131m}Xe, ^{83m}Kr, and higher energy backgrounds were derived with energy resolution raster scans and a Doke plot[4]. These were found to be consistent with the tuning. The tuning of bands in WS22 utilised a grid-scan of the χ^2 as only detector parameters were considered.

For WS24, the NEST microphysics parameters were needed to capture the full behaviour of the ER and NR models. The microphysics parameters for ERs and NRs, given in Table 2.2, are highly correlated. The large number of highly correlated parameters required the use of Markov Chain Monte Carlo (MCMC), from the emcee package, which performs well in high-dimension spaces [122]. The MCMC algorithm uses random-walks and rejection/acceptance based on the likelihood function, here the χ^2, at the current and proposed point. After many steps on these random walks the resulting distribution of parameters gives the probability distribution of them given the data - exactly what is required for parameter estimation. The dimensionality of the MCMC refers to the number of parameters that are fit. The only observables considered are the S1-S2 band mean and widths limited by the fact that the model needs to be re-sampled at

[4]A Doke plot assumes that the linear relationship between S2c and S1c $\frac{S2c}{E} = \frac{g_2}{W} - \frac{g_2}{g_1}\frac{S1c}{E}$, from re-arranging Equation 2.15, holds ERs. Then mono-energetic ERs at different energies are identified and used to plot this relation.

each step, which becomes prohibitively expensive with more observables. The calibrations team observed that the convergence to local minima was common for the WS24 tuning, probably because many configurations of the parameters produce a degenerate band mean and width, such that it is difficult for a walker to find the global minima.

2.3 Backgrounds

This section focuses on backgrounds that are relevant to the WIMP search analysis. Backgrounds that have uniform deposits are covered first, these are mixed xenon activation products, long-lived xenon radio-isotopes, and neutrinos. Following this, the author explores diffuse contaminants, which may be non-uniform depending on the mixing state, and highly non-uniform gamma and neutron backgrounds. Finally, the data driven accidentals background is covered. The goal of this section is to introduce fundamental concepts that are used in Chapter 5.2.9 and Chapter 6.

2.3.1 Activation of Xenon and the Electron Capture Model

Xenon can be activated when external radiation, protons, neutrons, and to a negligible degree muons, transforms a stable nucleus into short-lived radioactive isotopes with half-lives of the order of days. In WS22 this presented as a ^{37}Ar and ^{127}Xe backgrounds from activation by cosmic rays [123]. These had decayed away by WS24 but neutron calibration campaigns prior to and during data-collection introduced ^{127}Xe and ^{125}Xe, the latter having a half-life of hours and therefore had a negligible contribution in cosmic activation but could remain in trace amounts post-calibrations.

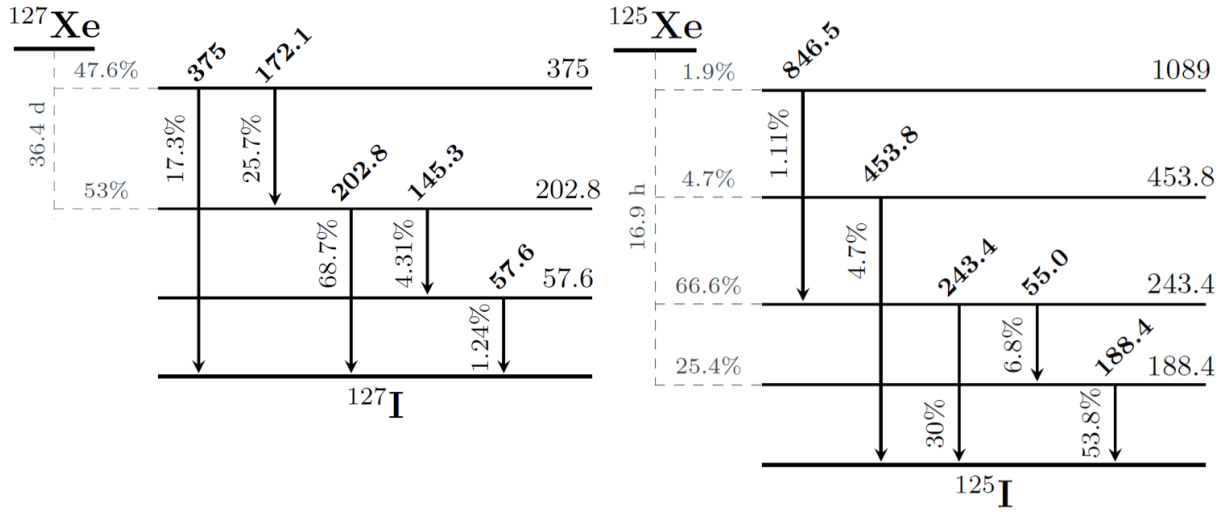


Figure 2.8: Decay schemes of the two dominant xenon activation products in LZ taken from Ref.[124]. The decay creates a SS produced by the electron capture to an excited state iodine, whose states and subsequent gamma decay are shown.

These two activation products, ^{127}Xe and ^{125}Xe, decay via electron capture (EC) to iodine in an excited state, which emits a prompt gamma upon relaxation. The decay schemes showing the prompt gamma energies are shown in Figure 2.8. The energy of the ER from the EC is determined by the shell of the captured electron, of which the L (5 keV) and M (1.1 keV) shells are relevant for WIMP searches. These are identical for both isotopes and the capture can form an Auger cascade. The prompt gamma ray provides a method to study the charge yield of Auger cascades by veto tagging or identifying a MS where the two S2s from the EC and gamma scatter can be resolved, for LZ this was studied in Ref.[124]. It was previously known that EC Auger cascades had a different structure after dedicated studies of ^{127}Xe SSs in Ref.[125]. It was also suggested that neutrino scattering off the inner shell may cause a similar response, shown in Figure 2.9.

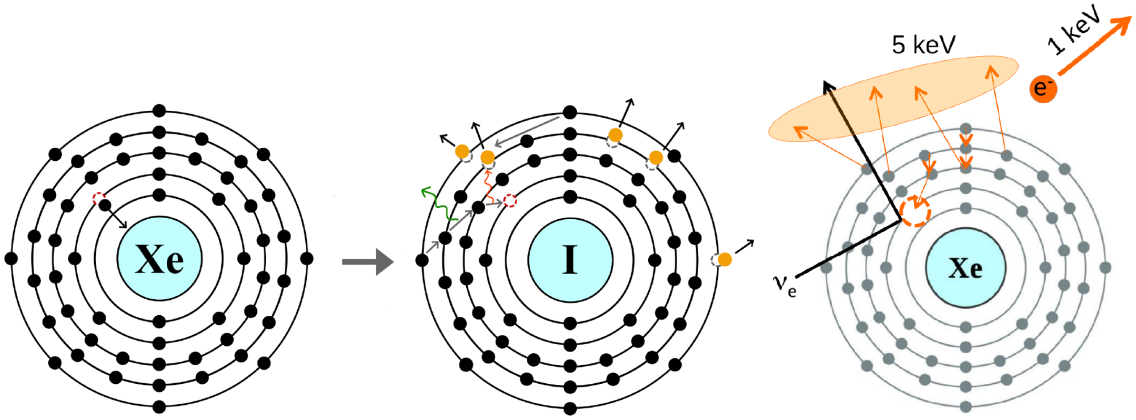


Figure 2.9: Diagram of inner shell vacancies caused by an electron capture or neutrino scatter taken from Ref. [124] and Ref. [125], respectively.

In LZ, this was modelled using the NEST ER simulation but with a custom charge yield weighed between the gamma and beta ER model. For WS22, this weighting, ω, was a function of the drift field and recoil energy tuned to match the charge yield to that was measured in Ref.[125].

$$Q_y^{weighted}(E,\mathcal{E}) = (1-\omega(E,\mathcal{E}))Q_y^{ER}(E,\mathcal{E},\vec{m}^{\beta}) + \omega(E,\mathcal{E})Q_y^{ER}(E,\mathcal{E},\vec{m}^{\gamma}). \tag{2.16}$$

The parameterisation of the WS22 model is available in FlameNEST and NEST [126][127].

2.3.2 Xenon Long Lived Isotopes

Long-lived isotopes of xenon produce rare two-neutrino double beta-decays or double electron captures (DECs). The two decays relevant for the WIMP search are ^{136}Xe double-beta and ^{124}Xe DEC, with half lives of $\mathcal{O}(10^{21})$ and $\mathcal{O}(10^{22})$ years, respectively [128][129]. In WS24 it was observed that the ^{124}Xe DEC exhibits enhanced recombination on top of the established Auger cascade from the L-shell ECs, as the two identical energy deposit are highly localised around a single site[90]. A model for its charge yield based on the TIB was developed in Ref.[124]

and its implementation for WS24 is discussed in Section 5.2.3 then reproduced in flameNEST by the author in Section 6. The ^{136}Xe decay is unlikely to present any significant enhanced recombination, as the two electrons will produce mostly separate tracks.

2.3.3 Neutrino Backgrounds

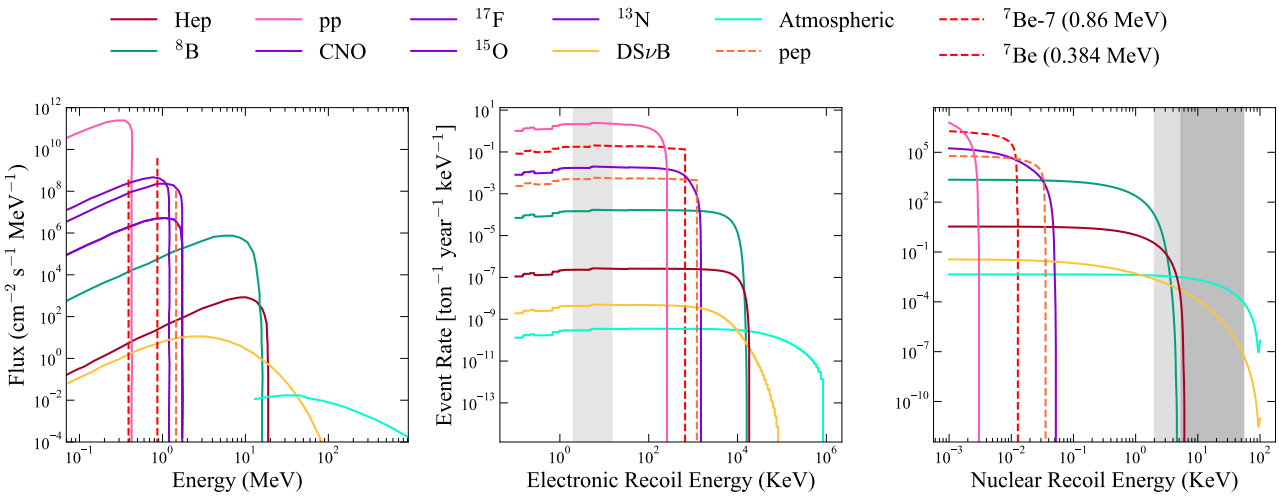


Figure 2.10: Sources of solar neutrinos backgrounds in LZ discussed in the text. The light and dark shaded regions indicate the NR energies with <50% and ≥50% in LZ's most recent science run tuned for WIMP masses greater than 10 GeV/c^2, taken from Ref.[90]. The ERs are only given an approximate indication as the spectrum is flat. Fluxes digitized from Figure 1 in Ref.[130]. Rates were produced in the LZ software DMCalc.

Neutrino backgrounds in LXe WIMP searches come from three main sources: the Sun's fusion processes, decay of the products from cosmic ray spallation in the atmosphere, and those from supernovae shown in Figure 2.10. The former two are called solar and atmospheric neutrinos. Neutrinos from supernovae consist mainly of the constant Diffuse Supernovae Neutrino Background (DSνB) from extra-Galactic supernovae. Galactic supernovae are exceptionally rare and have a clear signal, so they are not considered in the background model. Solar neutrinos are produced in a variety of fusion processes indicated by their labels, the simplest being proton-proton fusion and the most complex being the C-N-O cycle [130]. Finally, the energy spectrum of atmospheric neutrinos comes from well-understood particle physics processes, muon decay for example, but their low-energy flux relevant to WIMP searches is subject to several complex factors that are difficult to model, leading to a large uncertainty.

Scattering off the LXe target may occur via NR or ER for neutrinos; the recoil spectra for these processes are shown in Figure 2.10. Xenon's nuclear binding energy is $\mathcal{O}(1-10)$ GeV, so for relevant energies the only possible NR process is the Z^0 exchange in a Coherent Nuclear Neutrino Scatter (CEvNS). Neutrino absorption by the xenon nucleus can occur, but the rate is lower and xenon TPCs are not sensitive to them [131]. In figure 2.10, the only significant

contributions for the > 10 GeV/c^2 mass WIMP search in WS24 was atmospherics, however ^{8}B does contribute near the energy threshold and is a dominant background for low-mass WIMPs as the interaction is virtually indistinguishable from a 6 GeV/c^2 mass WIMP. For ERs, both Z^0 and $W^{-/+}$ scattering occurs and the kinematics of the scattering means that their threshold in terms of neutrino energy is lower than for NRs [130]. A neutrino may scatter off an inner shell electron and trigger an Auger cascade, like ECs, modifying its charge yield argued in [125]. As argued in Ref.[125], this has a negligible impact on WIMP sensitivity for the current generation solar neutrino ERs are modelled as normal NEST ERs in LZ.

The neutrino interactions discussed here are an irreducible background for future LXe dark matter detectors. Often referred to as the neutrino floor or fog, future WIMP searches would require observing an excess in the reconstructed energy spectra of a set of neutrino CEνNS from atmospheric and ^{8}B solar neutrinos, as shown in Figure 1.13 from Section 1.5. The next generation of LXe detectors are designed to penetrate this floor, and whilst doing so neutrinos will be the dominant ER and NR backgrounds due improvements in reducing radon contamination. Despite being backgrounds for the WIMP search, they still yield interesting potential to probe new physics. For example, the ^{8}B flux will constrain the low-energy Weinberg angle and the pp and CNO flux help understand the solar neutrino luminosity and therefore improve models of the Sun's composition and dynamics [131].

2.3.4 Diffuse Contaminants

This section discusses the origin of the two dominant diffuse contaminants, radon and krypton, in LZ and how they contribute to the WIMP search. The beta decays of ^{214}Pb, ^{85}Kr, and ^{212}Pb are the most dominant in LZ, in that order. These can be mitigated through a variety of techniques. For example, XENONnT utilises distillation columns to remove radon continuously and a separate intermittent system for krypton distillation [72]. In LZ, the flow is manipulated to reduce mixing between contaminated components and the central volume, as discussed in Section 2.1.2 [123]. This low-turbulence state allowed the utilisation of radon tagging in WS24, which will be discussed in Section 5.2.9 and was developed on WS22 data in Ref.[102].

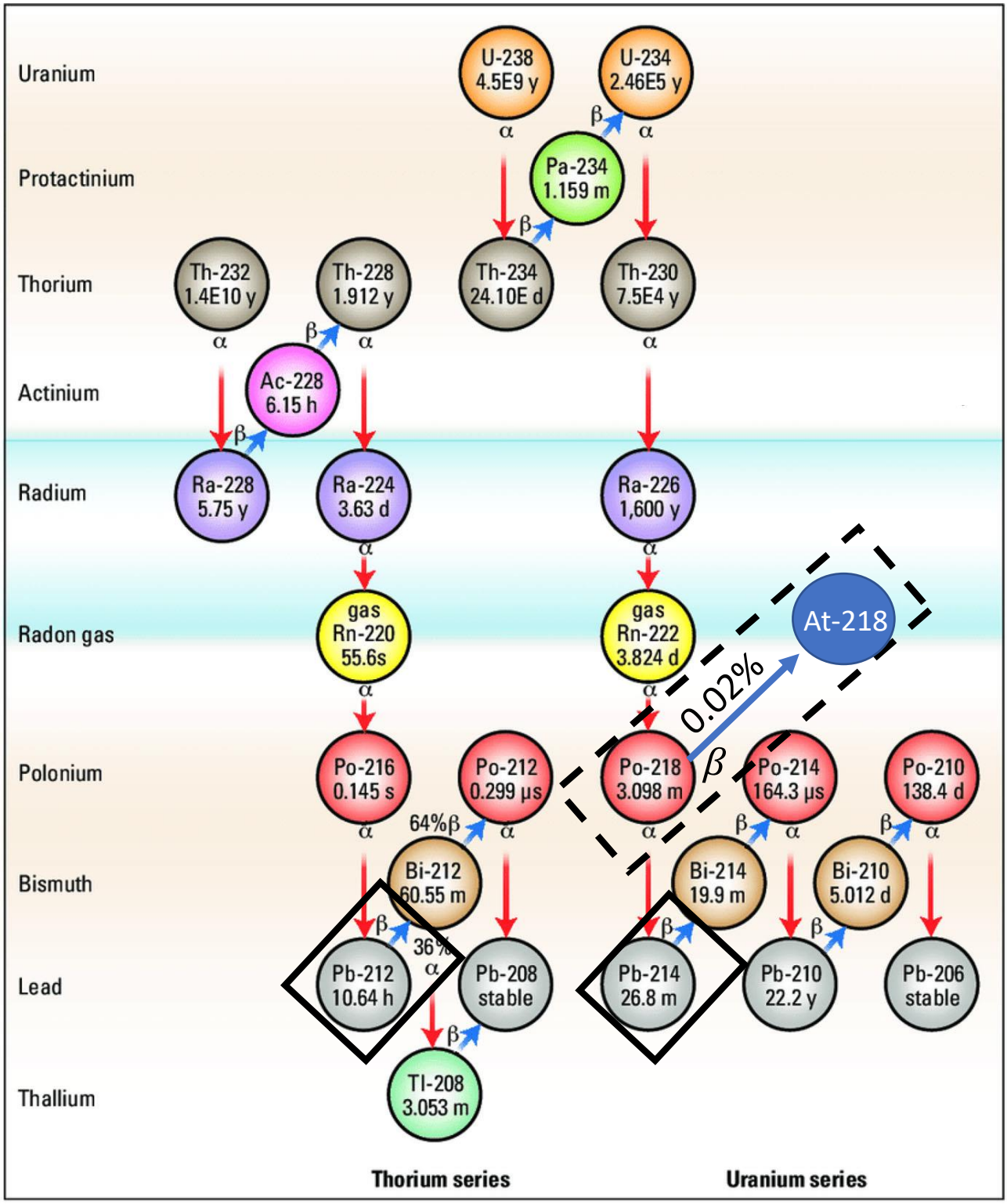


Figure 2.11: Modified figure of uranium and thorium series extracted from Ref.[132]. The decays relevant to a WIMP search are indicated in black boxes and the rare ^{218}Po decay has been added [133].

Radon enters the detector through the decay of isotopes in the uranium and thorium series that may become attached to the surfaces of components during manufacture and construction, or exist in the bulk of materials. In their decay, they produce radon, which can outgas from

the materials into the LXe target. Of the daughters of radon, only three actively contribute as backgrounds to a WIMP search due to their energies and topologies. In a beta-decay, the remaining daughter must be in a stable state to prevent a prompt gamma from triggering a veto or a multiple scatter, known as a naked decay. The end-point of this decay should also be $\leq \mathcal{O}(1)$ MeV so a significant fraction occurs in the ROI of $\mathcal{O}$(1-30) keV. For the thorium and radium series in Figure 2.11, the candidate isotopes are ^{214}Pb, ^{212}Pb, and a rare decay of ^{218}Po to ^{218}At, instead of ^{214}Pb, which was considered for the first time in WS24. The isotopes ^{214}Pb and ^{212}Pb beta-decay to the Bi ground state with branching ratios of 11% and 11.6%, respectively. This means that there is also a significant portion of non-naked decays occurring as well [134]. The ^{218}Po always decays into its ground state [133]. Note that the progenitors of ^{212}Pb have a shorter half-life than their ^{214}Pb counter parts, but that ^{212}Pb itself has a longer half-life than ^{214}Pb.

Natural krypton is found within commercial xenon and contains the ^{85}Kr which beta-decays to the daughter ground-state with a branching ratio >99% and $\mathcal{O}(10)$ year half-life [134]. For LZ, the xenon was purified using gas-phase charcoal chromatography, similar to how the iRRS functions, to reduce levels to $\leq$ 200 ppq [96]. A residual gas analyser sampling system measured this to be within the design specification [135]. It does evolve over time, due to outgassing from materials in contact with the liquid xenon and potential ingress, but the krypton concentration was observed to have plateaued in WS24. Another beta-decay isotope in natural xenon is ^{39}Ar, this was determined to be far less abundant, with abundance $<$ 1 ppq, and has a longer half-life of $\mathcal{O}(100)$ years, but is added to the rates for the background model [123].

2.3.5 Material Gamma and Neutron Backgrounds

Radio-isotopes within the detector materials can produce neutrons and gammas which contribute to the background of a WIMP search. The neutron background consists of (α,n) reactions from the uranium and thorium series, as well as Uranium Spontaneous Fission (USF). The detector ER component consists of prompt gamma rays from the uranium and thorium series, as well as naturally occurring ^{40}K and activated ^{60}Co[5]. These gamma rays typically have energies $\mathcal{O}$(few) MeV and therefore contribute to the WIMP search Region Of Interest (ROI) through Compton scattering. The beta-decays are modelled in simulations, but any background would be highly localised to the walls and not likely to contribute to the background model within the fiducial volume. The high-energy peaks may be used as a calibration to obtain the activity of each isotope in sets of components, which was obtained for WS22 in Ref.[123], further details in Aiham Al Musalhi's thesis in Ref.[136]. The prominent events being a low-energy tail, the number of individual components required, and the efficacy of shielding and vetoes makes both of these sources difficult to produce sufficient simulations. This is especially true for detector

[5]This may be activated through cosmogenic processes or neutron calibrations.

ERs, as gamma rays cannot penetrate deep into LXe, so very few simulated trajectories survive to contribute to the background model. Solutions for this are discussed later in Section 6.

2.3.6 Accidental Backgrounds

Accidental Coincidence (AC) events may be caused by many different sources; for WS22 and WS24 it was assumed that the sources of isolated S1 and S2 light are not correlated, and therefore AC events are just random pairings of these isolated S1 and S2. To emulate a random pairing the following procedure was used: identify an S2 triggered event with no S1 pulse and a random triggered event with a suitable S1 in the pre-trigger window, then replace the pre-trigger window of the S2 event with the pre-trigger window of the random event. This process is called Accidental Chop-Stich (ACS) and produces a replica of a randomly paired S1 and S2. This model is then compared to events with an Unphysical Drift Time (UDT), that is a drift time greater than the maximum. These UDT events must be ACs as they could not be produced by signals from the same physical event. The isolated S1 and S2s of AC events present unique topologies over physical events in the target. For example, physical S1 and S2 pulse shapes have defined relationships with drift time due to electron diffusion and S1 light transport. These differences are exploited in S1 and S2 based cuts, applicable to ACS and UDT data before building the model. Often, there are insufficient events to create a smooth PDF so Kernel Density Estimation (KDE) is used, details are given in Section 7.2. The rate of the model can be determined from the rate of the isolated S1 occurring within a physical drift time window of an isolated S2, adjusted by the efficiencies for the S1 and S2 based cuts derived from ACS.

2.4 Further Analysis Considerations

In addition to modelling physical events, analysers must: use analysis cuts to isolate signal from background, data quality cuts to remove noise or poorly behaved data, and employ procedures to mitigate bias. These are discussed in this section.

2.4.1 Analysis Cuts and Data Quality

The S1 and S2 based AC cuts are part of the data-quality cuts and may also reject physical events, which must be accounted for in modelling; this is done via S1 and S2 based acceptance curves. Data quality cuts can also remove non-AC events that are poorly or misclassified, such as MS resolved as an SS by LZap. To derive these acceptance curves in S1-based cuts CH_3T calibration data is used. For S2, as acceptance typically increases with uncorrected S2 within the ROI, the impact on the NR model must be emphasised. Therefore, fake NR data are produced again by chopstitch because NR calibration data is highly localised in the detector due to CSD and DD conduit positions. The process is unique from ACS, both CH_3T and DD or

AmLi data is used to form an S1 and S2 pair that match NEST prediction and the physical characteristics of a uniform NR source. The functional form of these efficiencies may be found in Figure 2.12.

Cutting only on the contents of the data is not sufficient, the quality of the environment in which the data was collected matters too. For example, there are many periods, particularly after a power-outage, where a temporary high rate of light emission localised to grids may occur in the detector, known as hot-spots. Including these periods in the science data would increase the chance of AC events or other aberrations. Another example is after a large S2, subsequent photo-ionisation and scintillation cause a series of isolated progeny S2 and S1s, colloquially referred to as electron and photon trains. Live-time cuts remove these and do not have an acceptance but must be accounted for in calculating rates and counts.

Analysers apply cuts based on RQs to remove regions of the observable space where background and signal are difficult to model. A Fiducial Volume (FV) cut removes events near the wall where some if not all of the S2 is lost. This is also required to remove backgrounds that increase near the edge, like neutrons and gammas, as the position of an event is not yet a viable observable with current analysis tools. To isolate the WIMP search from noise a 3-PMT coincidence requirement is imposed. Alongside this, cuts in S1c, S2c, and S2 define the ROI, for WS22 this was $3 \leq \text{cS1} \leq 80$ phd and $\text{S2} > 10 \times SE_{size} = 600$ phd. The impact of the data quality cuts and ROI on the fraction of NR events accepted within the FV is shown in Figure 2.12. Throughout this thesis the author refers to "threshold" or "near-threshold" sources and signals, this refers to those in the lower energy tail of the efficiency curve shown. For example, ^{8}B CEνNS falls into this category.

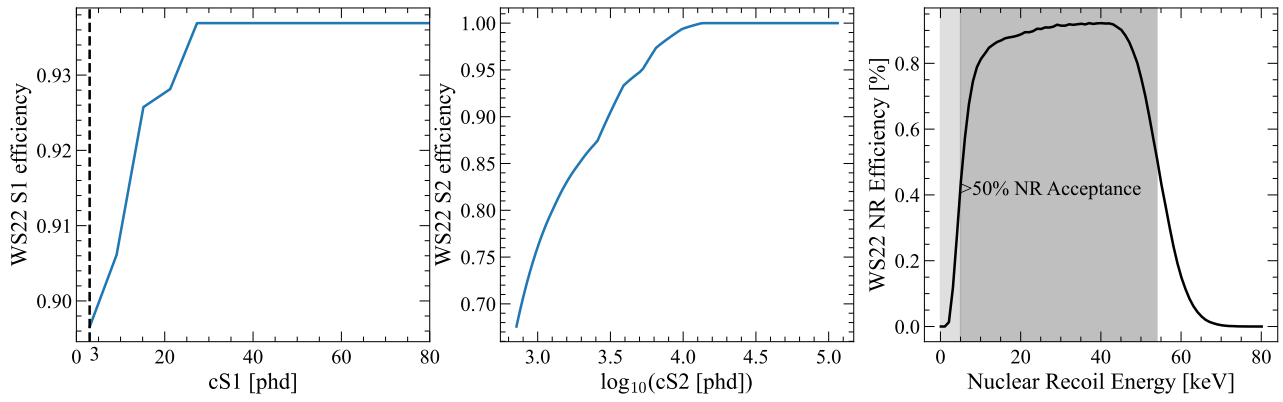


Figure 2.12: Acceptance curves for WS22 plotted in cS1 (left) and $\log_10$(cS2[phd]) (middle), these encapsulate the data quality and trigger efficiencies. The total efficiency defined as the fraction of NR events surviving all cuts within the FV (right). These were obtained from the FlameNEST code from Ref[111], the uncertainties can be found in the WS22 paper [95].

For calculating the rate of events, it is important to know the fiducial mass, that is, the

mass of the target inside the fiducial volume. In `BACCARAT` and `LZLAMA` the total active volume of LZ is assumed, and subsequent FV cuts are applied to the simulations to produce the correct rate. However, it is essential to confirm this volume and mass with a comparison to the data, typically using diffuse calibration sources.

2.4.2 Bias Mitigation and Salting

Analysers can use sidebands to investigate a model, such as UDTs for accidentals or higher energy regions to constrain beta-decays. In order to inspect the model in science data within the region of interest, without introducing analyser bias, LZ used a salting procedure, which is the introduction of fake signal in a random and unknown way. This, in contrast to the XENONnT collaboration's approach of blinding in Ref.[137], allows analysers to inspect potential issues with the background model before performing statistical inference. The procedure is identical to NR chopstitch for determining acceptances but injected directly into the data stream such that analysers cannot identify them from real data in any way. A chosen salt-keeper is the only member of LZ able to obtain the identity of the events.

Chapter 3

Statistical Inference

The goal of statistical inference is to either test a hypothesis or perform parameter estimation. Both of these require some way of defining probability which has two main camps; does it mean the frequency, given a hypothesis, of an event or does it represent a degree to which we believe some possible hypotheses. In the latter the probability of an observation is the conditional probability given the probability of the hypothesis and the probability of the observation under that hypothesis. These are frequentist and Bayesian statistics, respectively. As the dark matter community has chosen to use frequentist statistics, we will limit our discussions to that [138]. Unlike Bayesian, in frequentist statistics the probability of a hypothesis does not exist - only the comparison of different hypotheses given an observation. Whilst this can be conceptually frustrating, it is more robust against subjective choice of the initial probability of the hypothesis, called a 'prior', and in some cases against model misspecification.

3.1 Frequentist Hypothesis Tests and Confidence Intervals

The probability of any singular realisation of data consisting of continuous variables is not defined. However, the value of the probability density function (PDF) at a point, or a collection of points in an observation, could be thought of as a measure of likelihood. One could take a hypothesis and calculate the frequency at which the given observation or at least as unlikely occurs, and call this a frequentist probability. However, it is argued in Ref.[1] this has little meaning, as no matter how unlikely an outcome may be under the assumed hypothesis, if it is even more unlikely under any possible alternative hypothesis, there is no reason to reject it. Therefore, statistical tests for a hypothesis involve a null hypothesis H_0 and an alternative hypothesis H_1 with some measure that compares their likelihoods t. The frequency of an observation of t_{H_0} or more likely under the alternate hypothesis, under the assumption that the null hypothesis is true, is called the p-value. The p-value is then compared to a predefined critical value which determines whether or not a hypothesis is rejected, accepting the alternative.

For a hypothesis H and critical value, α, the p-value and criteria would be;

$$p_{H_0} = \begin{cases} \leq \alpha & \text{reject } \mathrm{H}_0 \text{ \& accept } \mathrm{H}_1, \\ > \alpha & \text{accept } \mathrm{H}_0 \text{ \& reject } \mathrm{H}_1 \ . \end{cases}$$

The critical value for a hypothesis test should be chosen before any outcome is measured. In particle physics, the p-value is often reported as a significance, according to one side of the normal distribution, and compared to community standards. For example, a p-value of 5(0.3)% yields 2(3)σ and the community standards are generally 3σ for claiming evidence and 5σ for discovery. These criteria are to avoid spurious claims and can be adjusted under certain conditions, for instance, one may wish for a greater significance where systematic uncertainties dominate or accept a lower one where there is an expected signal or a low 'degree of surprise' [139].

Experiments typically measure many events which can have many observables so are represented as a set of vectors, $\{\vec{s}_{obs}\}$, called a data set. The p-value is calculated for a given data set as the frequency of the observed test statistic evaluated on the data, $t(H_0|\{\vec{s}_{obs}\})$, or more in agreement with the alternate occurring assuming that the null hypothesis is true. As t is typically chosen to be zero for total agreement and to increase as the outcome becomes less likely under the null hypothesis, the p-value is written as

$$p_{H_0} = \int_{t(H_0|\{\vec{s}_{obs}\})}^{\infty} f(t(H_0)|H) dt_H, \tag{3.1}$$

where the function, $f(t(H_0)|H)$, drops the data set because it represents the distribution of possible $t(H_0)$ occurring under H. This test statistic has been left ambiguous, and the best choice will be explored in the next section.

Parameter estimation in frequentist statistics uses the confidence interval construction, which is the inversion of a hypothesis test. To understand this, we should parameterise the null and alternative hypothesis. For example, let's say we are searching for some new signal with rate μ. To test for the presence of this new signal the null and alternate hypotheses would be H_0:$\mu = 0$ and H_1:$\mu > 0$. After carrying out our test we may determine whether or not the new signal is present and either estimate its rate μ or determine an upper-limit. To do this we select a value of the rate μ' and choose that to be the null hypothesis H_0 :$\mu = \mu'$ we then vary until we reject it under an α. Once we obtain a region where the p-value is above the α we call this a confidence interval of confidence level(CL) 1-α. That is to say that, under repeated experiments and construction of the confidence interval, the parameter should be within it CL% of the time as we would reject every parameter outside it. The same stands for an upper limit, for which the lower bound would be zero.

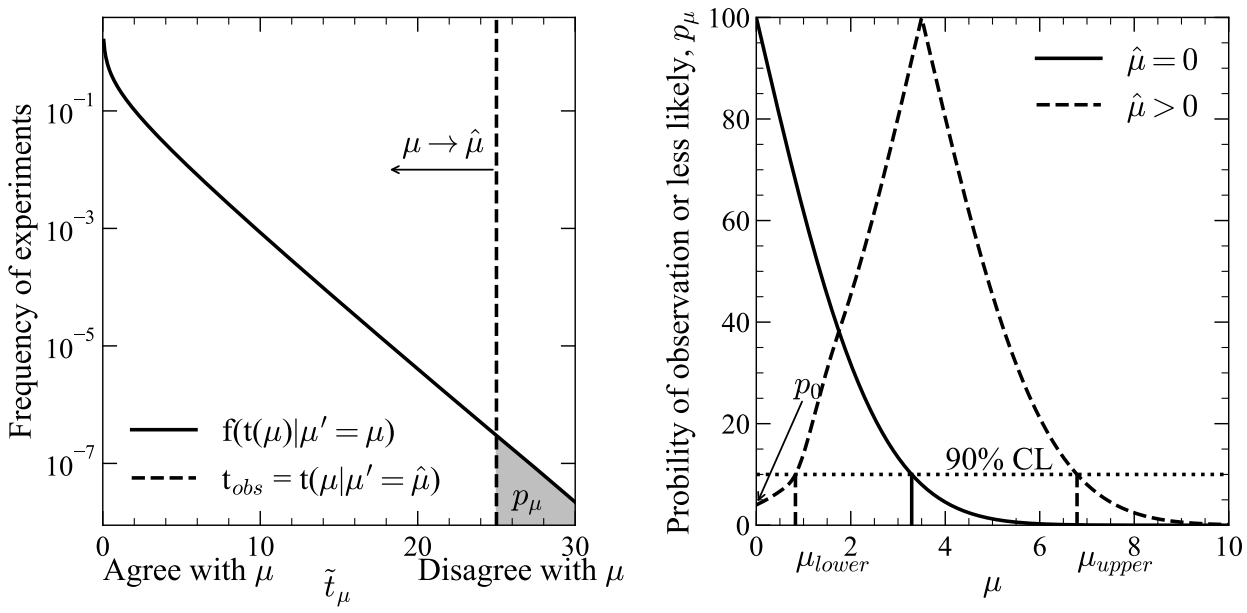


Figure 3.1: Plot of test statistic distribution using asymptotic formula in Ref.[140] and section 3.4.2 (left). Integral of the asymptotic distribution for different test values, μ, and two different cases where there is and is not a significant signal in data, indicated by the value of the best fit to data, $\hat{\mu}$.

In figure 3.1 an illustration of two tests with a scalar signal parameter μ, one with a non-zero signal and one without. The left figure indicates the distribution of experiments we expect under a given null hypothesis $H_0 : \mu' = \mu$ and the value of the observed test statistic. As the tested value approaches the observed the test statistic approaches zero. On the right are p-values from several tests at different $H_0 : \mu' = \mu$ plotted against μ. We see that as a non-zero signal is added the p-value curve at the left side falls below the CL to form an upper and lower limit, called lift off, and the intercept with zero gives the discovery p-value.

3.2 The Profile Likelihood Ratio and Power

As mentioned, the likelihood is the PDF of the observables under a given hypothesis evaluated on the observed data set. The PDF may have many parameters that are not observables, but it is not a PDF of the parameters only of the observables. When a data set is observed, the observables are fixed, and the PDF becomes a likelihood. As discussed, a set of parameters can differentiate H_0 and H_1, these are called the parameters of interest, $\vec{\mu}$. For example, signal counts can be a parameter of interest. These are not the only parameters of the model as H_0 and H_1 may be described by a set of nuisance parameters, $\vec{\theta}$, that are not of interest to the inference. There may be parameters unique to H_1 that are still treated as nuisance parameters. As our data set is a collection of independent events which follow the same distribution P,

$$L(\vec{\mu}, \vec{\theta}|\{\vec{s}_{obs}\}) = \prod_{\vec{s}_i \in \{\vec{s}_{obs}\}} P(\vec{s}_i|\vec{\mu}, \vec{\theta}). \tag{3.2}$$

The values which maximise this likelihood are denoted as $\hat{\vec{\mu}}$ and $\hat{\hat{\vec{\theta}}}$, this is often referred to as the unconditional best fit or unconditional Maximum Likelihood Estimators (MLE). Using the construction that follows, it can be shown in the large-sample limit these are best-estimators. Since the test statistic must be a function of $\vec{\mu}$, we can remove dependence on the nuisance parameters by maximising the likelihood for fixed values of $\vec{\mu}$. The subsequent best fit nuisance parameters are a function of $\vec{\mu}$ and labelled $\hat{\hat{\vec{\theta}}}(\vec{\mu})$. This process of removing the dependence on nuisance parameters is called profiling. Alternatively, one could use a prior to marginalise over the nuisance parameters, discussed in Ref.[141], but this is CPU intensive and the chosen standards in the dark matter community is the profiled likelihood [138].

The criteria for choosing a test statistic are based on two possible sources of error; type 1 of rejecting H_0 when it is true, and type 2 of accepting H_0 when it is false. These can both be classified by,

$$M_{\vec{\mu}'}(\vec{\mu}) = P(p_{\vec{\mu}} < \alpha | \vec{\mu}'), \tag{3.3}$$

where H is parametrised by $\vec{\mu}$. So, the probability of a type 1 error would be measured by $M_{\vec{0}}(\vec{0})$ and is theoretically the value α by definition. However, complications such as systematics not modelled in the likelihood can increase this causing spurious exclusion, hence the different standards on the critical value. The probability of a type 2 would be $1 - M_{\vec{\mu}'}(\vec{0})$, the second term is typically what is referred to as the power of the test. Neyman and Pearson found that the test-statistic that maximises power is the likelihood ratio,

$$\frac{L(H_0)}{L(H_1)}, \tag{3.4}$$

where parameters and data sets have been dropped. They also found that where there are many admissible alternate hypotheses, such as values of $\vec{\mu}$, the denominator should be replaced by the maximum likelihood case.Intuitively, this test statistic does maximise power as it compares a given H_0 to the most likely alternative given the data and therefore maximising the chance of rejecting H_0 when it is false [1]. So, in terms of equation 3.2 it becomes

$$\lambda(\vec{\mu}|\{\vec{s}\}) = \frac{L(\vec{\mu}, \hat{\hat{\vec{\theta}}}(\vec{\mu})|\{\vec{s}\})}{L(\hat{\vec{\mu}}, \hat{\hat{\vec{\theta}}}|\{\vec{s}\})}, \tag{3.5}$$

known as the profile likelihood ratio (PLR). It is bounded between 0 and 1, as the numerator cannot exceed the unconditional maximum likelihood. Any monotonic[1] function of it also maximises the power. The test statistic utilised throughout this thesis follows the recommended conventions by the dark matter community, as in ref [138],

$$t(\mu) = \begin{cases} -2ln\lambda(\mu) & \hat{\mu} \geq 0, \\ -2ln(\frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu)|\{\vec{s}\})}{L(0, \hat{\hat{\vec{\theta}}}(0)|\{\vec{s}\})}) & \hat{\mu} < 0, \end{cases} \tag{3.6}$$

[1]A function that reverses or preserves the order of its domain; y=$ln(x)$,$-x$, etc. Anything with an always negative or always positive gradient.

where we have dropped explicit dependence on the data for brevity. This was chosen such that rejecting the background-only hypothesis, requiring rejecting $\mu = 0$, to be evaluated using the same test-statistic as limit-setting, rejecting $\mu > 0$.

3.3 Combined and Extended Likelihood Function for Counting Experiments

For a counting experiment in particle physics the likelihood differs from equation 3.2 by introducing the Poisson term for the probability of observing the total n events,

$$L(\vec{\mu}, \vec{\theta}|\{\vec{s}_{obs}\}) = \frac{\nu^n e^{-\nu}}{n!} \prod_{\vec{s}_i \in \{\vec{s}_{obs}\}} P(\vec{s}_i|\vec{\mu}, \vec{\theta}),$$

where we use ν to separate expected counts from the other parameters. In reality, the model is constructed from multiple independent sources of events, each with a different rate, ν_j so the correct PDF to evaluate the events on becomes,

$$L(\vec{\mu}, \vec{\theta}|\{\vec{s}_{obs}\}) = \frac{\nu^n e^{-\nu}}{n!} \prod_{\vec{s}_i \in \{\vec{s}_{obs}\}} \sum_{j}^{sources} \frac{\nu_j f_j(\vec{s}_i|\vec{\theta})}{\nu}, \tag{3.7}$$

where f_j is the individual PDF of each source and $\nu = \sum_j \nu_j$. If there were multiple experiments, or ancillary datasets in one experiment called a sideband, with shared parameters, the total likelihood would be the product of both,

$$L(\vec{\mu}, \vec{\theta}) = \prod_d L(\vec{\mu}, \vec{\theta}|\{\vec{s}_{obs}\}_d),$$

where d labels separate data sets. Different variables can be used between the likelihoods, as long as the correlations between them are captured in an additional term.
In frequentist statistics, in order to constrain all the nuisance parameters of our model we should use ancillary measurements and a combined likelihood. However, these measurements are often independent from the main data set of the experiment, or region of interest, and can be treated as an individual constraint term,

$$L(\vec{\mu}, \vec{\theta}|\{\vec{s}_{obs}\}) = \frac{\nu e^{-\nu}}{n!} C(\vec{\theta}) \prod_{\vec{s}_i \in \{\vec{s}_{obs}\}} \frac{\nu_j f_j(\vec{s}_i|\vec{\mu}, \vec{\theta})}{\nu}. \tag{3.8}$$

This story is more complex when one considers parameters that have theoretical uncertainties. For proper frequentist statistics, the confidence interval should be presented as a function of the parameter of interest and these uncertain parameters, allowing future interpretation as information about the parameter improves. In reality, we add a constraint term for these parameters in the likelihood which is called a 'hybrid' approach as it is inherently Bayesian in nature - containing a prior which we update [142]. This approach makes results simpler, and consume less computational resources, but can make it more difficult to interpret under updated measurements of the uncertain parameters.

3.4 Practical Tests and Confidence Intervals

Now that we have a test-statistic, to practically implement it there are two approaches; asymptotic formulae and non-asymptotic methods. Here we present the conditions and methodology for both, as well as clarifying some terminology used in the rest of this thesis.

3.4.1 Sensitivities, Brazil Bands, and Power Constraints

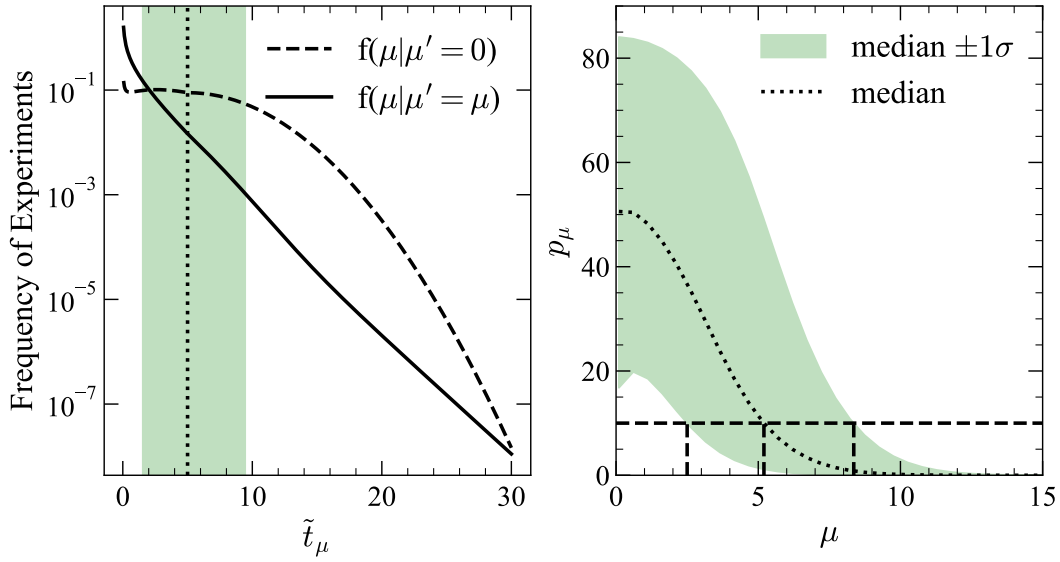


Figure 3.2: An example of sensitivity using asymptotic formula in Ref.[140]. The distribution of observed test statistics under the null hypothesis is shown (left) and that distribution translated into p-values (right). As lift-off in the p-value curve caused by signal-like background can artificially lower the limit, it is important to use the distribution of upper limits.

Apart from knowing your observed limits or discovery significance, sensitivity is useful for predicting the performance of an experiment. Sensitivity has a few definitions in a counting experiment, in this work we only need to know three of them:

- **Discovery Sensitivity** This is, given our predicted or assumed values for $\vec{\theta}_0$, the frequency of rejecting the background hypothesis by determining the distribution of p_0 under the assumption of a given signal, $f(p_0|\mu = \mu', \vec{\theta}_0)$. Where μ' is determined from predicted signal counts, or a range is chosen to encapsulate the possible signal counts, given previous upper limits.

- **Sensitivity** This is, given our predicted or assumed values for $\vec{\theta}_0$, the distribution of upper limits by evaluating the distribution of μ_{upper} under the background-only assumption, $f(\mu_{upper}|\mu = 0, \vec{\theta}_0)$.

- **Expected Upper Limits** This is identical to sensitivity but derived from the conditional best fit with zero signal to data $f(\mu_{upper}|\mu = 0, \hat{\vec{\theta}}(0))$. These are often called 'Brazil Bands' and are often presented with observed limits.

There is a fourth quantity we could evaluate, called the discovery power $M_\mu(0)$ in equation 3.3. Unfortunately, as we have not yet made discovery we will not cover that here. However, upper limits are the inversion of discovery so have an associated power;

$$M_0(\mu) = P(p_\mu < 1 - CL|0) = P(\mu_{upper} < \mu|0)$$

This states that if the observed data-sets were distributed under the alternative-hypothesis for the test, background only[2], what is the probability that we would reject this value of μ. Or what is the probability of accepting this value of μ given that it is false, a type 2 error. If we were unable to reject it in most cases, i.e the power $M_0(\mu)$ is less than some M_{min} , then we would not set an upper limit based on that. Therefore, the upper limit is defined as

$$\mu*_{upper} = max(\mu_{upper}, \mu_{min})$$

where $M_0(\mu_{min}) = M_{min}$, ensuring that we only set limits on values we have sufficient power to exclude. To obtain the power, one just needs to use the quantiles of expected upper limits. In the dark matter community the chosen power was $M_{min} = 0.16$, which is typically interpreted as not allowing our limits to be better than if the background fluctuated to -1σ - i.e we observed a cleaner than anticipated signal region.

3.4.2 Asymptotic Formulae

Asymptotic formulae for the test statistic in equation 3.6 were derived by Cowan et al. in ref [140] using the formalism of Wilks and Wald. The conditions for their accuracy are given in Ref [143] which we'll summarise here;

- **Asymptotic** there is a sufficiently large sample size, N, that the parameter of interest is normally distributed with variance $\mathcal{O}(1/\sqrt{N})$.
- **Interior** True values of the parameters of the likelihood lie within their admitted bounds.
- **Nested** The null hypothesis is a limiting case of the alternate.
- **Identifiable** Different sets of likelihood parameters represent distinct models.
- **Correct** The model accurately described the background and signal.

[2]Consistent with the above construction and maximum power, the actual alternative is the best fit assuming zero signal.

Our likelihood being interior and identifiable are modelling conditions, nested is a typical case for particle physics searches with no signal being a limiting case of the signal hypothesis. For determining if the model is correct we must investigate the data. To do so without biasinpaigns g ourselves anti-bias techniques are used to allow us to investigate the data using goodness-of-fit tests discussed in Section 5.1 . Finally, Asymptotic requires that there are sufficient signal and/or background events, which is not the case for rare event searches.

The asymptotic formula for the cumulative distribution of equation 3.6 is given by,

$$F(t(\mu),\mu) = \begin{cases} 2\Phi(\sqrt{t(\mu)}) - 1 & ,t(\mu) \leq (\frac{\mu}{\sigma})^2 \\ \Phi(\sqrt{t(\mu)}) + \Phi(\frac{t(\mu)-(\frac{\mu}{\sigma})^2}{2\frac{\mu}{\sigma}}) & ,t(\mu) > (\frac{\mu}{\sigma})^2 \end{cases}$$

where σ is the uncertainty of $\hat{\mu}$ and in the asymptotic limit can be approximated from the hessian of the likelihood,

$$V_{ij} = -(\frac{\partial^2 L}{\partial\nu_i \partial\nu_j})|_{\hat{\vec{\nu}}},$$

where $\vec{\nu} = (\mu, \vec{\theta})$ and therefore $\sigma = \sqrt{V_{00}}$. The special case for $\mu = 0$ yields a half-chi-squared distribution of $f(t(0)|0)$ which has a normal cumulative distribution ,

$$f(t(0)|0) = \frac{1}{\sqrt{2\pi t(0)}} e^{-t(0)} \tag{3.9}$$

$$F(t(0)|0) = \Phi(\sqrt{t(0)}). \tag{3.10}$$

This gives the definition of asymptotic discovery significance to be $\sigma_{disco} = \sqrt{t(0)}$.

3.4.3 Non-Asymptotic Monte Carlo Methods

The Monte-Carlo method for approximating the test statistic distribution only requires that the model is correct. Usually, interior is preferred as numerical methods to obtain the MLEs can perform poorly when the true value of the parameter lies on the boundary of its admitted space. The procedure to obtain $F(t(\mu)|\mu)$ is as follows;

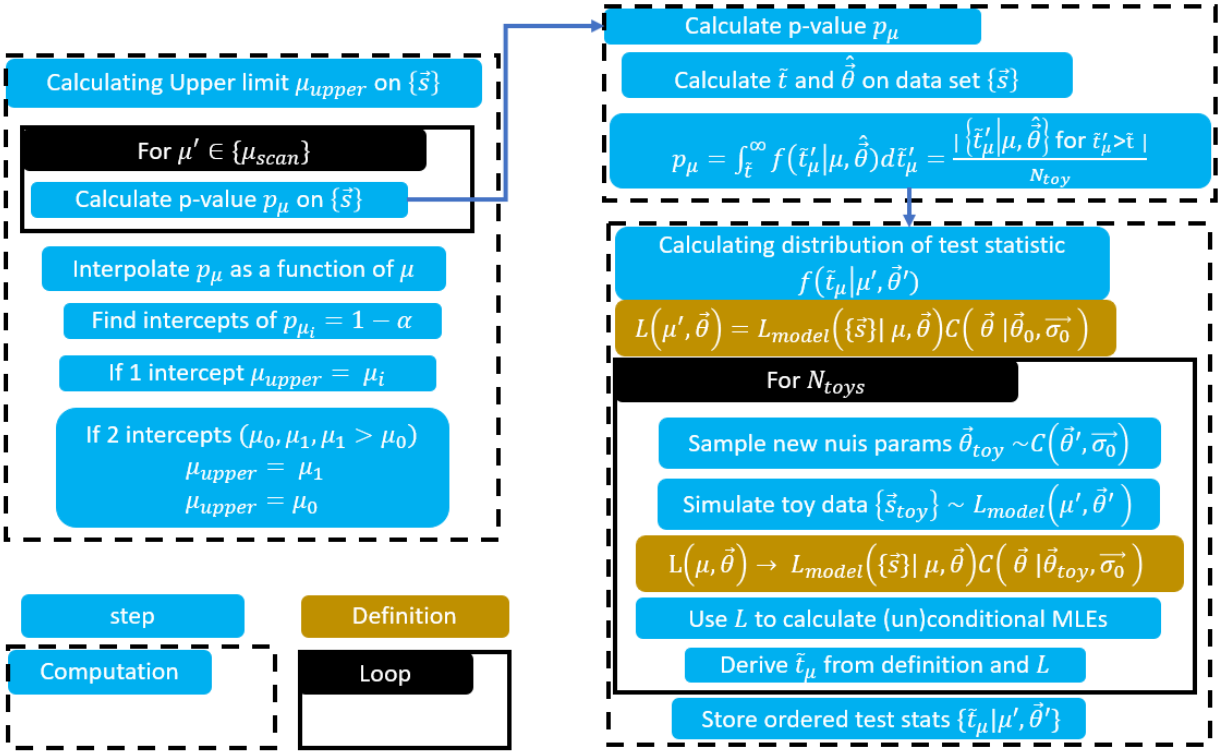


Figure 3.3: The non-asymptotic calculation of a confidence interval, which requires the p-value and therefore the test statistic distribution. Noted are some redefinitions specific to the implementation in the dark matter and high energy physics community, particularly in the treatment of nuisance parameters in throwing toys. Outside the calculation of the confidence interval the only p-value of concern is p_0 for discovery.

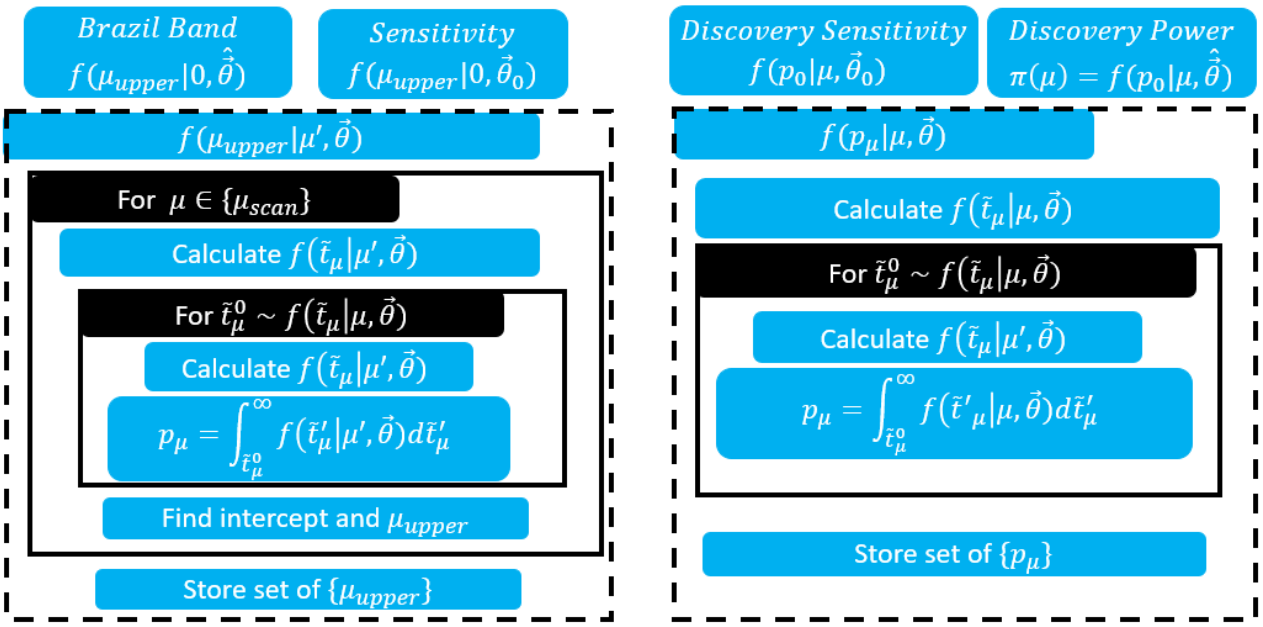


Figure 3.4: Non-asymptotically calculating the sensitivities discussed in Section 3.4.1 using the test-statistic distributions from Figure 3.3. The confidence interval is explicitly separated, rather than calculating the distribution of predicted test statistics and evaluating their p-value, the p-value curve for each predicted data set is calculated and then the intercepts are found. This is important in upper limits as a background-only data-set as a lower limit may be caused by background faking the signal, artificially lowering the predicted upper limit. In those cases, the larger intercept is used as an upper limit.

The non-asymptotic inference methods for various values of interest are shown in Figure 3.3. The maximum likelihood fits required to calculate each test-statistic and conditional best fit are implicit when those values are shown. To sample the likelihood, where possible, sample the parameters from their constraint terms then the observables from the rest of the likelihood. Uncertainties are not updated to the best fit, as it must resemble a repeat of the experiment and whilst the value may have shifted our ancillary measurements would be no more precise than before. It is argued in Ref. [142] that the toys for the ancillary measurements should also be thrown, but in LZ they are well-modelled by their chosen constraint terms.

Chapter 4

Explicit Likelihood For Physics Models

Particle physicists must evaluate the probability of events given a model. This is often done using Monte Carlo (MC) simulations, which are used to populate a normalised histogram with simulated events to make a PDF, referred to as a template. As more parameters or observables are added to the model, the number of simulated events required to obtain an accurate estimate of the probability scales exponentially, known as the curse of dimensionality. In LZ this is carried out using a C++ and ROOT based software known as `LZLAMA` interfaced with the C++ based NEST, and whose outputs are processed through an `ALPACA` modules - this can be opaque and difficult to modify. `flamedisx` is a Python based object-oriented package that can produce the same final models as NEST, `LZLAMA`, and `ALPACA` tools. Its structure allows for easy modification of background and signal models for a specific application using inheritance. Models constructed in it can carry out simulation using `scipy`, matching the performance of the above tools [144]. However, the main advantage of `flamedisx` comes from the explicit inference methods, which will be referred to here as the `flamedisx` method. In essence, this uses the observed data to inform model evaluations and will be explained by example in Section 4.2.1. This method is written in `TensorFlow`, allowing for the use of Graphical Processing Units (GPUs) and auto-differentiation. The former are optimised for repeated calculations of the same structure with varying inputs, which is ideal for optimisation problems. The latter allows optimisation algorithms to use explicit gradients rather than estimating them with steps in the parameter domain, which reduces the error and time required to find the optimal point.

`flamedisx` was developed by the authors of Ref.[145] from the XENON collaboration. Later NEST and parts of `LZLAMA` were integrated by Dr Robert James and Dr Jordan Palmer from the LZ collaboration in Ref.[127], this was called FlameNEST. This was applied as a proof of principle by Dr Robert James to several analyses, including WS22, in Ref.[111]. The author contributed to the WS22 FlameNEST analysis by investigating and deriving the position distributions of relevant sources and validating models against standard LZ simulations. At the time the analysis was unable to fully utilise `flamedisx`, as will be discussed, the structure of the NEST models required extremely large memory usage. Instead, a limited implementation which could only fit rates, and not other model parameters, was used. This chapter describes

the `flamedisx` methodology, what caused the performance issues in FlameNEST, and the work by the author to resolve this and allow full calculation. Unlocking the full calculation, the author had to resolve several issues in FlameNEST where the defined functions were not fully differentiable. In the flamedisx method, there still remains the requirement MC simulations to estimate total rates of sources, which are smaller than those required to produce a full PDF. The author discusses several possible resolutions and preliminary work on these problems, but development is ongoing.

The author makes some distinctions here for the rest of the work. Observables are random variables of the likelihood model of an experiment, e.g. the measured energy of an event in an experiment, and shape-varying parameters determine the shape of the model, e.g the energy resolution and rate of events, and for the sake of MC simulations both constitute dimensions of the model. A further distinction here of separable and in-separable dimensions is required; two separable dimensions would allow the model to be broken into a product of two independent models. For example, if the resolution of a detector changed with time, the measured energy and time would be inseparable. If the model is accurate, the inseparable parameters would be correlated, but observed data can still correlate separable parameters.

4.1 Why go beyond Monte Carlo?

Many experiments, including LZ, have detector and microphysics models that can change over the geometry of the detector or time. For example, in LZ often only (c)S1 and (c)S2 are considered as observables in inference and calibrations, but LCEs have a position and time dependence, and the drift field may vary impacting the microphysics. Whilst an analyser can try to correct out these variations, for example tuning a model to an average field and using spatially and temporally correct cS1 and cS2, this will come with some systematic modelling error. If the conditions requiring these corrections change rapidly enough in some observable it may introduce some statistical uncertainty. The correction route also requires lost live time to carry out sufficient calibrations or the use of extrapolations from different regions of the observable space.

In addition to observables these models often come with parameters that must be calibrated, for example the detector parameters g_1 and g^1_{gas} or the NR charge yield microphysics parameters in Table 2.2 of Section 2.2.1. In a rare event search these often require dedicated calibration campaigns, as the observed science data provide insufficient information to constrain them using standard tools. This comes with some uncertainty that is often difficult to propagate to the science inference, due to the limitation of MC tools, so excessive calibrations are carried out to achieve high precision. Even with a reduced set of observables, computational constraints require using modes of distributions such as the bands used in Section 2.2.3 which require longer

calibrations with higher rates to derive the modes of the distribution. Parameters such as the total rate of each background source are easy to implement as scaling of model PDFs. However, shape-varying parameters that vary the shape of those PDFs in any of the observables are a unique problem, as one needs a new MC simulation for each value, in essence, a new dimension.

Whilst shape-varying parameters and observables are distinct in inference, the latter being fixed to the observed data, they both constitute dimensions of the model. This is because if MC were used to create the model PDFs the scaling to re-evaluate the MC for each considered parameter is identical to that of another observable. Usually, the likelihood is defined as a combination of PDFs, which are normalised templates of the results of MC simulations that describe the source, signal or background, in some observable space for a fixed set of parameters. Once created, these are quick to evaluate. For example, in LZ's 2024 WIMP search, which had a likelihood with approximately 12 background models and 2k events, a MC-template based fit took $\mathcal{O}(1)$s to fit. To include shape-varying parameters, templates may be generated at fixed points in the parameter range and interpolated over, which is not necessarily prohibitively slow. The prohibitive step is to generate the required templates, especially with more observables, the number of simulated events required to create a smooth template scales exponentially, $\mathcal{O}(n^d)$ for d dimensions and n simulations. Generation of these simulations requires up to geological times with an increase in dimensions required [145].

To summarise, the limitation of time to produce MC simulations usually means that the parameters are calibrated to high precision using large calibration campaigns and fixed in science-objective inference. These calibration campaigns take time from data collection for science, such as the WIMP search, and frequent calibrations must be taken to track any time variation. Furthermore, due to the curse of dimensionality, these calibrations require using a reduced set of observables. The reduced observables mean that the model used in science searches is averaged over the lost observables, and variations require correction of the data. It is possible that this procedure will come with systematic and statistical uncertainty, which would be important to include in the final inference. Furthermore, being able to use the background data in your ROI to calibrate detector parameters would be a powerful tool to reduce the time lost to calibration runs. An example of how observables and shape varying parameters are inter-twined and may even impact corrected observables in LZ is given in Figure 4.1.

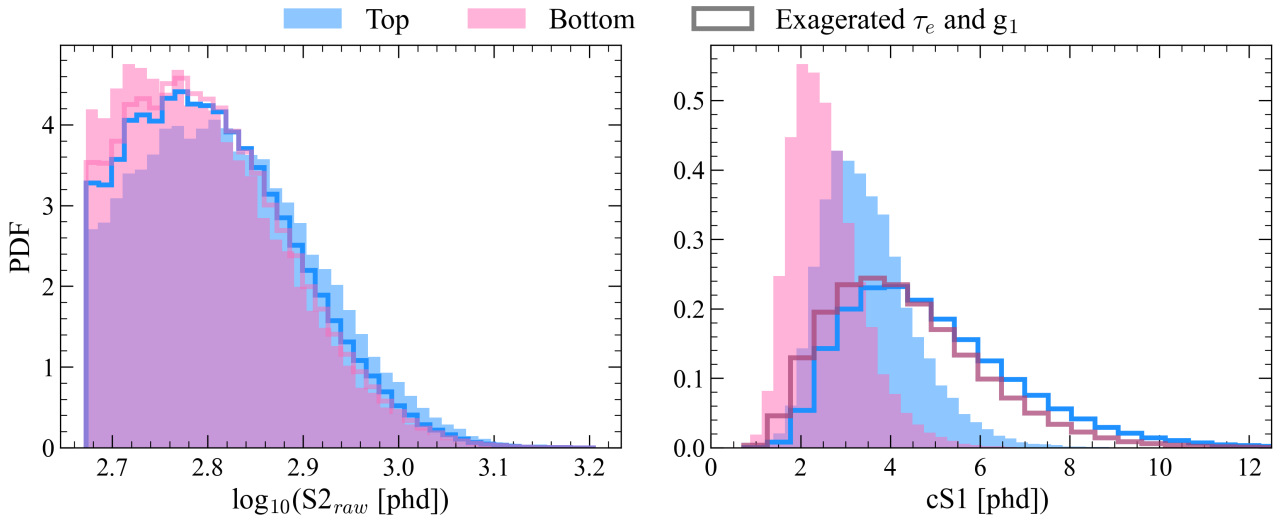


Figure 4.1: Projections in $S2_{raw}$ and cS1, for a ^{8}B CEνNS source, which is near threshold, using WS22 configuration with no ROI cuts and 3-coincidence requirement. The high value S1 LCE map at the bottom of the detector enhances the number of low cS1 events passing the 3-coincidence requirement. The finite electron lifetime reduces the $S2_{raw}$ of events from the bottom of the detector. Exaggerating the electron lifetime, τ_e, and the S1 total light collection efficiency, g_1, to non-physical values reduces the impact. This shows the interplay of dimensionality and shape-varying parameters required to accurately model this source.

4.2 Flamedisx method and `flameNEST`

The flamedisx method aims to eliminate the need for MC simulation by evaluating the models explicitly; it is better to demonstrate this method in contrast to MC templates. This explicit calculation requires information from the data set to constrain the number of evaluations of the model. Once a data set is available the calculation can be constructed in an auto-differentiable codebase, here `TensorFlow`, for fast reevaluation and gradients for maximising the likelihood. The new construction requires a modified log-likelihood term and still requires MC simulations, albeit smaller than for a full PDF, for estimating the total rates in a Poisson likelihood. In this section, the author covers an example of the flamedisx methodology, the new log-likelihood function, the key aspect of NEST implementation in `flameNEST` relevant for this work, and rate estimation.

4.2.1 `flamedisx`: Explicit Likelihoods vs Monte Carlo

To explain the explicit likelihood method it is best to begin with a simple experiment, as in Ref. [145]: a radioactive source produces n true events according to a Poisson distribution with a rate, λ, and an analogue measuring equipment Gaussian smears the number of events measured by some resolution, σ, to produce a signal, s, resulting in the likelihood of seeing s

being;

$$L(s) = \sum_{n} Gaus(s|n, \sigma) \times Pois(n|\lambda). \tag{4.1}$$

The MC method to evaluate this likelihood would be to have some theoretical value of lambda, then sample from the Poisson an inverse cumulative distribution for true events, n, then use those true events to sample observed events, s, from the Gaussian. This would require approximately $\mathcal{O}(10^3 - 10^6)$ MC simulations, depending on the accuracy required. If the parameters needed to be determined from a fit to observed data, multiple templates are produced across the parameter values and interpolated between.

The flamedisx method takes a different approach; the observed s is a random variable and a probable range of n is chosen, which would significantly contribute to the likelihood, here $n \in [s - 5\sqrt{s}, ..., s - \sqrt{s}, s, s + \sqrt{s}, ..., s + 5\sqrt{s}]$. Then Equation 4.1 is evaluated in this range, under the assumption that all values outside this range have a negligible contribution to the sum. This assumption is more complex for different responses, and a more general approach using a hybrid with some small MC simulations was developed in Ref.[127]. It is important to note that the flamedisx method will scale exponentially with the dimension of these guesses, but in many applications observables like position and time are definite and can be treated as fixed without a guessed range.

To plot the PDF in Figure 4.2, with $\lambda = 1$ and $\sigma = 0.5$, only a maximum of 42 terms of the sum were required for each of the 84 signal points considered, for a total of $\mathcal{O}(10^2 - 10^3)$. For MC simulations $\mathcal{O}(10^5)$ toys were used, but any number of signal points at those parameter values could be evaluated once generated.

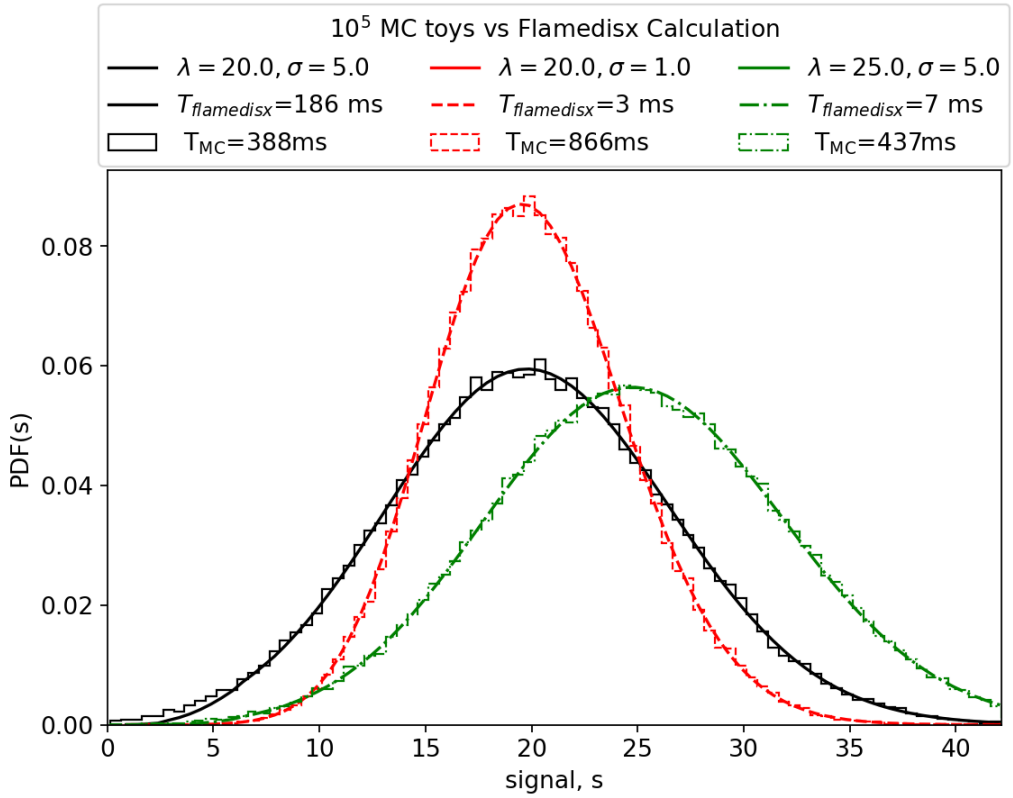


Figure 4.2: Calculating the probability for a range of signals following equation 4.1 using the method vs MC. The separation at low signal is solved using a continuity correction not applied in this toy model. The legend shows varying parameters of the model and reduced re-computation time of the flamedisx method due to the use of the saved computation graph compared to using MC.

The true power of this method comes when evaluating with different parameters of the model, if one needed to determine them from observations or there was some uncertainty that needed to be factored in for inference. In this example, fewer calculations are required per λ, and this process is faster since an explicit function of the parameters exists once the range of n is chosen. However, this may scale poorly with larger data sets or more complicated models, as the calculation needs to be carried out for each observed signal.

Even further optimisation can be achieved by representing the calculation as a linear algebra problem, with the range of possible n as a vector, then the equation 4.1 is a tensor product of two vectors. From here the range of values for the contracted variables in a model is called the quanta tensor. Evaluating this in `TensorFlow`'s differentiable programming creates a graph of primitive operations with known gradients, giving several advantages;

- Primitive functions are GPU-optimised operations.
- Graph representation allows for rapid recalculation of the function (only connecting nodes).

- Auto-differentiation using known derivatives of primitive functions and chain/product rule.

In figure 4.2, the flamedisx method is similar to MC for the first calculation, but subsequent calculations for different values of the parameters far outperform MC by two orders of magnitude. This also introduces a key aspect of using `TensorFlow`, evaluating the first set of parameters builds the graph for the first time but subsequent calls are $50\times$ faster; this is called tracing time. Using auto-differentiation constructs a second graph of the derivative of the first, which again requires tracing. The subsequent gradient graph can use more memory, as the chain and product rules expand the terms to consider. However, these derivatives allow for faster optimisation algorithms, which is especially useful with a large number of parameters.

4.2.2 New Extended Log-Likelihood Function

As this method requires the energy spectrum to produce accurate guesses for the quanta tensors it has the sources' differential energy spectrum folded into the calculation. This requires a different object from PDFs to be used, called the differential rate, $R_j(\vec{s},\vec{\theta})$. The two can be compared by,

$$\nu_j(\vec{\theta}) = \rho \int R_j(\vec{s},\vec{\theta}) d^D\vec{s} = \nu_j(\vec{\theta}) \int f_j(\vec{s},\vec{\theta}) d^D\vec{s} \tag{4.2}$$

where ν_j,R_j, and f_j are the expected counts, differential rate, and PDF from source j, respectively. The ρ term indicates exposure, mass, or density depending on the definition of rate used. As in Chapter 3, $\vec{s}$ indicates a vector of D observables and $\vec{\theta}$ is the parameters of the source model, e.g microphysics and detector parameters. From this the extended likelihood in equation 3.7 becomes

$$L(\vec{\nu}^{r.m},\vec{\theta}|\{\vec{s}_{obs}\}) = \frac{\nu(\vec{\theta})^n e^{-\nu(\vec{\theta})}}{n!} C(\vec{\nu}^{r.m},\vec{\theta}) \prod_{\vec{s}_i \in \{\vec{s}_{obs}\}} \sum_j^{sources} \frac{\nu_j^{r.m} R_j(\vec{\theta})}{\nu(\vec{\theta})}.$$

where $C(\vec{\nu}^{r.m},\vec{\theta})$ is a constraint on the rates and parameters of the model and n is the total observed counts. In order to allow the rate of each source to be scaled, a simple modification of $\nu_j^{r.m}$ is added to allow the predicted counts and differential rate to be uniformly scaled as the expected rate changes. The total expected rate would be given by $\nu(\vec{\theta}) = \sum_j \nu_j^{r.m}\nu_j(\vec{\theta})$. Finally, the log-likelihood can be expressed as

$$lnL(\vec{\mu},\vec{\theta}|\{\vec{s}_{obs}\}) = \sum_j \nu_j^{r.m}\nu_j(\vec{\theta}) - \sum_{\vec{s}_i \in \{\vec{s}_{obs}\}} ln(\sum_j^{sources} \nu_j^{r.m} R_j(\vec{\theta})) + ln(C(\vec{\nu}_{r.m},\vec{\theta})). \tag{4.3}$$

Constants have been ignored because they do not impact inference. Throughout the rest of the work discussed here, this object is used in fitting.

4.2.3 `flameNEST`: using the NEST and `LZLAMA` Models

Like in the example above, the mapping from simulation to explicit inference requires assuming the existence of a guess of values for the relevant random variables, the quanta tensor. The NEST models introduced an increased level of complexity to these quanta tensors, which caused significant performance issues. It helps to diagrammatically compare the simple models of `flamedisx` LXe microphysics to those of the NEST models. Considering the models from a MC perspective gives,

Simple Model	NEST ER [NR] Model
$E \sim R_0(E)$	$E \sim R_0(E)$
$n_q \sim \text{Binom}(f(E))$	$n_q[n_{ion}] \sim \text{Norm}(f(E))$
$n_{el} \sim \text{Binom}(f(n_q))$	$n_{ion}[n_q] \sim \text{Binom[Norm]}(f(E, n_q[n_{ion}]))$
$n_{ph} = n_q - n_{el}$	$n_{el} \sim \text{Norm}(f(E, n_{ion}), n_{el} > n_{ion})$
	$n_{ph} = n_q - n_{el}$

As only the microphysics is discussed here, and not detection, the produced label of n_k^{prod} is dropped. Here, the dependencies on previous simulated variables have been absorbed into a short hand, f, which would be different for each step ignoring parameters of the model. The `flamedisx` representation of these models would be,

$$R_{simple}(n_{ph}, n_{el}) = \delta(n_q = n_{ph} + n_{el}) \sum_{n_q} Binom(n_{el}|n_q) \sum_{E} Binom(n_q|f(E))R_0(E),$$

$$R_{NEST}(n_{ph}, n_{el}) = \delta(n_q = n_{ph} + n_{el}) \sum_{n_q, n_{ion}, E} Norm(n_{ion}|f(E, n_{ion}), n_{el} > n_{ion}) \times Binom(n_{ion}|f(E, n_q))Norm(n_q|f(E))R_0(E).$$

Where the values n_k are now the quanta tensor of guesses for those species, and the species $k \in \{el, ph, ion, q\}$ labels the electrons produced, photons produced, ions produced, and total quanta produced . The simple model may be broken, as shown in the sums, into the contraction of 3 functions evaluated on quanta tensors of dimensions 2D, 2D, and 1D with dimensions d_{el} x d_q, d_q x d_E, and d_E, respectively. Here the distinction of dimension of the quanta tensor and quanta tensor is made. The quanta tensor n_{el} represents in the calculation the number of electrons produced, but it has many guessed values of d_{el}. As the model block depends on quanta produced, they must both be constructed in a tensor of $d_{el} \times d_q$.

The sum terms have been moved to explicitly show where contractions occur. Unfortunately, the same cannot be said for the NEST model where, due to their energy dependence, the summation cannot be separated and the ions introduce a new quanta dimension which causes a single 4D tensor of d_{el} x d_q x d_{ion} x d_E. As discussed in chapter 2.2, the NEST models are empirical, but the energy dependence of each step of the microphysics probably reflects the size and

shape of the energy deposit and the ions dimension is required to properly model recombination, which would also depend on the size of the deposit. Perhaps it is possible to reduce the dimensionality and still capture all the required behaviour, but this was not explored in this work.

`TensorFlow` is best optimised for operations over dimensions of a tensor. This means that in constructing the quanta tensors and evaluating models for a data set of n events, instead of processing n quanta tensors it is better to add an additional dimension n_{events} to each and processes n/n_{events} of the larger tensors. This is known as batching, where n_{events} is the batch size, and is limited by the increased memory consumption for now $n_{events}\times$ larger tensors. Within a batch, events can have varying guessed quanta values, as long as the dimensionality of each and the energy guess is the same. The impact on computation time may vary, and is quantified in the following section.

4.2.4 Rate Estimation

Here the author focuses on improving the calculation of differential rates, $R_j(\vec{\theta})$ in Equation 4.3. The total rate term, $\nu_j(\vec{\theta})$, still needs to be estimated by MC ideally by

$$\nu_j(\vec{\theta}) = \frac{1}{N}\sum_i \frac{R(\vec{\theta},\vec{s}_i)}{Prior(\vec{s}_i)}, \vec{s}_i \sim Prior(\vec{s}_i) \tag{4.4}$$

using some prior. This would allow any number of parameters to be varied using the explicit calculation of $R(\vec{\theta},\vec{s}_i)$. To accurately estimate the integral one would need to calculate $R(\vec{\theta},\vec{s}_i)$ for $\mathcal{O}(10^5)$ simulated events for sub 1% variance, as the variance scales as $N^{-1/2}$. This is currently not feasible, so each event is simulated from the full response, so the result is essentially

$$\nu_j(\vec{\theta}) = \frac{N_{survive}}{N}, N \sim R_j(\vec{\theta}) \tag{4.5}$$

where N events are simulated but only $N_{survive}$ pass a Bernoulli trial from the various acceptances. In general, far fewer events are required to form a precise estimate of the total rate than a smooth PDF or differential rate histogram. The variance in the total rate scales as $\frac{1}{\sqrt{N}}$ and does not scale with observables, but the accuracy of the integral can depend on a constant of integration which scales with observables and inversely with the size of simulations; see discussion on this in book Ref.[146] and public note Ref.[147]. So, it is not completely accurate to say that the accuracy estimation for a single point in the parameter space $\vec{\theta}$ is independent of the number of observables. Nevertheless, it has been found that 10^6 events are sufficient to accurately estimate the total rate of the 6D `flameNEST` model, which requires $\mathcal{O}$(10s) to evaluate. In `flameNEST` results are calculated over a grid and linearly interpolated in `TensorFlow`.

The issue with MC rate estimation is scaling with inseparable parameters. For separable parameters two single functions can be evaluated $\nu(\vec{\theta}_1,\vec{\theta}_2) = \nu(\vec{\theta}_1)\times\nu(\vec{\theta}_2)$ so the computation

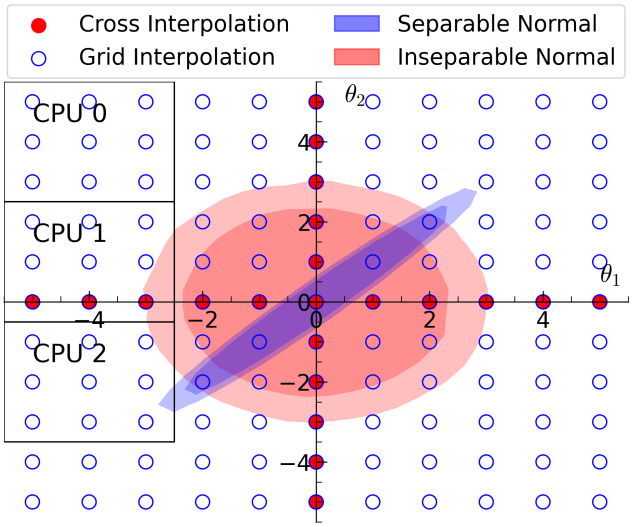


Figure 4.3: Diagram of separable and inseparable variables where the axis into the page represents the total rate. Also shown is the batching of the grid for deployment on multiple threads.

scales with $\mathcal{O}(n_{sep} \times n)$ where n_{sep} is the number of separable parameters and n the simulation size. However, a grid is required for inseparable parameters, scaling as $\mathcal{O}(n^{n_{insep}})$ where n_{insep} is the number of inseparable parameters. This is presented diagrammatically in Figure 4.3. In order to reduce the required wall time, the author created a custom grid-interpolator that supported parallel processing for generation of the anchor points. The yield parameters in NEST, presented in Table 2.2, all impact the NR charge and light yields, making them analytically inseparable. Perhaps, if two parameters greatly impact different energy domains of the yield they could be treated as separable. An example of separable parameters in total rate, would be g^1 and g^1_{gas} which each independently impact S1 and S2, respectively. Even then, one could imagine they would become inseparable if a line-cut was implemented as a function of the two or a source was near a threshold in both. It should be noted that separable parameters can be correlated in a fit to a data set including shape information, i.e the differential rates, may well show correlation between two parameters. For example g^1 and g^1_{gas} are always anti-correlated in this manner.

4.2.5 Summary

The `flamedisx` method is superior to MC when there are multiple observables, especially if only a subset of these require quanta tensors. For example, in NEST if there is negligible position and time uncertainty, only the (S1,S2) models need to be evaluated with observed position and time fixed. Unlike MC it scales linearly in the calculation of differential rates, or the shape term in the likelihood, with the number of inseparable shape varying parameters. It is worth noting that it also scales linearly with the observed data and requires creating quanta tensors for each new data set. Therefore, if few observables are required and the data is shape insensitive,

MC PDFs may be better for using non-asymptotic methods which require evaluating fits on $> \mathcal{O}(10^4)$ toy data sets.

4.3 FlameNEST model optimisation

The author quantified the performance of `flameNEST` and determined the largest bottleneck. They then implemented an upgrade to resolve this without changing the existing structure of the calculation. The author also took care to make sure that it was compatible with future applications. Finally, a test case was performed and the remaining outstanding issue for `flameNEST` applications are identified and discussed.

4.3.1 Quantifying Performance

To quantify performance, there are three things to consider; time to run a program, available resources, and memory consumed. The available resources are Central Processing Units (CPUs) and GPUs, often the former are more available on computing clusters but are less tuned to repetitive computation than the latter. If the structure of a calculation is changing over a program, such as constructing new data structures in simulations, then CPUs are preferred. The available resources control the computation time as if a GPU runs 4 times faster but you have 10 times more CPUs available, and your calculation can be broken into separate parallel threads; the latter is preferred. Memory constrains the number of resources available; there is a fixed number of CPU or GPU cores that share a set of memory. Therefore, the consumed memory of each process directly constrains the number of parallel processes that one can run. This logic also applies to the parallel processing on threads in GPUs, and it is important to consume as much of its resources, and therefore memory.

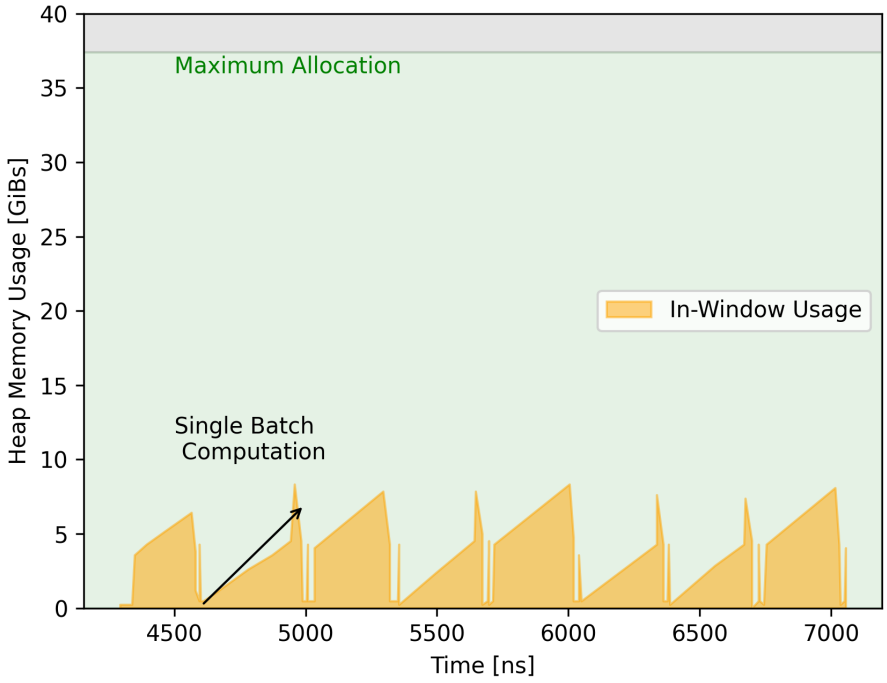


Figure 4.4: Example of profiling the differential rates. Heap memory is that which is used dynamically over a program, and here it grows in each batch as the parallel sum over energies occurs. There is a pre-window heap usage, which is caused by tracing the function and can be the largest.

The memory profile for calculating the differential rates in `flameNEST` is shown in Figure 4.4. As can be seen there is a peak in memory, the origin is discussed in the next section, the size of this peak determines how much of the GPU resources can be used. It should be noted that ideally this would be a flat consumption, as the peak is consuming resources intermittently, but the number of processes are limited by it. The pre-window memory consumption appears to dominate in `flameNEST`, particularly after the updates applied in this section, due to the tracing time. Future improvements to the differential rate calculation should focus on reducing the pre- and in-window memory peaks, even at the cost of higher total memory consumption and increased computation time. The author leaves distinction of pre and post window for Appendix A. For this work only the computation time and maximum memory of either window is considered, as that is what constrains the amount of resources available.

4.3.2 Performance of `flameNEST`

The implementation of the NEST models leads to significantly higher memory consumption and computation time for differential rates. This is not only from the increase in the dimensionality of the quanta tensor, but from the increased complexity of the computation, shown graphically in Figure 4.5. The energy dimension is usually the largest, so a parallel sum over energy was implemented to improve the time performance, but this further adds to the memory

consumption. For the following discussion, the author refers to this section of the calculation concerning microphysics as the Central Block (CB).

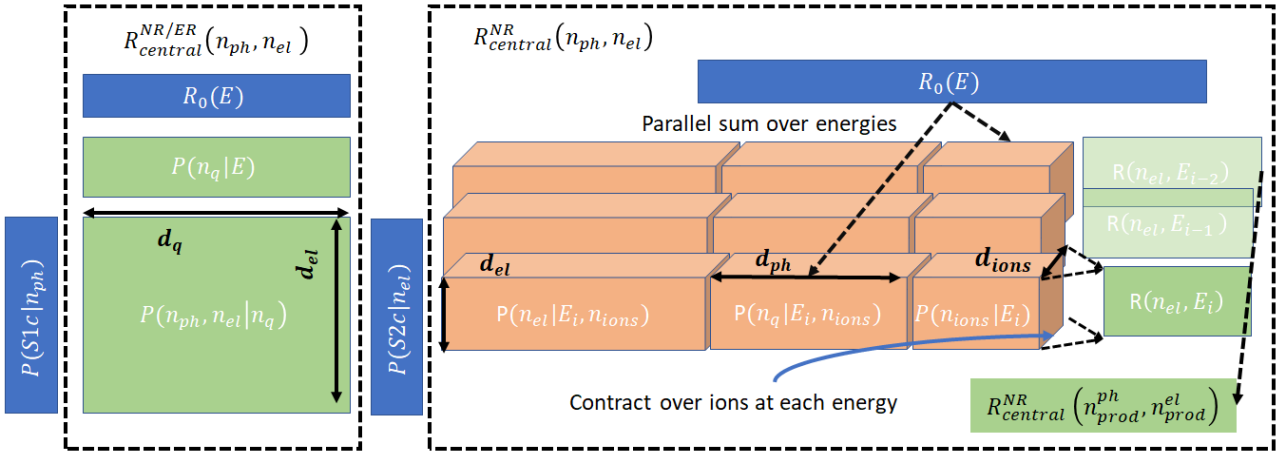


Figure 4.5: Diagram of the original `flamedisx` models (left) and implementation of the NEST models (right). Double edged solid arrows indicate a dimension of the quanta tensors and dashed arrows on the right indicate a parallelised sum over the energies contributing to an event and contraction over the ion dimension per energy. Blue, green and orange indicate 1,2 and 3 dimensional tensors, respectively. Batching is treated implicitly here.

This method worked for fitting detector parameters outside the CB to calibration data as `TensorFlow` would only need to compute the complex CB once, as long as the parameters inside it were fixed and not allowed to be variables in the likelihood. This is because `TensorFlow` can choose the complexity of the graph it needs to construct and save based on the input signatures of a given function[148]. Once the parameters of the CB models were added to the function as inputs, memory consumption grew exceptionally large, as seen in Table 4.1.

Differential Rates	Peak Memory (GiBs)	Trace time (MM:SS)	ex.time (SS)	Top Operations
det.param g1 batch size 5	12.0	02:05	02	CB TFP functions tensor ops
yield.param α batch size 1	28.0	04:13	02	CB TFP functions and gradients of them

Table 4.1: Profiles of memory usage of calculating the values of the differential rate, not differentiating, of all events in a data set with floating a shape parameter in the detector functions, g1, and one in the yields. A data set of size 50 was used for the former, and 10 for the latter to avoid CPU memory issues. Captured utilising `TensorFlow` profiling tool, tensorboard [148]. `TensorFlow` Probabilities (TFP) package represents model evaluations.

The memory usage presented in Table 4.1 is too large; the state-of-the-art computing cluster at the National Energy Research Scientific computing Centre (NERSC) provides GPUs with maximum random access memory of 40 GiBs [149]. As memory usage is directly proportional to batch size, processing any more than 1 event at a time would crash the execution, and doing

so for any appreciably sized data set would lead to memory issues inside the CPU-GPU memory exchange. Further, whilst the memory is manageable for non-CB parameters, Dr Robert James in Ref.[111] required significant compromises to allow just multidimensional inference and only managed to fit detector parameters in a calibration fit by reducing the size of the energy tensor, compromising accuracy. Here, it is important to note that Table 4.1 is only the calculation of the value of the differential rate, not its first or second order gradients, which would consume significantly more memory.

4.3.3 The Solution: Reduce Degeneracy

In Figure 4.5 (right) one can see that the quanta tensors for the `flameNEST` models are implemented with the dimensions of $n_{events} \times d_{el} \times d_{ph} \times d_{ion}$ in a parallel sum over energy, where n_{events} represents the number of events in the batch. To achieve this construction, for example, the range of the number of ions is guessed for each event in the batch, the n_{ions} quanta tensor has dimensions $n_{events} \times d_{ion}$ but is repeated across $d_{el} \times d_{ph}$ so that the total matrix product could be performed. The same is true for the electrons and photons produced. The important thing to note is that for a given energy, the models do not operate explicitly on all three quanta dimensions. Taking the n_{events} dimension as implicit, the actual dimensions are a composite of 1D $P^{NR}_{ions}(n_{ion}|E_i)$, 2D $P^{NR}_{quanta}(n_q|E_i, n_{ion})$, and 2D $P^{NR}_{electrons}(n_{el}|E_i, n_{ion})$. This means that many of the evaluations of the function are repeated unnecessarily, i.e there is degeneracy. To understand why this has caused a memory issue, recall that each evaluation of a function in auto-differentiable programming represents a graph of concrete functions. So, each element corresponds to a larger memory consumption in the graphical computation than just its value. This grows as the model is differentiated and derivative graphs are built.

In order to improve performance, the author focused on reducing degeneracy and calculating only unique elements. The ions tensor was the easiest, as it was a simple task of not repeating it along the other dimensions. The quanta tensor was slightly more technical, it is derived from the sum of n_{el} and n_{ph} so it technically has dimensions $n_{events} \times d_{el} \times d_{ph}$. However, since n_{el} is just repeated along d_{ph}, and vice versa, there are many repeated elements of the produced tensor. Failing to find a nice matrix operation to manage this, the author used `TensorFlow`'s unique function to identify the unique elements. This produced an index that could project back to the original dimensions with the gather function. This would actually reduce the n_{events} dimension, leading to another performance improvement as, if events with similar energies were batched together, they would have very similar total quanta values. However, this increases the complexity of the gather operation with batch size, potentially contributing to diminishing returns on performance with batch size seen later in Figure 4.13. A similar procedure was applied to the electrons produced quanta tensor, post total quanta calculation, also improving performance.

The above discussion omits technical details of how the indices were used which can be found in the Appendix A. The NR calculation is profiled and tested here, but results for the ER are shown in the appendix. Here, the author presents a case where the CB is uniform across events, this is the case in NEST for a uniform drift field, the non-uniform case is also handled in the appendix. The current implementation of this carries out all the indexing and mapping in the same single parallel sum over energies. In the new paradigm, it may be more efficient to remove this parallel sum of the CB and operate on tensors with an energy dimension. This would trade increased time for reduced and flattened memory consumption. As the computation is sufficient for the purposes of this work, this is left to future developers.

4.3.4 Quantifying Solution Performance

As seen in Table 4.2, the above implementation reduces the memory usage when using CB parameters by a factor of 28 and when using non-CB parameters by a factor of 6. Tracing the graph takes longer, as expected with the construction of the unique tensors and indices, but the subsequent execution time does not change significantly. This makes computation of CB parameters feasible, but also speeds up non-CB parameters, as the batch size can be increased with the decreased memory usage to process a dataset faster.

Differential Rates	Peak Memory (GiBs)	Trace time (MM:SS)	ex.time (SS)	Top Operations
det.param g1 batch size 5	2.0	08:16	02	gather/tensordot tensor ops
yield.param α batch size 1	1.0	04:25	01	gather/tensordot tensor ops

Table 4.2: Identical profiling to table 4.1 with updated calculation. Gather and tensordot represent the mapping and scalar product discussed in section 4.3.3.

To verify that the modifications had not introduced an error in the computation of differential rates, the author repeated the consistency checks from Ref.[127]. These use the MC simulate feature, which were unmodified here, to produce a binned differential rate from events at the centre of the detector, given by

$$R(cs1_i, cs2_i)|_{(\vec{r}=\vec{r_0})} = \frac{\mu_0}{N} \frac{H_i}{\Delta cs1_i \Delta cs2_i}$$

for the i-th bin. Here μ_0 is the rate before the physics and detector response, given by the integral over the energy spectrum, and N is the number of simulated events, and $\vec{r}$ represents the space-time coordinates, or other observables. Fixing the other observables simply makes the computation of the differential rate via MC feasible, whilst the response varies over these observables nothing about the structure of the computation of differential rates does. If the structure were to change, for example there was some condition for high and low drift fields, one may wish to construct an alternative combination of dimensions for this test. The results

of the test are shown in Figure 4.6, showing that the new implementation is still consistent with the simulate step.

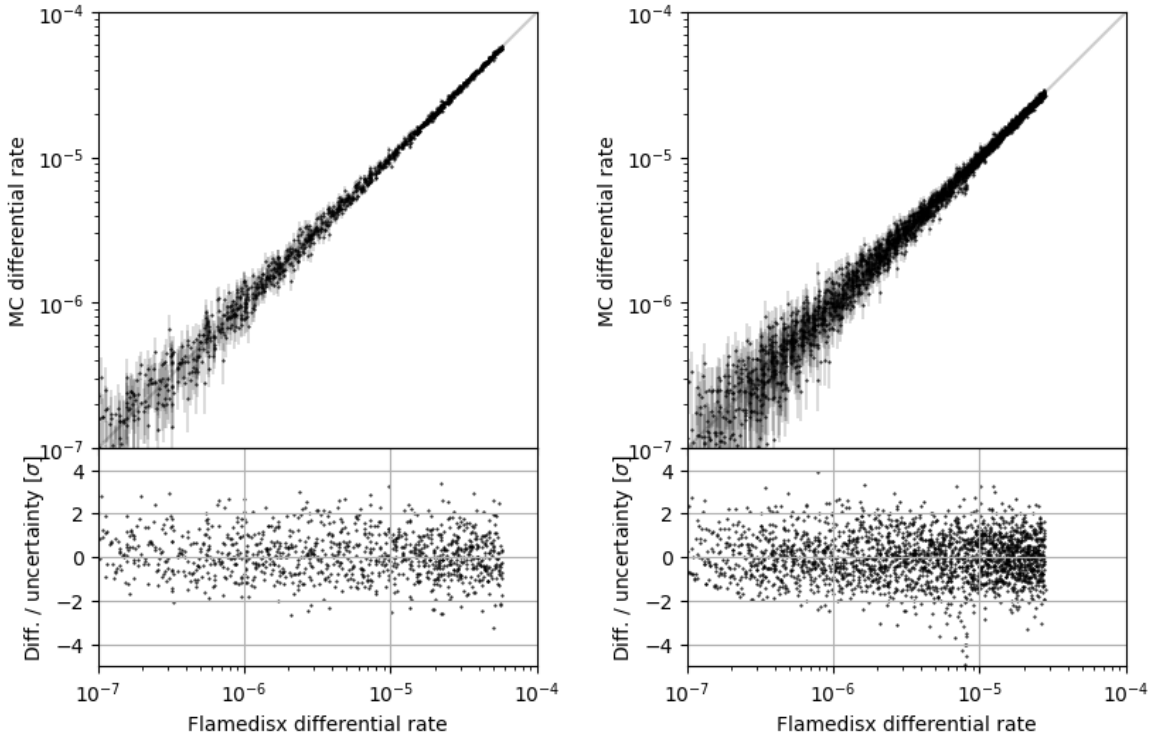


Figure 4.6: Comparison of differential rates derived from MC templates and from `flameNEST` differential rate for a mono-energetic 20 keV NR (left) and 5 keV ER (right). The templates used 10^6 simulated events.

In order to check that the differential rates values and derivatives, first gradient, and the Hessians were consistent with previous versions within numerical error. This check is important as modifications to the computation of differential rates may not change the value but perturb the gradients. This was checked by producing a simulated data set of $\mathcal{O}(10^3)$ events and calculating the three results several times. Then repeating with the updated computation. The variance from the original calculation is shown in Figure 4.7, showing that there is no significant change between the two versions of the calculation that are consistent within their accuracy.

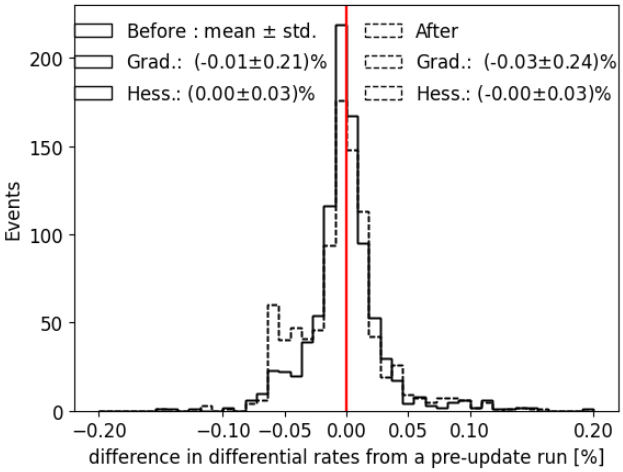


Figure 4.7: Variance of multiple calculations of the differential rate, gradient, and Hessians between calculations before and after the updates. All are compared to a single pre-update calculation.

Whilst developing the above checks the author resolved several issues with the implementation of the CB NEST models in `TensorFlow` ranging from undefined behaviour of core functions to models with divergent integrals. Here, the tests for the total model are shown, but the author notes that future developers should focus some resources on creating consistency checks for gradients and Hessians of each block of the model and their sub-functions.

4.4 Testing with Example Implementation

For the upgraded FlameNEST models, the author used a test case that would be difficult to model with existing MCMC band-fitting tools, discussed in Section 2.2.3, and using CB parameters that were previously impossible. The author chose a flat low-energy NR spectrum at 0-10 keV to model a near-threshold calibration source with a small sample size of 650 events in the region of interest of $S1c > 3$. $S2c > 10e^-$. As the parameters were highly correlated, only three were chosen that give the greatest impact on the rate within their uncertainty, as quoted in Table 2.2, which were α, β, and ϵ. These vary the total quanta yield and ϵ controls the point at which the charge yield changes its power law in Figure 4.8. Other parameters did not have a significant impact on the shape or rate for this source.

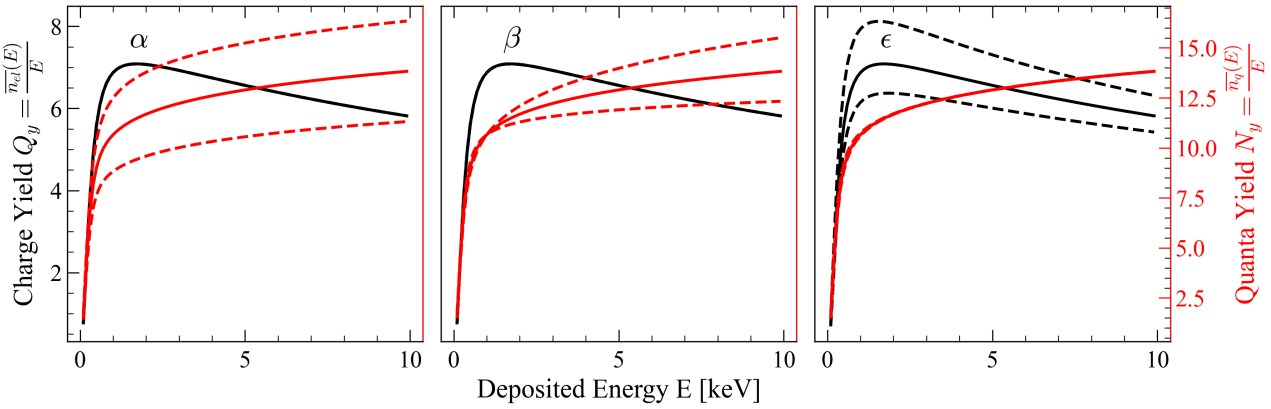


Figure 4.8: Impact of varying chosen parameters on the total quanta (right axis, red) and charge yield (left axis, black) within $\pm 1\sigma$ of the default NEST uncertainties. For asymmetric uncertainties, the largest is chosen. The parameters α and β control the total quanta yield where ϵ controls the charge yield

The parameters have an in-separable impact on the total rate term $\nu(\vec{\theta})$, as they all impact the same two variables n_{el} and n_{ph} over a similar domain. Therefore, a multi-dimensional linear grid was constructed spanning $3\sigma_{\pm}$ and 8 anchor points in each parameter, i.e using their asymmetric uncertainties. A total of 10^7 events were simulated per 8^3 anchor points, it is possible this was not an optimal choice. The results are shown in Figure 4.9.

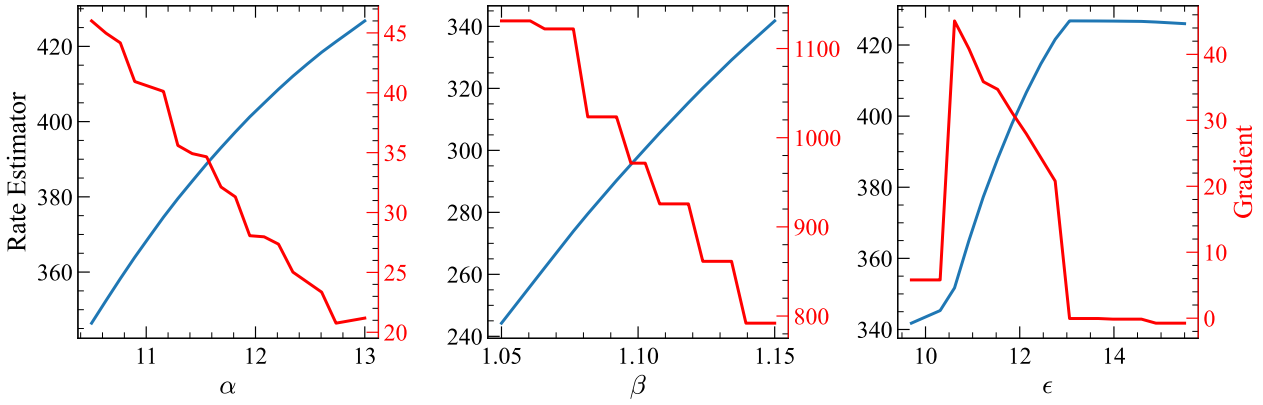


Figure 4.9: Projection in each parameter of the rate estimator that is a linear interpolation over a 3D grid. Also shown is the gradient with respect to each parameter, by definition the second order gradient, and therefore any contribution to the Hessians, will be zero.

In order to check the accuracy of the method, the author created a pull test. To produce these, samples of model parameters, $\vec{\theta}_0$, were generated using some prior information, here their constraint term. Then data sets are generated from those parameters and the Maximum Likelihood Estimators (MLEs), $\hat{\vec{\theta}}$, and estimate the uncertainty with the covariance matrix, $\vec{\sigma}$, calculated using the likelihood model. By definition, $\hat{\vec{\theta}}$ should be normally distributed around $\vec{\theta}_0$ with variance $\vec{\sigma}$. Technically, this means that the distribution used to sample $\vec{\theta}_0$ should not impact the result, but using the constraints weighs values to those more likely a priori.

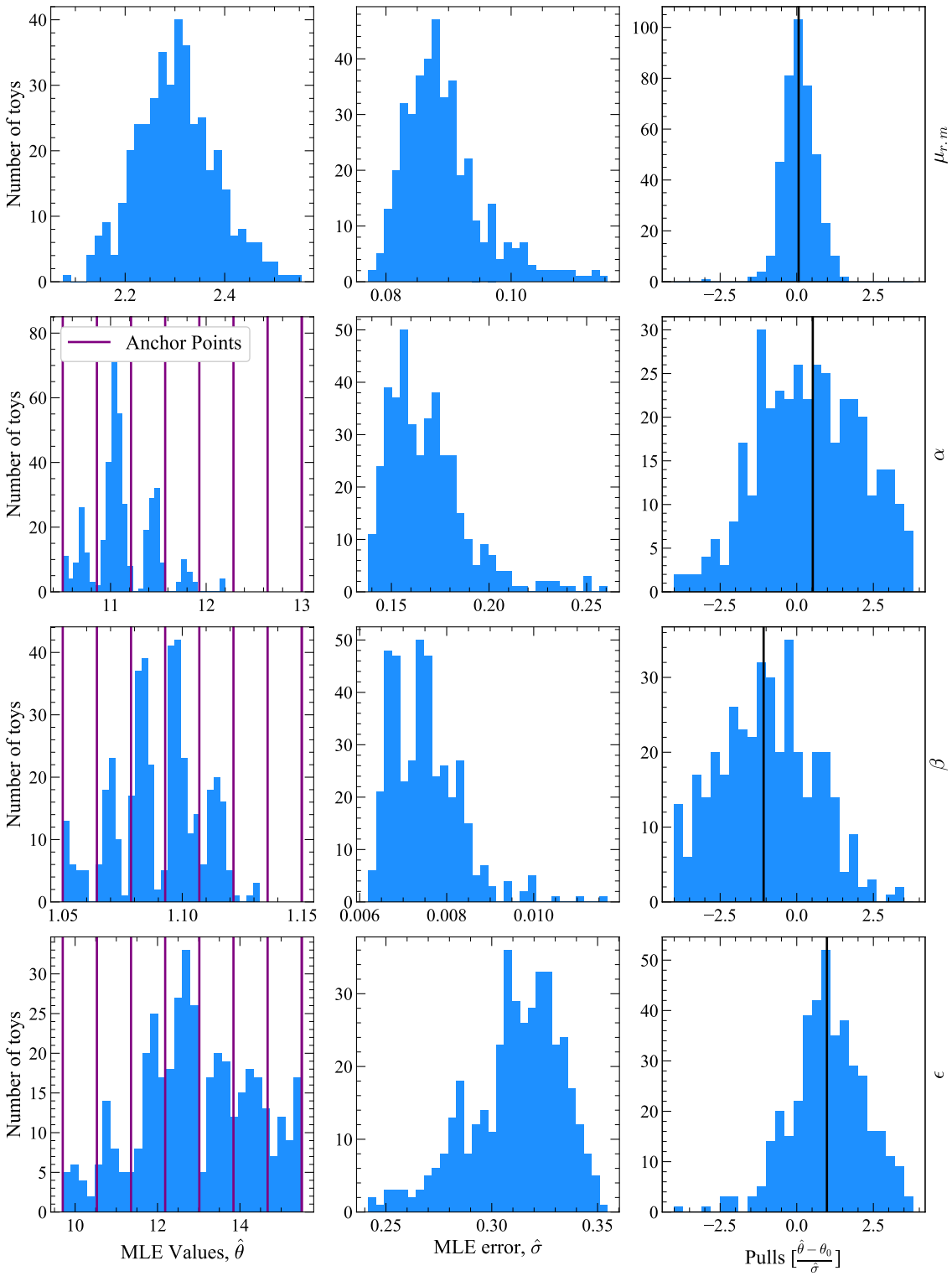


Figure 4.10: Pull plots of the test-case, each row shows a different parameter. The rate multiplier has a smaller variance as its sample value θ_0 was generated from the size of the data set and rate estimator at sampled values, i.e from the observed and not sampled value. That is from the observation not the prediction. The left column gives the varying MLEs with the value of the anchor points overlaid. The middle shows uncertainty on the MLE derived from the Hessian of the likelihood. The right column gives the pull, which should be a normal distribution.

The results of the pull test in Figure 4.10 show significant biases in the pulls and the MLEs. The latter appear to be distributed in a step-wise fashion correlated with the anchor points. This is due to the use of linear interpolation of the rate estimation, in Figure 4.9 the stepwise gradient and poorly defined Hessian would interfere with the optimiser. Turning off the gradients and Hessians for a multi-parameter fit would greatly increase the time required. For the case of separable parameters, this is readily resolved; in Section 5.2.3 the author uses cubic splines to resolve a similar issue in template morphing and in Section 6 a combination of 1D polynomial fit to separable parameters is introduced. For this problem, a 3D B-spline interpolator would be required as the S shape in ϵ is not well modelled by a multivariate polynomial or ordinary cubic spline. This is feasible within `TensorFlow`, but technically challenging and not necessary for the scope of this work.

As a presentation of the performance, a single example from the above pull test gives the computation time and shows how `flameNEST` achieves a shape varying fit with some bias in the parameters. As shown in Figure 4.11, the fit converges within the uncertainty on the pulled values, with a significant shape difference. The fit was performed in 6 observables and cost 11 minutes on Perlmutter's A100 NVIDIA 40GB GPU, a further 30 minutes of a CPU node utilising 128 threads was required for building the rate estimator.

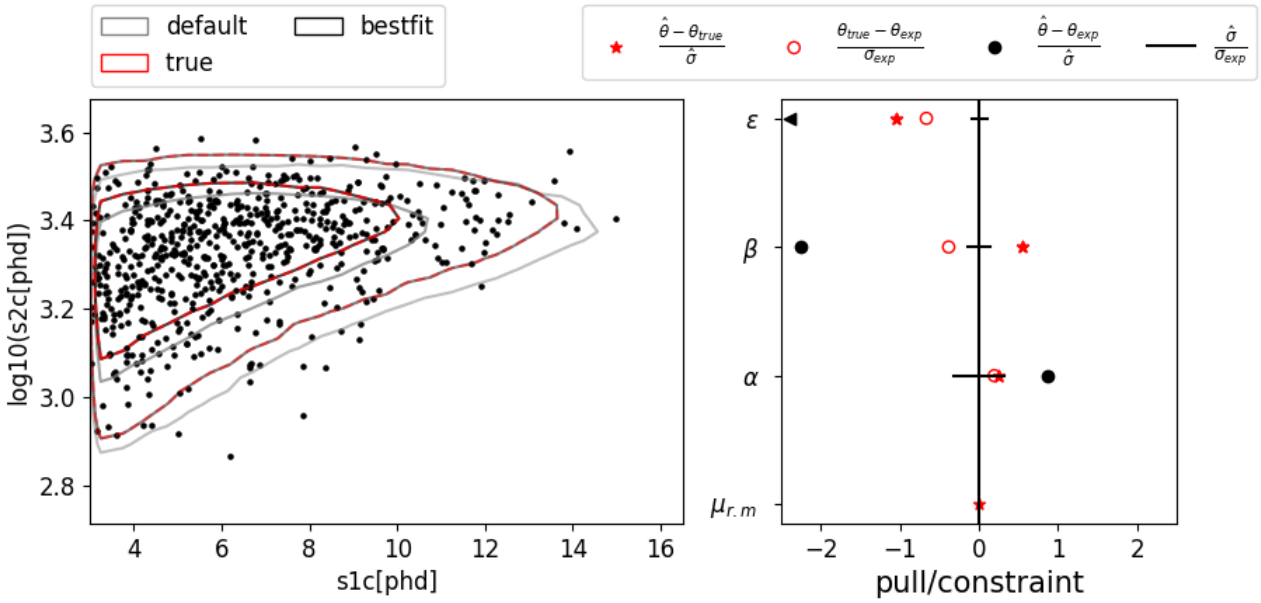


Figure 4.11: Performing a fit with the test source using perturbed parameters. The simulated data with contours showing the default, best fit, and true shape for this sample (left). The pull plot with several definitions of the pull (right). For this, the true value of a parameter used to generate the data, θ_{true}, is sampled from a Gaussian distribution centred on the pre-fit value, θ_{exp}, and uncertainty, σ_{exp}. The fit gives the MLE, or post-fit, value, $\hat{\theta}$ and uncertainty, $\hat{\sigma}$. The red hollow points give the perturbation to the data, the stars how close the fit is to the true value, and black points the distance from the expected. The horizontal lines give how much this data set constrained that parameter.

Using this single example, another potential issue with the rate-estimators can be investigated; it is possible that an insufficient number of anchor points can affect the fit. In Figure 4.12, there is a kink in the parameter scan of the likelihood suggesting that the differential rate term is changing between anchor points of the total rate. This may also impact the uncertainty calculated from the Hessian of the likelihood, as shown the intercept with $2\Delta lnL = 1$ does not line up with the derived uncertainty as expected.

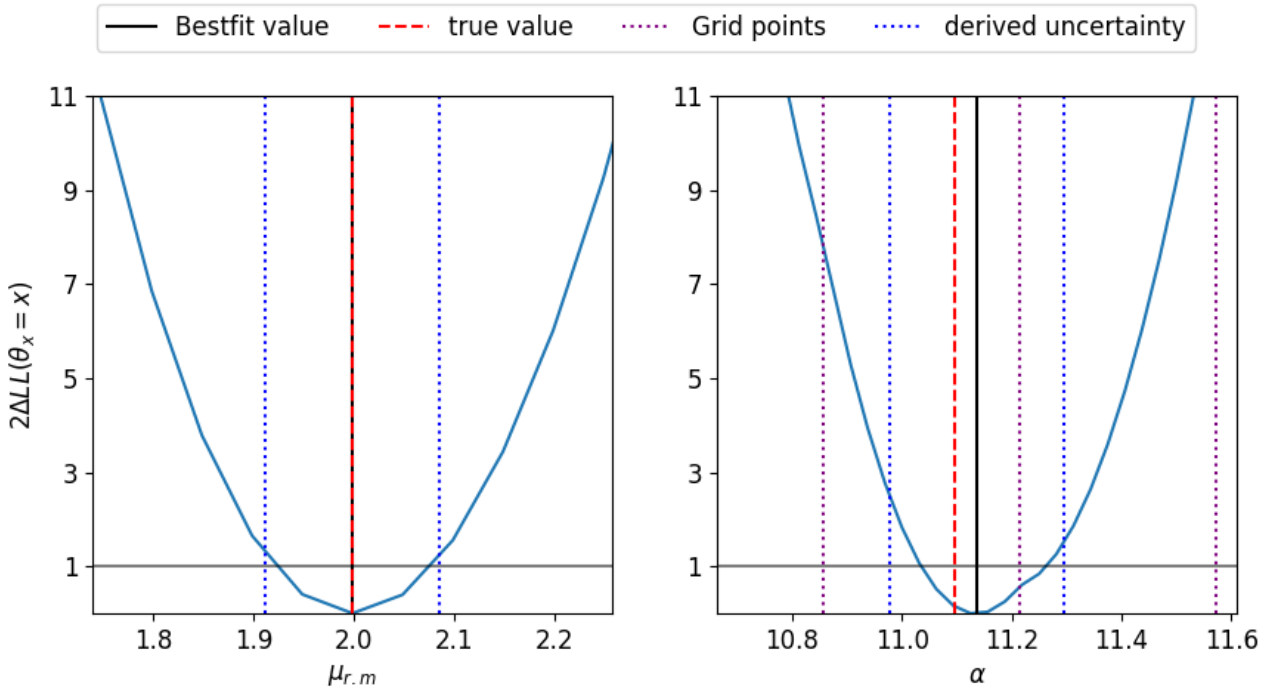


Figure 4.12: Likelihood scan over the rate multiplier and one of the parameters α. The horizontal line gives the approximate intercept of the 68% confidence interval if the likelihood were Gaussian. This is not a profiled likelihood scan so these do not intercept with the uncertainty derived from the Hessian. In the scan of α there is a kink in the likelihood near the minimum, this indicates that the differential rate is varying in a way that is not captured by the interpolated total rate term.

Finally, the author observed that there was a diminishing return in calculating the value, gradient, and Hessian of the likelihood on the entire data set with batch size within the updated `flameNEST`, as shown in Figure 4.13. This may be due to the fact that whilst increasing the batch size reduces computation for the entire data set, it would also increase the size of the unique quanta and electron dimensions counteracting gains after a certain point. This is not a prohibitive behaviour, and may have been present in previous versions, but is important to note when choosing optimal batch sizes.

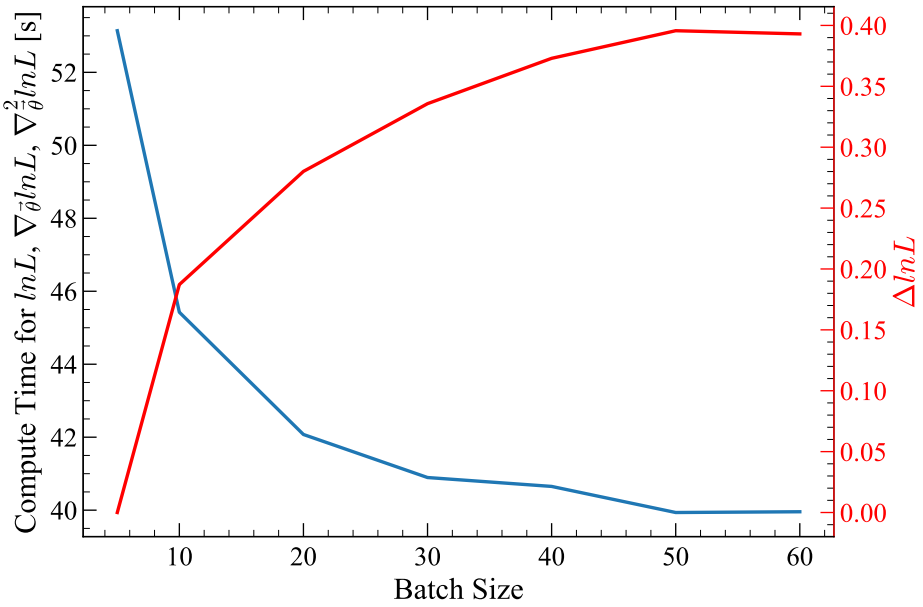


Figure 4.13: Timing the calculation of the value, gradient, and Hessian of the likelihood from above test for the entire data set of 650 events with varying batch size. Increasing the batch size may cause some error in the computation shown on the right axis.

4.5 Future Development and Implementation

There are two avenues for development in `flameNEST`, the computation of the total rate term and further optimisation of the differential rate term. The latter could be resolved by further modifications of the authors' upgrades; for example, it may be possible in the reduced dimensionality to remove the parallelisation over energy. This may help to flatten the peaky memory profile and allow for greater batch sizes, though it may not help in the light of the diminishing returns found by the author. For using the total rate term in inference the number of shape-varying parameters required in a fit to data can often be controlled with a re-parametrisation. For example, the eigenvectors of the covariance matrix, if it exists from calibration fits, can be used[150]. Then, using tools the author has developed, rate estimators can be produced for a sufficient subset of parameters for inference on observed data. However, the creation of better interpolators is important, as seen linear interpolation is insufficient in some cases. `flamedisx` would be greatly improved by a standard toolkit for auto-differentiable interpolation of the total rate that is continuous in gradients and Hessians.

The rate estimators become the dominant issue for calibration fitting where all the parameters of the model must be included. The author notes that utilising the Markov Chain Monte Carlo (MCMC) software discussed in Section 2.1 with `flameNEST` to calculate the differential rates and total rate simulated at every step could out-perform the current band-fitting tools.

Initial tests by the author show that this may be slower than the existing tools, requiring $\mathcal{O}(1)$min per step to evaluate the value of the differential rates for $\mathcal{O}(10^3)$ data points on the CPU, with multiple walkers requiring at least $\mathcal{O}(10^3)$ steps. The CPUs were tested as many parallel walkers would be required. However, it would carry the benefit of full dimensionality and shape information, unlike bands or MC templates, and may require less calibration events. This process may be further improved in `flameNEST` by Taylor expanding the differential rates to higher orders on a GPU, once the Taylor series was determined it would be extremely cheap to re-evaluate on a CPU compared to the explicit function.

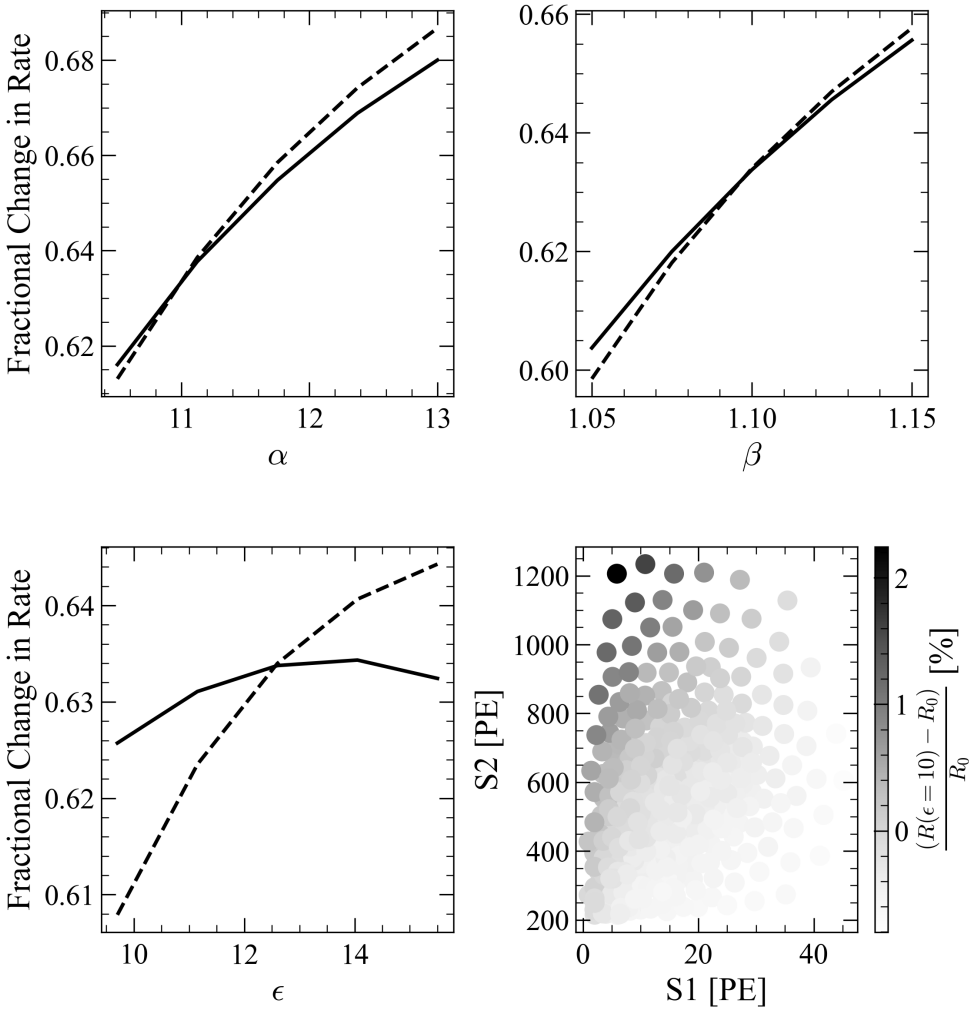


Figure 4.14: Using a modified sample k-mean clustered over many different α values. Rate estimation was then performed using a modified version of Equation 4.4, weighing calculated points by their volumes. Cubic volumes are used except for the non-linear space of S1 & S2 which used Voronoi tessellation. A sample size of 400 points was used and the function normalised to a single MC estimate at default NEST values. The approximation holds for a short range of α and β but breaks down when ϵ, which controls the electron yield and creates a different change in the shape of the model than what is captured in the sampling.

Ideally, the explicit calculation of differential rates and their derivatives could be implemented with improved numerical tools to aid in this calculation. For example, the author

explored using voronoi volumes, as shown in Figure 4.14, but this performed poorly when the parameters differed from those initially sampled. This method could be further improved with interpolation to better estimate the integral over voronoi volumes and importance sampling to construct an improved sample. The goal of such work should be to construct an explicit calculation of the total rate using the differential rates, allowing for well defined derivatives. Perhaps with current performance, a single point of $\mathcal{O}(10^6)$ events could be used to calculate this once and define a Taylor series. From Figure 4.13 one can estimate that this would take a total of $\mathcal{O}(20)$ hours for a second-order Taylor series, which was not pursued in this work and it may struggle with sinusoidal shapes such as ϵ shown in Figure 4.9.

4.6 Summary

The author resolved major memory issues in floating CB parameters, allowing fitting of NEST's scattering physics models. This will reduce the size of calibrations required for the model and help unfold detector effects from real physics, helping predictability. In doing so, the author has significantly reduced the memory required for standard fits, which allows for a speed up of the original functionality. The solutions here were necessary for NEST implementation but may extend to other models utilising `flamedisx`. Even with improved interpolation, the total rate estimation is a bottleneck for many inseparable parameters. This may be resolved by a hybridisation of the differential rate calculation with MC, feasible given improvements by the author, or by improved numerical methods, which the author explored but work is ongoing. For calibrations and not frequentist inference, `flameNEST` may be integrated with MCMC tools allowing the rate to be estimated by MC at every step. The author did some preliminary benchmarks for this, and proposed several implementations, but did not continue as the LZ collaboration was interested in applying `flameNEST` to full science searches. The latter would be a more robust test of the functionality of `flameNEST`, so the author explored this in Section 6.11.

Chapter 5

Statistical Inference for LZ in WS24

The author was pivotal to the development of a new Python-based inference tool, called `LZFlameFit`, that allowed for flexibility in the construction of a complex combined likelihood model that would not have been feasible in the `LZStats` tool. As part of the statistics team the author developed and used this tool for the combined analysis for LZ's first and second WIMP searches, WS22 and WS24. The Python-based `LZFlameFit` package was written to implement sources and likelihoods from `flamedisx`, if only using MC templates imported from existing LZ simulations tools. The clear Python object-oriented framework allowed for construction of the complex combined likelihood model for WS24, discussed in Section 5.2.9. The python implementation allowed for easy use of public packages like goodness of fit tools discussed in Section 5.1.2.

5.1 LZFlameFit: Inference Package

In Chapter 4, `flamedisx` was presented as a tool for explicit likelihood models in a tensor-based construction. In reality, it can also manage traditional templates with an array of tools for asymptotic and non-asymptotic inference. In Ref.[111], Dr Robert James developed some of these tools and created a package, `LZFlameFit`, to test and store LZ-specific analyses. As part of the statistics team[1] , the author expanded on this work making `LZFlameFit` manage combined likelihood, new inputs, and template morphing, making it the central inference tool in LZ.

5.1.1 Modified `flamedisx` Implementation

While `flamedisx` provided a package to build likelihoods and perform non-asymptotic inference, in reality a great deal of information is required to build an analysis beyond this, such as all the inputs to the likelihood, model testing, and goodness-of-fit checks. It was also found that for analyses in LZ some modifications to the inference routines and `flamedisx` objects were

[1]Dr Robert James, Dr Joe B. McLaughlin, Joshua Green (the author), Makayla Trask, Leah Wolf, Jacopo Siniscalco, and Dr Asher Kaboth

required. Therefore, the LZ statistics team developed the `LZFlameFit` package. The author in particular devised template based sources that used counts instead of rate multipliers, the combined likelihood structure, and introduced smooth template morphing discussed in Section 5.2.3. Practically, the author contributed to every aspect in terms of quality assurance, producing results for WS24, and debugging. Some examples are; the author created a method to limit memory consumption in inference routines by producing multiple files in a single task, created the fit overviews in the goodness of fit package, and implemented sensitivity into the inference helper. In addition, the author was responsible for `flamedisx` aspects, such as debugging the source code or specifics about implementation. Some details of the author's contributions are discussed here, and details of the package can be found in Appendix B.

In order to use normalised MC templates and counts as an input, the differential rate for each source, R_j in Equation 4.3 must be replaced with the PDFs,f_j, such that

$$\nu^j_{r.m} R_j(\vec{s}, \vec{\theta}) \rightarrow C_j \nu_j(\vec{\theta}) f_j(\vec{s}, \vec{\theta})$$
$$\nu^j_{r.m} \nu_j(\vec{\theta}) \rightarrow C_j \nu_j(\vec{\theta})$$

where $\nu^j_{r.m}$ is the rate multiplier, C_j is the counts, and ν_j is transformed from a function that represents the total rate to a fractional change in the counts with respect to some model parameters, $\vec{\theta}$. The $\vec{s}$ represents the observables of the model. A function connecting the change in rate in both the PDF 'shape' term and total rate term is required as, for example, this could change the proportion of a source in two separate spaces or likelihoods according to some efficiency parameter. So, for templates and counts, the `flamedisx` source objects are over-written to produce,

$$L(\mu, \vec{\theta}) = \sum_j C_j \nu_j(\vec{\theta}) - \sum_i ln(\sum_j C_j \nu_j(\vec{\theta}) f_j(\vec{s}_i, \vec{\theta})). \tag{5.1}$$

As the gradients with respect to different counts are simple functions, the use of `TensorFlow` for auto-differentiation to give explicit gradients in `flamedisx` was retained. This allowed for fast optimisation of a many parameter model without the need for excessive model evaluations.

The `flamedisx` likelihood object has the capability to handle multiple sources; however, the handling of parameters between sources and datasets is not well adjusted to specific problems. For example, in the case of WIMP counts, one wants to have a single signal parameter that scales the counts between two spaces with different exposures and signal acceptance. To solve this, the author created a simple object that inherited the `flamedisx` log-likelihood class but stored multiple individual log-likelihoods, each with its own dataset and shared or unique parameters. Combining the results and derivatives is a simple addition, with some care for the structure of the derivative tensors. Dr Robert James generalised this method to pass multiple

generic likelihoods and scan each likelihood for shared and unique parameters, making adaptation easier. This allowed for a much simpler framework to build the full WS24 multiple-space model and combine it with WS22. This package, with further developments by the relevant groups, was used in LZ for light dark matter and ^{8}B CEvNS searches in Ref.[151] and low energy ER searches in Ref.[152]. Further, Dr Joe McLaughlin and the author, have been working on adding binned likelihoods for high energy ER searches.

5.1.2 Goodness of Fit

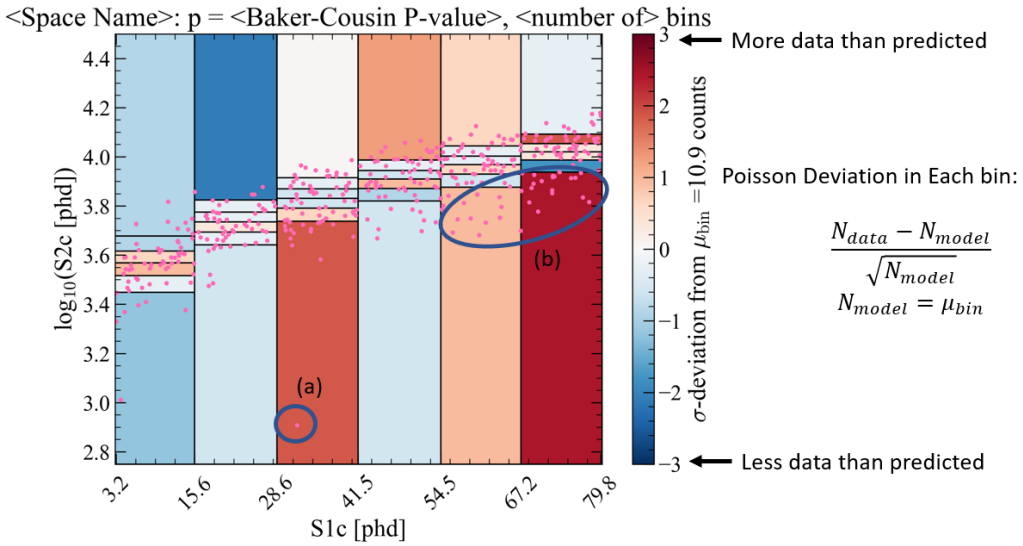


Figure 5.1: Example of the 2D GOF plots shown in this chapter. This space had an acceptable p-value but the position of deviations can inform us. Without knowing the model we can determine that the event at (a) is difficult to cover and the region (b) is showing some over-fluctuations under the ER band.

In the literature, the goodness of fit (GOF) refers to the p-value of the background-only hypothesis. As calculating this would be equivalent to calculating the discovery significance in our dark-matter searches, we use a different test to check that the shape of the background model is sufficiently well modelled. Although the author was involved in discussions and cross-checks, these tests were predominantly developed by Leah Wolf using the XENON collaboration's `EvaluateGOF` package from Ref.[153]. A binned Poisson likelihood ratio is used as recommended in Ref.[154] given by using either pre-determined bins for 1D checks or `EvaluateGOF`'s equiprobable bins with fixed counts for 2D,

$$\chi^2_m = -2\ln(\frac{L(\vec{y}|\vec{n})}{L(\vec{n}|\vec{n})}) = 2\sum_i y_i - n_i + n_i ln(\frac{n_i}{y_i}),$$

where $\vec{n}$ and $\vec{y}$ are vectors of observed and predicted bin-counts, respectively, and i indicates the sum over m bins. This is a basic version of the profile likelihood ratio as discussed in Chapter 3, where $\vec{n}$ is the maximum likelihood parameter of the model, i.e the expectation for each bin

is its observed value. This is sometimes called the "Baker-Cousins χ^2" as, being a likelihood ratio, it asymptotically follows the χ^2 distribution with m-degrees of freedom. Technically the degrees of freedom should be modified by free-parameters and constraints, but non-asymptotic methods were used to calculate the p-value by throwing toys Poisson distributed around a histogram representing the median realization of the model given some parameters. As we wish to test the GOF from the MLE of our unbinned likelihood using observed data, this expectation histogram is derived from the median of simulating data from the MLE estimators of the model parameters and their covariance matrix. An example of the 2D GOF test, with indications of where each value is, is provided in Figure 5.1.

5.1.3 Fit Parameter Summary

In addition to the GOF it is useful to inspect the resultant parameter values, their uncertainties, and their correlations. Although a fit table and covariance matrix is sufficient, it helps when interpreting results to use a plot with the pulls and constraints, as discussed in the high energy physics manifesto on reporting results in Ref.[155]. Here, the pull gives the difference between pre- and post-fit parameters normalised by their pre- or post-fit uncertainty. The pre-fit uncertainty tells us how much the observation deviates from the expectation. The choice of post-fit can be interpreted as a residual plot. It is useful in development, or where an uncertainty estimate does not exist, which is not the case here. The constraint is the ratio of post-fit and pre-fit uncertainty, where defined, and gives an indication of how much the observed data is improving the constraint on a parameter. Finally, the correlation matrix gives the correlation coefficient or how much the observed data and the model implies two variables depend on each other. It may also be intuitive to consider this as the change required to restore the maximum likelihood should one value change i.e, if the Solar Neutrino counts increase, one must decrease the ^{214}Pb counts to conserve the number of events in the ER band.

Another recommendation in Ref.[155] is the impact of each parameter, which is how much a perturbation of each parameter by a chosen amount changes the best fit result. The author introduced these in their early work but they typically require non-zero signal events so are not applicable to a discovery experiment. In practice some high-energy physics experiments use an Asimov data-set, that is the average observed data set, but this is not well defined for an un-binned likelihood model.

5.2 WIMP Search 2024 with MC Templates

Due to the novelty of `flameNEST`, it was decided that for the main WIMP search of 2024 (WS24) only template-based functionality of LZFlameFit would be used, limiting the dimensions and number of parameters available in the model. The author contributed to establishing this 2D (S1c,S2c) template-based likelihood, implemented template morphing for shape varying

parameters, helped construct the final likelihood, performed various checks, and calculated the final results. To avoid analyser bias, fake WIMP events were added to the WS24 data set in a process called salting. This takes random S1 and S2 signals from calibration data and merges them according to NEST predictions for a WIMP of specific mass and a random number of counts. The parameters for the salt model would remain a secret from the analysers, but it was known that a flat pedestal was added to the energy spectrum for higher-energy searches.

In this section, we will cover the components of the combined WS24 and WS22 analysis, how they were combined, and the specifics of the model. The author worked in a close collaboration with the statistics team, some plots in this section were taken from the technical note, and whilst the author was involved in their creation, they were not generated by him.

5.2.1 Combined Likelihood Spaces of WS24

There are five distinct spaces in the WS24 likelihood: the neutron veto sideband, mixed state, isolated state radon tag inactivate, radon tagged, and radon untagged. Mixed and isolated refer to the high and low mixing states of the xenon in the TPC, the latter allowing radon tagging to take place, though it was not always active. In addition, the first science run, WS22, was added to the likelihood to maximise the physics reach. In the following sections, the data and models relevant for the discussed space are presented at the beginning.

WS22: the First Science Run

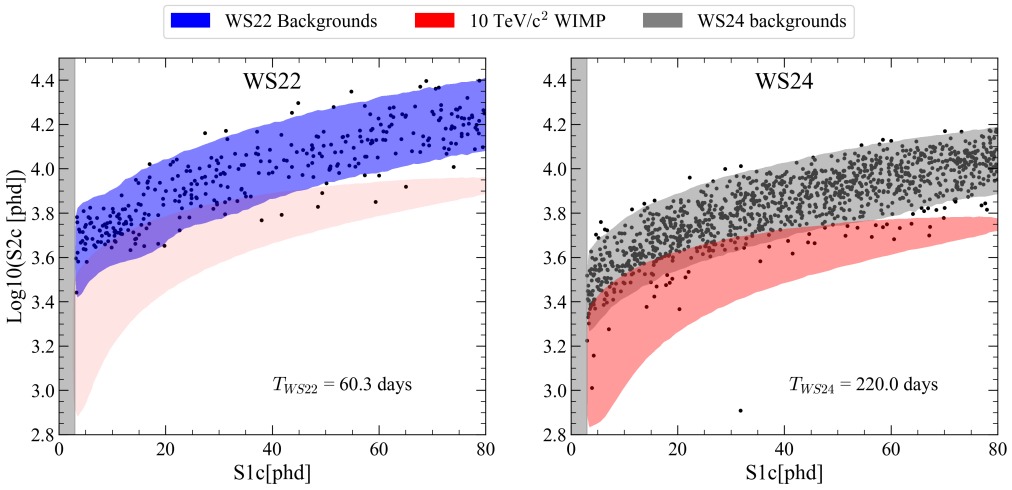


Figure 5.2: The WS22 and WS24 ROIs and 95% contours of various sources. The intensity of the contour diagrammatically indicates the rate of a source in that space, here the WIMP rate is much greater in the WS24 space due to the larger exposure. The beta bands are in separate colors as these backgrounds were uncorrelated.

While WS22 was not salted, it was decided to add this to our final search as an important constraint on the WIMP signal, but to avoid bias in the current analysis the model was frozen

to its condition in ref [95]. Note that due to the cosmogenic activation of the xenon on the surface, WS22 had significant contributions from ^{127}Xe and ^{37}Ar, the latter of which would have decayed to a negligible rate by the time WS24 data taking began. As it is difficult to predict the impact of changing conditions, it was assumed that common background rates were not correlated with their WS24 counterparts. Unlike WS24, the constraint on neutrons from the neutron-veto in WS22 was modelled by an approximate likelihood derived from scanning over fixed neutron counts in likelihood fits to the veto side-band.

5.2.1.1 Veto

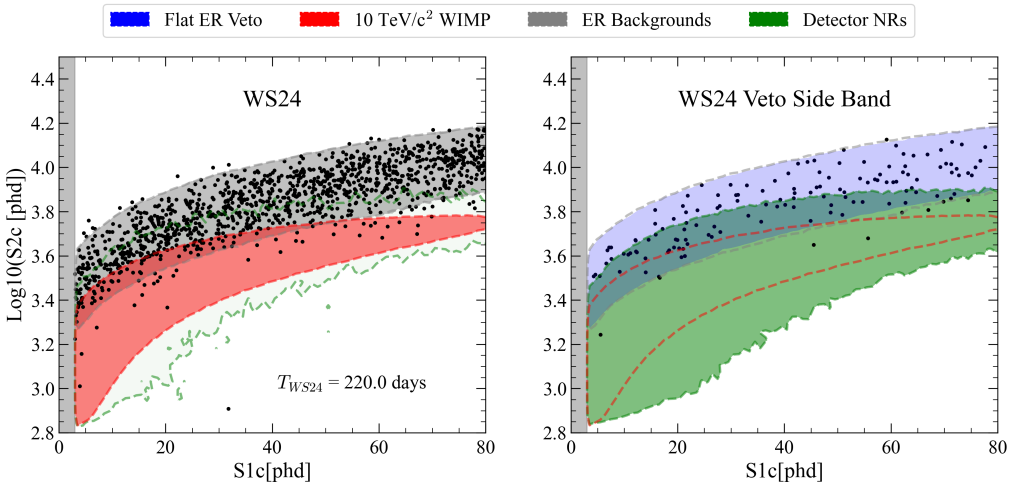


Figure 5.3: The WS24 ROI and vetoed spaces with 95% contours of various sources. The intensity of the contours indicate the ratios of rate between the two spaces. For an efficiency 90% on average 1 neutron in the ROI is accompanied by 10 in the veto, and for a dead time of 3% every 100 WIMPs 3 are observed in the veto. Only the edges of the WIMP contour are added to the veto space to show its negligible contribution.

Recalling chapter 2.1.3, the LZ experiment has skin and OD detectors to veto gamma rays and neutrons. The energy thresholds and time window of these vetoes are tuned to minimise dead time, λ_{veto}, which is the fraction of signals in the TPC that would be randomly vetoed, and maximise efficiency, ϵ_{veto}, which is the fraction of neutrons correctly vetoed. The veto is implemented so that it is possible to observe WIMPs in the veto side band via the dead time, the dead time is so small, and enhancement of neutrons through efficiency so large, that it is virtually impossible to fit any WIMP signal in the veto.

The veto efficiency here is in terms of neutrons, for gamma-rays the efficiency is either fixed or the veto and ROI sources are uncorrelated. For example, the enhanced rate between the veto and ROI of the detector ERs due to tagging of the gamma rays was assigned to an uncorrelated ER source with a flat energy spectrum in the veto. This is only viable because the detector ERs are dominated by ^{214}Pb in the ROI. Although vetoed detector ERs are the majority of

the rate of the flat ER veto source, there is a small contribution from non-naked ^{214}Pb, where the ^{214}Bi is left in an excited state after the beta decay and relaxes to emit a gamma ray. Due to neutron calibrations during WS24, ^{127}Xe and ^{125}Xe were introduced by activation. They always occur with a coincident gamma-ray from the relaxation of the daughter iodine nucleus. This yields negligible rates in the ROI but a significant contribution to the veto.

5.2.1.2 Mixed and Isolated Flow States

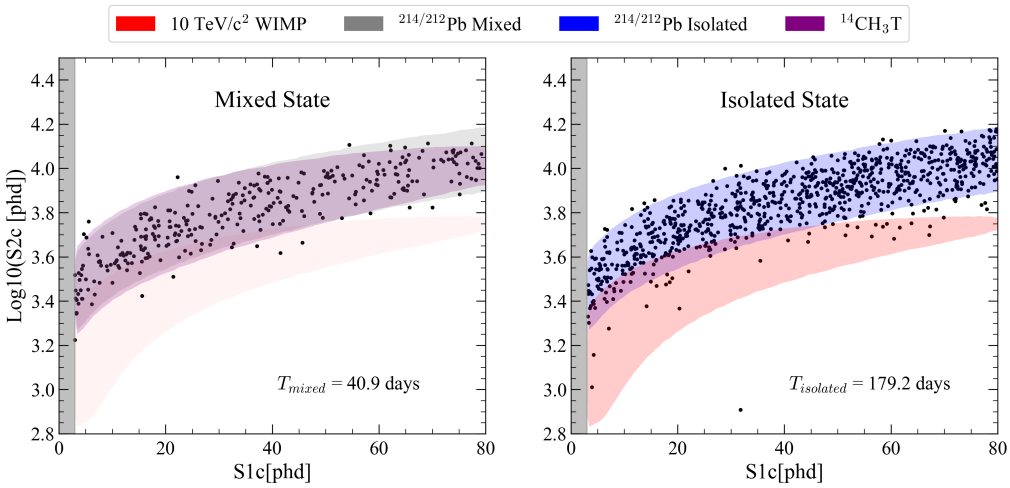


Figure 5.4: Mixed and Isolated data sets showing the unique $^{14}CH_3T$ source and that the mixed and isolated radon daughters,^{214}Pb/^{212}Pb, are treated as distinct and uncorrelated sources.

As discussed in Section 2.1.2, the flow in LZ can be manipulated to reduce the mixing of the centre of the liquid xenon target with the volume boundaries, which emanate radon mainly from PTFE as discussed in Section 2.3.4, This results in a reduced activity of ^{214}Pb/^{212}Pb in the TPC as the radon flows to the top where it decays outside the FV. However, turbulent flow is still present, and each decay in the radon chain may produce a positively charged daughter, which would drift along the field back into the TPC. For ^{222}Rn $\rightarrow$ ^{218}Po $\rightarrow$ ^{214}Pb the half-lives of the progenitors give sufficient time to drift back into the FV, leading to a consistent rate between the mixed and isolated state. For ^{220}Rn $\rightarrow$ ^{216}Po $\rightarrow$ ^{212}Pb, the half-lives of the progenitors are too short, so very few make it into the FV and the isolated state has a lower rate. Whilst WS24 is mostly in the isolated state, there was a period of the mixed state added to the search whereupon calibrations took place. These two states were treated as separate spaces.

The mixed state space contains an additional background of tritiated methane that was also activated to contain carbon-14 injected during calibrations. As they are from the same source, the sources are combined into a single histogram weighted by their predicted rates and are referred to as ^{14}CH$_3$T. Due to the removal time of methane in the getter, these sources have a half-life of $\approx$3 days in the liquid xenon before being removed, so none remain in the isolated

state. Furthermore, neutron calibrations during the calibration period introduced ^{125}Xe and ^{127}Xe as a background. The neutron calibrations also occurred during the isolated state, so these backgrounds are in all periods.

5.2.1.3 Radon tagging

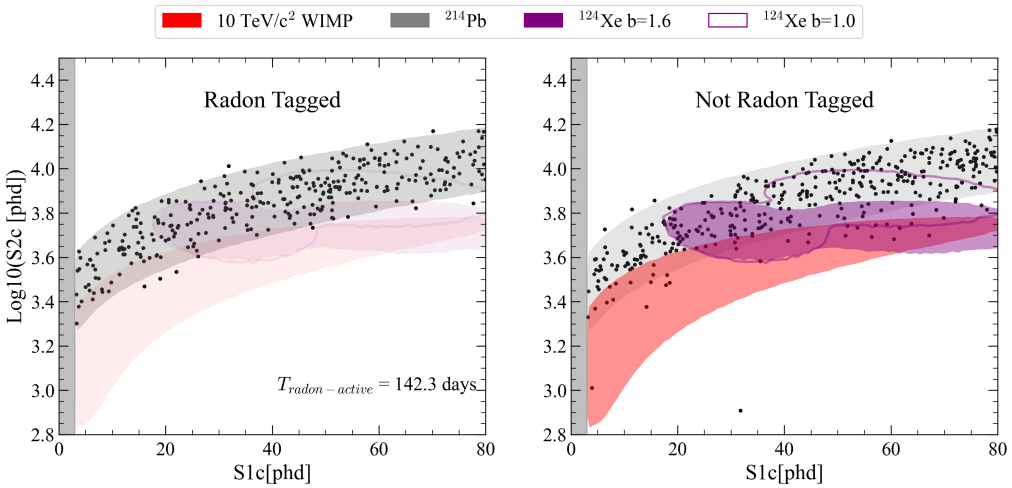


Figure 5.5: Diagram of the radon tagged and untagged space, as presented in Table 5.2 15% of non-^{214}Pb sources are in the tagged space and approximately 70% of ^{214}Pb. This makes the ^{124}Xe DEC clearer in the untagged space, which is discussed in Section 5.2.3 and Section 5.2.5.

The isolated state reduced the ^{212}Pb rate, but unfortunately the dominant ^{214}Pb remained consistent between the flow states. Fortunately, as the flow is easier to model in the isolated state, one can attempt to tag a ^{214}Pb by pairing the distinct alpha decay of its progenitor ^{218}Po and ^{222}Rn. This was possible for most of the isolated state period, but was not calculated for a sub-sample called the inactive period. This tag operates by identifying the RnPo pair of the progenitors and then tagging any decay that occurs in a time-evolving volume determined by a model of the flow and drift field as more likely to be a ^{214}Pb decay. The tag has a tagging efficiency, ϵ_{Rn}, defined as the ratio of tagged ^{214}Pb to the total decays and is determined by decays of ^{214}Pb at energies outside of the ROI. It also has a dead time, λ_{Rn}, determined by calculating the ratio of the tagged exposure to the total in the active period. The final value of λ_{Rn} was cross-checked against uniform ratios of $^{131m/129m}$Xe decays outside the ROI in the tagged and untagged datasets. Even if ^{212}Pb were not subdominant, this tag would not be appropriate as, despite the longer half-life of its progenitors, its half-life is ≈24x greater than ^{214}Pb. This would cause the exclusion period to be extended much further in time, greatly decreasing efficiency and increasing dead time. Like the veto, the tag is implemented as a separate space with a ^{214}Pb rate enhanced by the efficiency and rate of all other sources adjusted by the dead time.

5.2.1.4 The combined likelihood

Parameter	**Mixed**	**Inactive**	**Rn tagged**	**Rn untagged**	**Veto tagged**
^{214}Pb Mixed	1	—	—	—	$\frac{\lambda_v}{1-\lambda_v}\frac{C^{veto}}{C^{ROI}}$
Rn Other Mixed	1	—	—	—	$\frac{\lambda_v}{1-\lambda_v}\frac{C^{veto}}{C^{ROI}}$
^{214}Pb Isolated	—	$\frac{C^{\rm iso}_{\rm ina}}{C^{\rm iso}}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm iso}}\epsilon_{Rn}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm iso}}(1-\epsilon_{Rn})$	$\frac{\lambda_v}{1-\lambda_v}\frac{C^{veto}}{C^{ROI}}$
Rn Other Isolated	—	$\frac{C^{\rm iso}_{\rm ina}}{C^{\rm iso}}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm iso}}\epsilon_{Rn}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm iso}}(1-\epsilon_{Rn})$	$\frac{\lambda_v}{1-\lambda_v}\frac{C^{veto}}{C^{ROI}}$
CH_3C	1	—	—	—	$\frac{\lambda_v}{1-\lambda_v}$
^{14}C	1	—	—	—	$\frac{\lambda_v}{1-\lambda_v}$
All others	$\frac{C^{\rm mix}}{C^{\rm mix}+C^{\rm iso}}$	$\frac{C^{\rm iso}_{\rm ina}}{C^{\rm mix}+C^{\rm iso}}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm mix}+C^{\rm iso}}\lambda_{Rn}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm mix}+C^{\rm iso}}(1-\lambda_{Rn})$	$\frac{\lambda_v}{1-\lambda_v}\frac{C^{veto}}{C^{ROI}}$
Detector NR	$\frac{C^{\rm mix}}{C^{\rm mix}+C^{\rm iso}}$	$\frac{C^{\rm iso}_{\rm ina}}{C^{\rm mix}+C^{\rm iso}}$	$\frac{C^{\rm mix}}{C^{\rm mix}+C^{\rm iso}}\lambda_{Rn}$	$\frac{C^{\rm iso}_{\rm a}}{C^{\rm mix}+C^{\rm iso}}(1-\lambda_{Rn})$	$\frac{\epsilon_v}{1-\epsilon_v}\frac{C^{veto}}{C^{ROI}}$
Flat Veto ER	–	–	–	–	1

Table 5.1: Table showing scaling factors for rates between each SR3 sample. Here, C^{mix} and C^{iso} represent the predicted counts in the mixed and isolated state, respectively. C^{iso}_a and C^{iso}_{ina} represent the periods of time where the radon tag was active and inactive, such that $C^{iso} = C^{iso}_a + C^{iso}_{ina}$. For the veto a new parameter $C^{veto/ROI}$ represents the predicted counts of that source in the veto/ROI and spans total exposure so that $C^{ROI} = C^{iso} + C^{mix}$.

All spaces are included in a combined likelihood as described in Chapter 3. Each source of WS24 has a single parameter in all 5 spaces, the total counts in the WS ROI, given in Table 5.2. Exceptions to this are the counts of $^{214/212}$Pb,^{14}C, and ^{3}H between mixed and isolated states and the additional Flat ERs in the veto. Note that in some cases, the acceptance of a source may vary between each space and so using the predicted counts for each space, C, accounts for the acceptance as well as the live time of each space. So, for a source that is in all spaces, the scaling of the total counts parameter in the mixed state space would be

$$\frac{\epsilon^{mix}_s T^{mix}}{\epsilon^{mix}_s T^{mix} + \epsilon^{iso}_s T^{iso}} = \frac{C^{mix}}{C^{mix} + C^{iso}},$$

where $T^{mix/iso}$ represent exposures and $\epsilon^{mix/iso}_s$ a simulation based efficiency of the source in that space, nominally identical. The rate before efficiencies of the source cancel in the ratio.

As the full likelihood is too long to write here, it is useful to explore two examples of how

to use Table 5.1. First, the expression for the combined likelihood of WS22 and WS24 is studied, then the combination of the WS24 Veto and ROI. The only source to be shared between them was the WIMP signal source, which would have identical scaling to 'Other' in Table 5.1, except for an additional C^{WS22} in the denominators, meaning that the signal parameter is the total WIMP counts across WS22 and WS24. The expression would be

$$\begin{aligned} l(\mu, \vec{\theta}^{WS24}, \vec{\theta}^{WS22}) =& l(\frac{C^{WS24}}{C^{WS22}+C^{WS24}}\mu, \vec{\theta}^{WS24}|\{\vec{x}\}^{WS24}) \\ &+ l(\frac{C^{WS22}}{C^{WS22}+C^{WS24}}\mu, \vec{\theta}^{WS24}|\{\vec{x}\}^{WS24}) \end{aligned}$$

where $C^{WS24} = C^{\text{mix}} + C^{\text{iso}}$. Another example is the veto and WS ROI in WS24, which is a bit more complex as they have the same exposure but sources may have different efficiencies in the veto or in the WS ROI. For example, one may imagine that the Detector ER may have a different simulation derived efficiency for the FV cut in the veto and WS ROI as the veto would preferentially select events outside the FV. The form can be written as

$$\begin{aligned} l(\vec{\theta}, \theta^{veto}_{Flat.ER}) =& l\left([\vec{\theta}, \theta_{DetNR}]|\{\vec{x}\}^{WS}\right) \\ &+ l\left(\frac{C^{veto}}{C^{WS}}\left[\frac{\lambda_{veto}}{1-\lambda_{veto}}\vec{\theta}, \frac{\epsilon_{veto}}{1-\epsilon_{veto}}\theta_{DetNR}\right], \theta^{veto}_{Flat.ER}|\{\vec{x}\}^{veto}\right) \end{aligned}$$

where the ratio of counts is not representing a difference in live time, as the veto is a sideband for the total exposure, but difference in simulation derived efficiency of the sources. Note that it is treated as an operator here for simplicity. Every θ is the counts in the WS24 ROI, except the unique $\theta^{veto}_{Flat.ER}$ that models additional events from sources in the veto assumed to be uncorrelated with the WS, as discussed before. An important aspect of the veto space, not listed in Table 5.1 and Table 5.2, is the ^{127}Xe+^{125}Xe have a unique veto dead-time of 78% due to their relaxation gamma being tagged. This was derived from simulations and fixed in the fit.

5.2.2 Modelling Signal and Backgrounds

All backgrounds except Accidental Coincidences (ACs) were modelled using the `LZLAMA` simulation package, either with custom generators or `BACCARAT` outputs, and processed through the `BGSkimmer ALPACA` module, as described in Section 2.2. The models used for WS22 were given in Ref. [95] and Ref. [123], imported to `LZFlameFit` by Dr Robert James in Ref.[111], and fixed in this analysis. The details of the updated models used are discussed when the `flameNEST` model is updated to match them in Section 6.1. The Accidental Coincidence (AC) model was produced in the data-driven method described in Section 2.3 on updated WS24 data.

Recall that the entire active volume is simulated in `LZLAMA`, or `BACCARAT` and `LZLAMA`, assuming a fixed rate that can be rescaled later by a calibrated rate. This only produced limited

RQs, used here are (S1(c), S2(c), x_{obs}, y_{obs}, drift time). The data has cuts applied to it in other RQs so efficiencies are derived as a function of the simulated ones in order to model their effect. Cuts in the simulated RQs and efficiencies are applied by the `BGSkimmer ALPACA` module, such that the weighted sum of events produces the expected rate in events per unit time. The cuts applied impact the expected rate, namely the FV cut. It is important that the cut is tuned such that the simulated and measured FV matches, in WS24 this introduced a 3% uncertainty [156]. The differential rates are calculated by normalising 2D histograms, or templates, of S1c and $\log_{10}$S2c with a 3 to 80 with 77 bins and 2 to 5 with 80 bins, respectively. These template bins are then normalised by their volumes S1c - $\log_{10}$S2c to produce a differential rate function that is interpolated to calculate the differential rates for events in data set by `flamedisx`'s TemplateSource. These differential rate functions are normalised to unity, as the counts from simulations are not used directly in the inference to reduce human error. The exception to this is the expected WIMP counts from an assumed cross-section. These are stored and used to derive the cross-section from the upper limit on observed counts. In the case of Detector ER and NR, BACCARAT is used to generate deposits in the skin and OD volumes, which `LZLAMA` can model the response of. For Detector NRs, there were sufficient surviving simulated events to produce a smooth template. For Detector ERs, the simulations could be used to derive an event rate but not a smooth template so it is replaced with a Flat ER beta source.

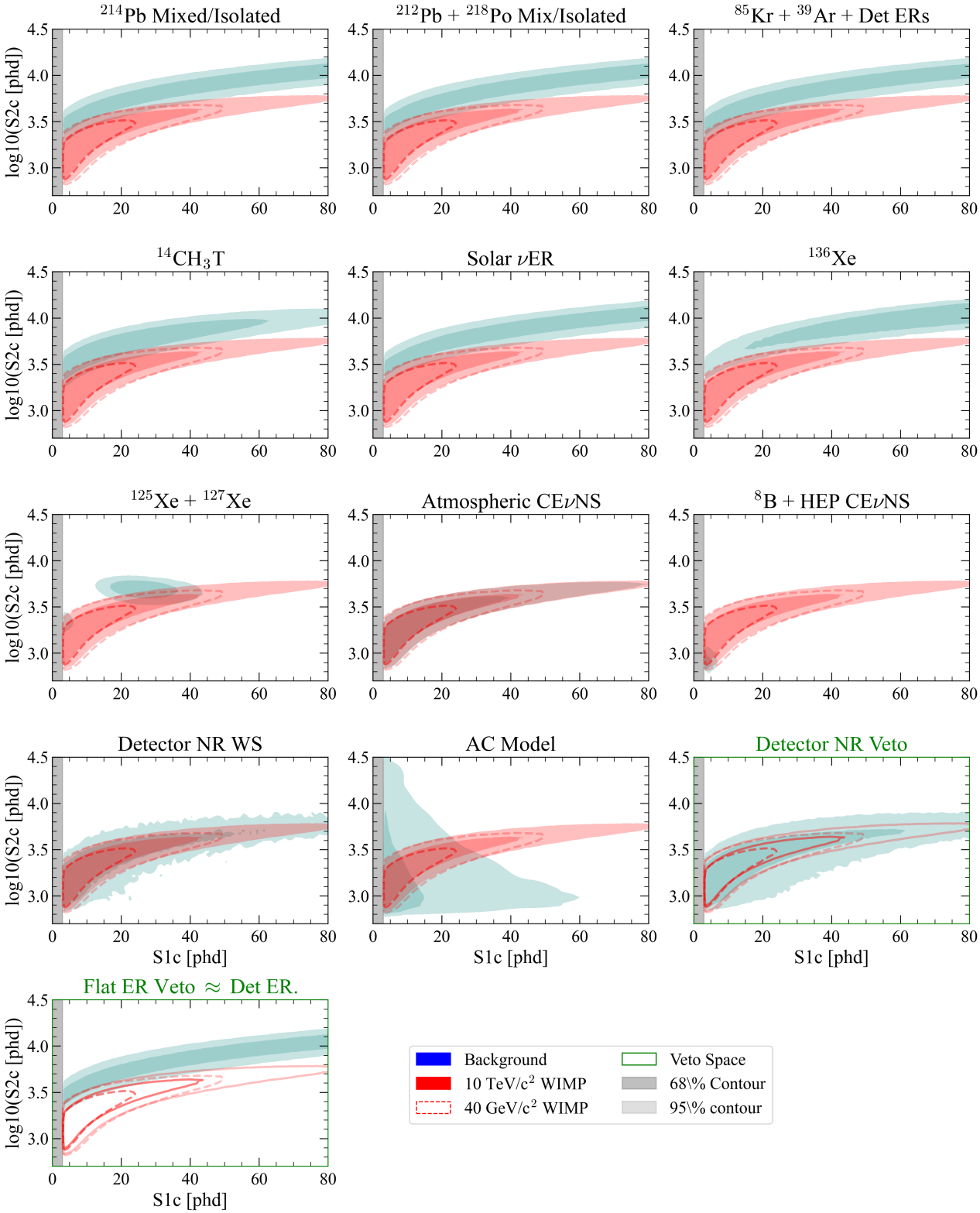


Figure 5.6: Plots of the various backgrounds with the WIMP signal. Atmospheric neutrinos have a negligible rate, see table 5.2. Both WIMP signals are not filled in the veto to indicate that they have a negligible contribution.

The final input parameters for WS24 are in Table 5.2. The expected signal counts are still derived from the `BGSkimmer` template normalisation, in evts/day, and live-days but are handed

off from the backgrounds team. In Figure 5.6, the overlap of each background with signals from two WIMP masses is shown. The shape varying source ^{124}Xe is shown for different assumptions of the box parameter in Figure 5.5. It can be seen that the detector NR source has low statistics in the WS because this is generated directly from `BACCARAT` deposits using `LZLAMA`.

In Table 5.2 there is an indication of how the values for each background were derived by the background's team, and therefore their primary source of uncertainty. Theoretical indicates they are derived from either a known flux and cross-section, or abundance of and half-life of an isotope. Calibrated means this quantity is directly calibrated to measure, e.g, neutron sources were used to measure the veto efficiency. Extrapolated means that we know the source of the background i.e, activation in the case of unstable Xenon isotopes, which can be extrapolated from their introduction to the science run. Measured indicates that the quantity is a direct measurement; for example, the radon tagging deadtime is simply how much of the exposure was tagged with no uncertainty. Finally, sideband means this quantity can be measured during the science run from a region of the data separated from the ROI.

Parameter (Symbol)	Value	Uncertainty (Rel.)	Units	Floated	Derivation
Parameters exclusively in mixed state					
^{214}Pb β counts	136.8	13.0%	evts	Y	Sideband
^{212}Pb β counts	19.8	24.0%	evts	Y†	–"–
^{218}Po β counts	2.4	11.7%	evts	Y†	–"–
^{14}C β counts	1.30	5.80%	evts	Y*	Extrapolated
CH_3T β counts	57.0	5.80%	evts	Y*	–"–
Livedays	40.87	—	days	N	Measured
Parameters exclusively in isolated state					
^{214}Pb β counts	606.1	13.0%	evts	Y	Sideband
^{212}Pb β counts	31.5	18.0%	evts	Y†	–"–
^{218}Po β counts	9.0	10.8%	evts	Y†	–"–
Radon tagging efficiency	60%	4% (abs.)	pct.	Y	Sideband
Radon tagging deadtime	15%	—	pct.	N	Measured
Livedays	179.16	—	days	N	Measured
Radon tag inactive livedays	36.85	—	days	N	–"–
Radon tag active livedays	142.31	—	days	N	–"–
Parameters exclusively in veto sideband					
Flat ER Veto	139.7	30.0%	evts	Y	Mix
Veto efficiency	92.2%	4.3% (abs.)	pct.	Y	Calibrated
Veto tagging deadtime	2.87%	—	pct.	N	Measured
Shared parameters					
^{85}Kr β counts	153.5	14.6%	evts	Y‡	Theoretical
^{39}Ar β counts	2.8	16.5%	evts	Y‡	–"–
Detector ER counts	6.0	30.0%	evts	Y‡	–"–
Solar ER counts	102.3	6.0%	evts	Y	–"–
^{136}Xe $\beta\beta$ counts	55.6	15%	evts	Y	Measured
^{127}Xe + ^{125}Xe ECs counts	3.2	21.2%	evts	Y	Extrapolated
^{124}Xe counts	19.4	20.0%	evts	Y	Sideband
^{8}B CEνNS counts	4.0×10^{-2}	4.0%	evts	Y**	Theoretical
HEP CEνNS counts	2.0×10^{-2}	30%	evts	Y**	–"–
Atmospheric CEνNS counts	0.1	20.0%	evts	Y	–"–
AC counts	2.75	23.1%	evts	Y	Sideband
^{124}Xe LL box parameter	1.0	[1,1.6]	—	flat prior	Theoretical

Table 5.2: Summary of parameters entering WS24 likelihood components. If a parameter is floated that means it is a parameter of the likelihood, it may be constrained. The uncertainties are incorporated as a Gaussian constraint term.
† Combined in fixed proportion to expected counts as *Rn Daughter Beta* parameters.
* Combined in fixed proportion to expected counts as a CH_3T + ^{14}C parameter.
‡ Combined in fixed proportion to expected counts as a ^{85}Kr + ^{39}Ar + *Gammas* parameter.
** Combined in fixed proportion to expected counts as a ^{8}B + *HEP* parameter.

5.2.3 The ^{124}Xe Model and Template Morphing

In the radon untagged space of Figure 5.5, anomalous leakage at high S1c can be observed. As it is more prominent in the radon not tagged than the tagged space, it is likely not from ^{214}Pb events and therefore not a general behaviour of beta-decay ERs. Further, there were no spatial or temporal deviations that could not be precisely corrected for in this science run, further details can be found in the calibrations technical note Ref.[157]. Therefore, it was hypothesised that the events were from unique behaviour of non beta-decays, particularly those for which the charge yield could be suppressed. Although this is true for EC backgrounds, deep scatters of solar neutrinos or double beta-decays, the most prominent candidate was ^{124}Xe double electron capture (DEC) whose LL peak lies near the leakage. The author worked on determining the leakage, testing some of the initial candidates, and implementing template morphing for the new ^{124}Xe model, but not on its derivation. The author later implemented the model in `flameNEST`, but it was first discovered in the template analysis, so we will describe it here.

Recall that the ^{127}Xe and ^{125}Xe backgrounds discussed in section 2.3.1 were cases of Electron Capture (EC) where an electron is absorbed by a proton in the nucleus leading to an Auger cascade of electrons filling the empty orbitals and subsequent X-ray emission. The longest lived experimentally observed isotope, ^{124}Xe, consists of a DEC directly to the daughter's ground state[158]. The half-life of order 10^{22} years makes this decay exceptionally rare, predicting just 19 events in the 3.3 tyr exposure of WS24, and without a prompt gamma ray like for ^{127}Xe it is impossible to tag. Like the ECs, the DECs are classified by the shell of the captured electron, and relevant in the ROI are the LM and LL DECs. As the M-shell energy is only 1.1 keV, and the L-shell 5.5 keV, it can be approximated that the LM DEC has an identical charge yield to the L-shell but at a slightly different energy as the M-shell has a negligible impact. The latter L-shell charge yield is well modelled by the γ-β weighted model in NEST, as it can be studied in ^{127}Xe by tagging the prompt gammas. However, for the LL decay one can expect a far greater increase in the density of the deposited energy such that the charge yield cannot be well modelled by the L-shell in NEST. The higher density results in greater recombination, which would suppress the charge yield and enhance the light yield, causing the LL peak to move down in S1c-S2c space and into the NR band as seen in Figure 5.5. The DEC charge yield model discussed here, and calibrations for ECs in LZ, can be found in Ref.[124].

Due to the high density of DEC, the Thomas-Imel-Box (TIB) model is used. Recall from Section 2.2.1, that the recombination parameter depends on a single model parameter ζ which is proportional to the density of primordial ions allowing one to introduce a box parameter, b, which scales it to give the charge yield of,

$$Q_y^{LL} = \frac{ln(1 - b\zeta^L)}{b\zeta^L W(1 + n^{ex}/n^i)}. \tag{5.2}$$

Here all the symbols are identical to Section 2.2.1 and ζ^L is the measured model parameter for the L-peak. The ratio of TIBs for with and without the box parameter is used to modify the NEST weighted model for ECs,

$$Q_y = Q_y^{\beta-\gamma}\frac{Q_y^{LL}}{Q_y^L} = Q_y^{\beta-\gamma}\frac{ln(1-b\zeta^L)}{bln(1-\zeta^L)}. \tag{5.3}$$

To implement this, `LZLAMA` simulations were produced at anchor points of b ranging from 1-1.6 in steps of 0.1, which would be sufficient since the uncertainty on the parameter was expected to be large with such few events. Note that 1.6 is the maximum parameter since $\zeta \propto \rho^{2/3}$ means that doubling the density gives a b of 1.6. Figure 5.5 shows how the parameter causes the LL peak to vary.

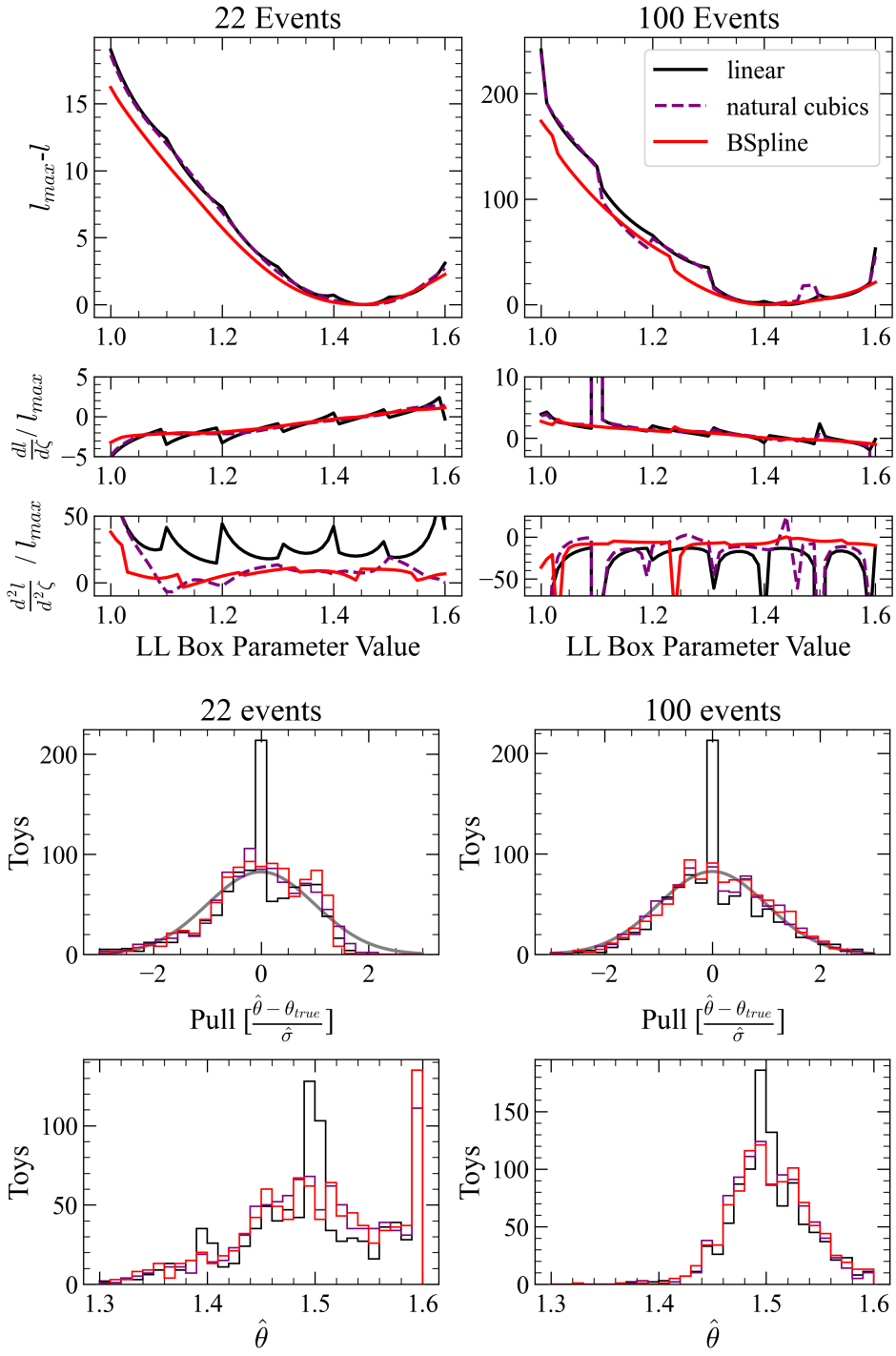


Figure 5.7: Log likelihood value, its gradients, and second order gradients with respect to the box parameter for different interpolation types on simulated data for 22 and 100 events (top). Pull plots sampling toys from a box parameter of 1.5 and the same number of events (bottom). The splines remove a bias of the optimiser to remain near anchor points, caused by a discontinuous likelihood, and retain nominal behaviour of pulls.

In order to carry out template-morphing, the author originally implemented a linear interpolation between templates. Unfortunately, this showed several anomalies as the stepwise behaviour, see Figure 5.7, would lead to unrealistic fits in the optimiser and ignoring them

would cause a significant slow-down for the many-parameter likelihood. Therefore, the author implemented a B-Splines interpolation adapted from the `TensorFlow Graphics` package in [159]. These give continuity up to second gradients and provide local control of the function . The latter meant that a single bin fluctuation in the simulations would not impact the entire function for an event but that B-Splines can cause errors as the function does not converge on anchor points. This was not significant for WS24 but for future applications the author created two more options; extrapolating more points to force B-Splines closer to the anchor points and natural cubic splines, which force the function through the values at anchor points. This would allow the end user to choose the appropriate tool for the situation. Both natural cubics and B-Splines perform well for the WS24 case of $\mathcal{O}(10)$ events, but one may need to reconsider with more events or a situation where small biases could cause significant changes in inference. It can also be seen that for a lower number of events, there is a bias in the pull plot near the boundary due to the physical edge of the parameter.

5.2.4 WS24 Template Likelihood Pull Tests

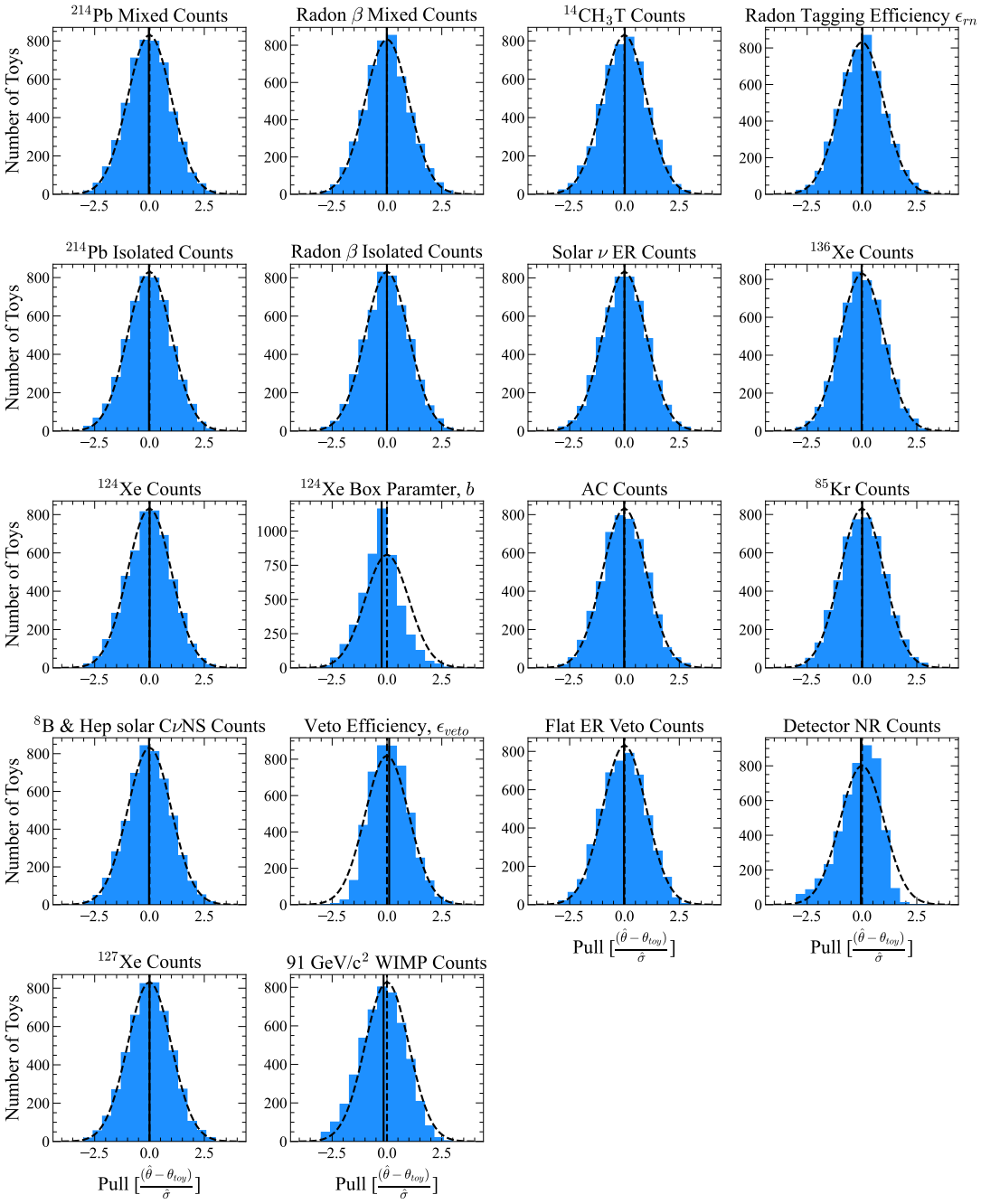


Figure 5.8: Pull test of the WS24 model, all parameters were sampled from the pre-fit Gaussian constraint terms except from the ^{124}Xe box parameter, WIMP signal, and Detector NRs. The box parameter was sampled uniformly within its bounds and Detector NRs from [0,5]. The WIMP was assigned 20 events, and all sampled counts are subject to Poisson fluctuations inside the likelihood model. These three all show mild negative biases in their pulls, due to boundary effects. Both the box parameter and Detector NRs pulls are negatively correlated with the WIMPs giving correlation coefficients of -0.08 and -0.19, respectively, imposing a small bias. No other parameters show a bias.

The author performed a pull test to verify that the fit was performing as expected using toy data produced using nuisance parameters sampled from a choice of priors that represent likely values, as seen in Figure 5.8. The three parameters that present small biases are the ^{124}Xe box parameter, the WIMP counts, and the detector NR counts. However, the positive bias is essentially weakened as the ER band can absorb any variation, so the pulls have a negative bias. For detector NRs, the sampled values can be near the boundary at 0 counts and impose a slight positive bias. The WIMP signal was sampled at values great enough to avoid the boundary at 0, but a strong anti-correlation with detector NRs produces a slight downward bias. This is all expected behaviour, thus the model was deemed to be acceptable. The box parameter could be seeing a similar bias but with the lower boundary being softened by the ER band. However, it could also be an artefact of the BSpline interpolation bias seen in Figure 5.7.

5.2.5 Salted WS24 Results

The advantage of salting is that it gives the ability to check the background model in the ROI, which is only marginally impacted by the small number of signal-like events, without biasing the analyser. In part, this gave motivation for the ^{124}Xe DEC model. This section discusses a subset of the goodness-of-fit tests prior to un-salting which the author performed and motivated the decision to un-salt.

The first step to investigating the salted data is to check the best fit values for the parameters of the model; as it was known that the salt can use any WIMP mass, and has a flat pedestal, a 10 Tev/c^2 WIMP signal was used to cover the entire possible spectrum. In Table 5.3 it can be seen that there is a significant number of best fit WIMPs in the total WS24 exposure and that the box-parameter is greater than 1., as expected. In Figure 5.9, the two strongest pulls are ^{214}Pb in the mixed state and Flat ER Veto from Detector ERs and other veto-tagged gamma components and both of these have a weak impact on the WIMP sensitivity. The deviation of the Flat ER Veto may be due to an over-estimate of the gamma tagging efficiency of the veto, but inclusion of this in the model is unlikely to yield any improvement of sensitivity due to the few events predicted in the ROI being dominated by other Betas and the very weak impact on signal counts through the Detector NRs. The data significantly constrained the dominant sources of a space, i.e ^{214}Pb isolated/mixed and the Flat ER Veto, along with the radon tagging efficiency, which is correlated with the former.

Parameter	Predicted	Best fit
^{212}Pb +^{218}Po mixed	22.2 ± 4.8	23.8 ± 4.7
^{214}Pb mixed	137 ± 18	158 ± 13
Solar νER	102.3 ± 6.1	102.2 ± 6.1
^{136}Xe	55.6 ± 8.3	55.8 ± 8.2
^{127}Xe	3.15 ± 0.67	2.64 ± 0.64
Atmospheric CEνNS	0.124 ± 0.025	0.124 ± 0.025
^{85}Kr ^{39}Ar Gammas	162 ± 22	162 ± 21
$^{14}CH_3T$	58.3 ± 3.3	59.7 ± 3.3
AC	2.75 ± 0.64	2.76 ± 0.61
^{124}Xe	19.4 ± 2.5	20.4 ± 2.4
^{124}Xe Box Param, b		$1.44^{+0.12}_{-0.12}$
^{214}Pb isolated	606 ± 86	576^{+32}_{-31}
^{212}Pb +^{218}Po isolated	40.5 ± 5.8	39.9 ± 5.7
Rn Tagging Efficiency, ϵ_{rn}	0.600 ± 0.040	0.623 ± 0.026
Rn Tagging Deadtime, λ_{rn}	0.15	
^{8}B + HEP CEνNS	0.0626 ± 0.0065	0.0626 ± 0.0065
WIMP 10TeV/c^2		$4.7^{+4.1}_{-3.0}$
Total counts	1211 ± 92	1208 ± 33
Observed		1228
Veto Side band		
None Unique Sources	35.6 ± 2.7	35.6 ± 1.3
^{124}Xe	11.2 ± 2.4	9.4 ± 2.3
Det. NR		$0.0^{+2.6}_{-0.0}$
Flat ER Veto	140 ± 42	96^{+12}_{-11}
Veto Efficiency, ϵ_v	0.922 ± 0.043	$0.922^{+0.043}_{-0.043}$
Veto Deadtime, λ_v	0.0287	
Total Counts Veto	198 ± 42	141 ± 12
Observed Veto		138

Table 5.3: Table of the parameters from fit to salted WS24 data. The mixed/isolated and radon tagging separation is included but only the total counts for the WS24 is shown. The scalings in Table 5.1 may be used to derive the counts in each space. For example, none unique sources is the sum counts of those that simply scale to the veto sideband with the dead time.

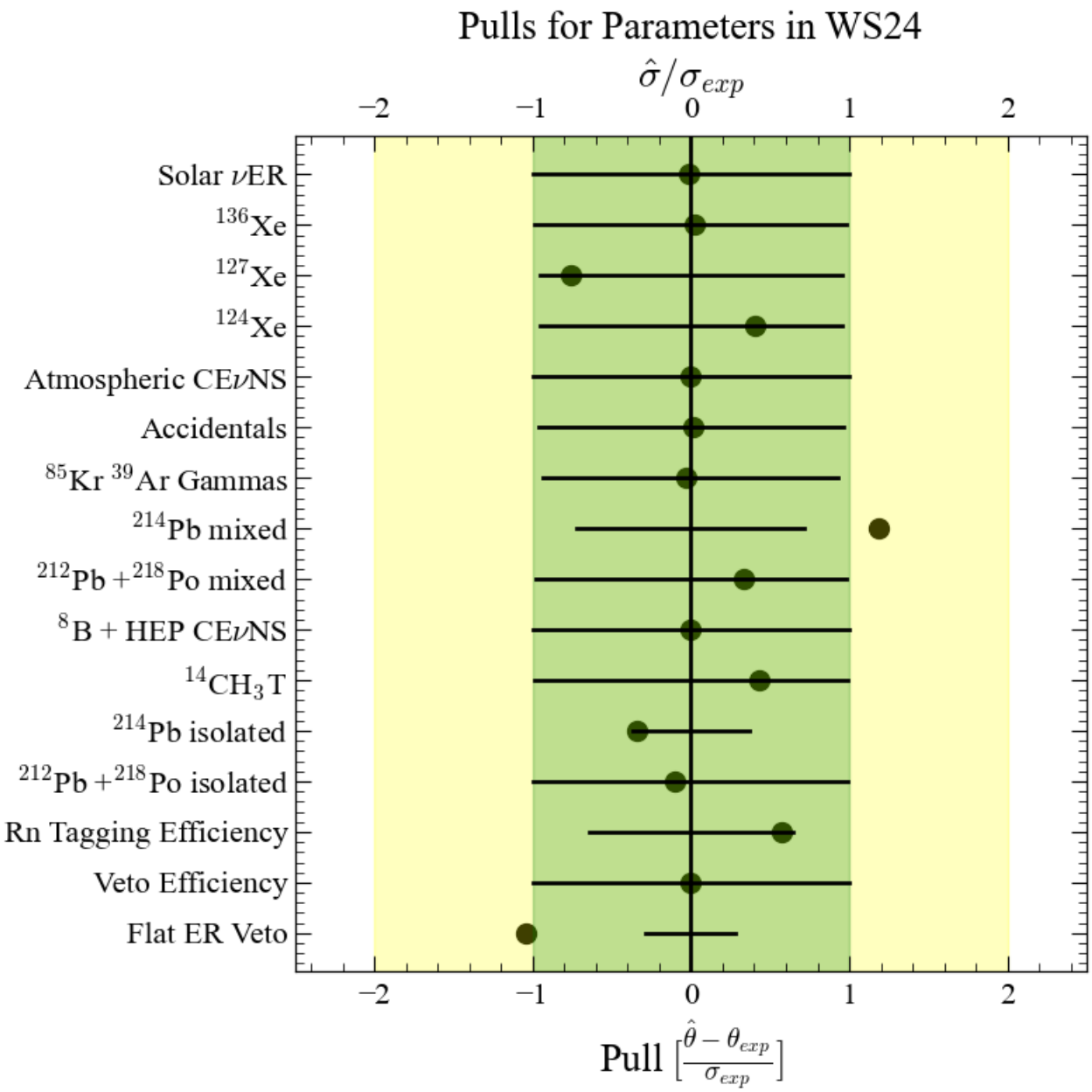


Figure 5.9: Pull plot of the parameters shown in Table 5.3 for which an expected value exists. Only the largest sources, ^{214}Pb isolated and mixed, Flat ER Veto, and radon tagging efficiency, linked to ^{214}Pb, are constrained by the fit itself. There is a significant over-fluctuation in the mixed state of ^{214}Pb and under-fluctuation of the Flat ER in the veto.

The GOF check in Figure 5.10 shows that there were significant deviations from the model, which included a 10 TeV/c^2 WIMP, in several places. First, an event in the radon untagged space at approximately (30,2.8) that is far from any background or signal model, except the AC, yields a strong over-fluctuation in the data. Despite the ^{124}Xe box parameter being free, the untagged space still shows strong over-fluctuations at high S1c. With the acceptable p-value, and complementary under-fluctuations in the ER bulk, this may just be a fluctuation. It is also possible that an event from the pedestal component of the salt model is contributing to this.

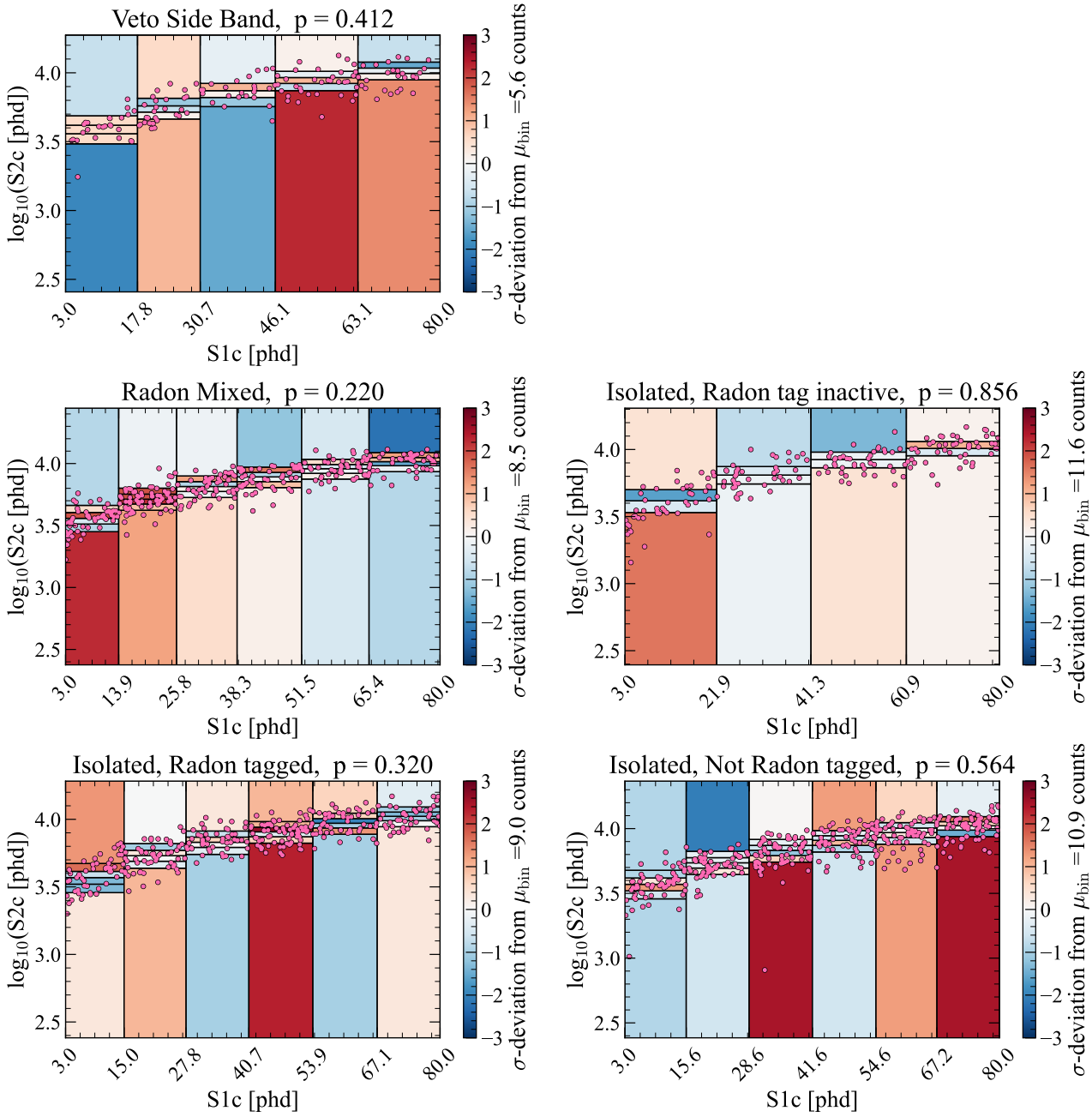


Figure 5.10: The 2D GOF of all spaces in WS24 from a fit using a WIMP mass of 10 TeV/c^2. All spaces pass the p-value requirement. There are strong over-fluctuations in the data below the ER band at high S1c in the untagged space and an apparent anomalous event at S1c $\sim$ 30 phd and $\log_{10}$(S2c [phd]) $\sim$ 2.9.

Given that there is still over-fluctuation in its relevant region, it was prudent to check the impact of the box parameter. In Figure 5.11, it is clear that the inclusion of the box parameter did not significantly improve the GOF. This is not surprising, the GOF is designed to check the background model with many events, not signal-like leakage, especially with a WIMP model present to absorb any deviation. Instead, a profile likelihood scan can show that both the best fit is adequate and optimiser is performing well, and that the model is appropriately absorbing WIMP events that are from some mixture of salt, signal, or genuine leakage caused by ^{124}Xe

DEC or some other process. The model cannot continue to absorb events due to the diminishing returns caused by the constraint on the ^{124}Xe counts, the backgrounds shape, and structure of the leakage.

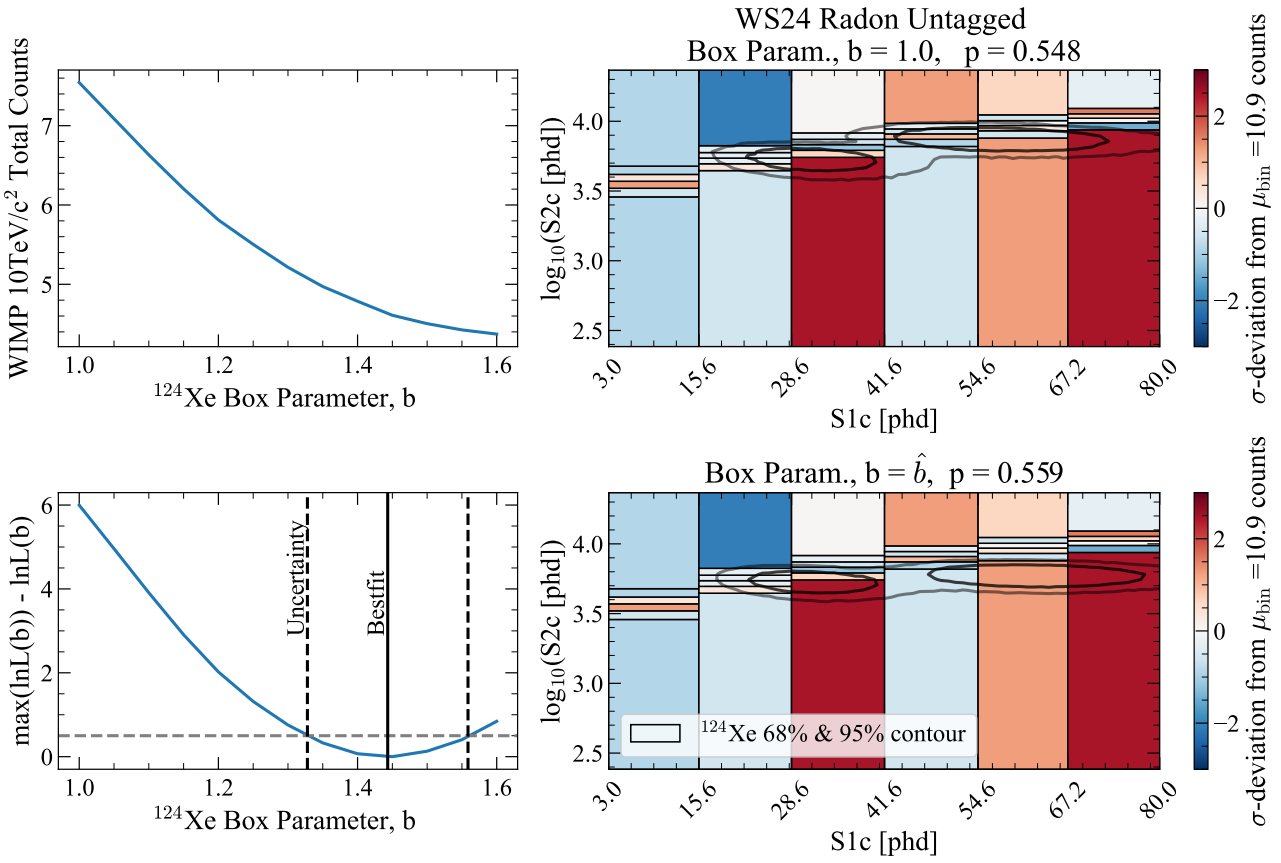


Figure 5.11: A profile scan of the likelihood (left) shows that the box parameter absorbs events otherwise attributed to WIMPs (top) and the data shows a preference for a model with approximately 1.7× greater ionisation density (bottom). The 2D GOF of a fit to data including a WIMP of mass 10 TeV/c^2 (right) with the box parameter fixed to a EC response (top) and floated (bottom). The test is not very sensitive to the box parameter.

With an adequate model, the inference routines for discovery and limit setting may be tested. As can be seen in Figure 5.12 there is a lower limit on the number of WIMP events and a significance of 2σ above WIMP masses of 1 TeV/c^2. With no clear errors in the model, just an unlikely configuration of the AC events and novel presentation of the DEC, the observed WIMPs are either salt or true WIMP events and the decision was made to un-salt. There were other thresholds set for un-salting based on pre-determined tests, such as the 1D GOF for event distance from the ER band median and in reconstructed energy. The author assisted in performing some of these checks, but all conditions of the model were satisfied and are discussed in the statistical analysis technical note in Ref.[160]. The information those tests provide is essentially redundant with the discussion above.

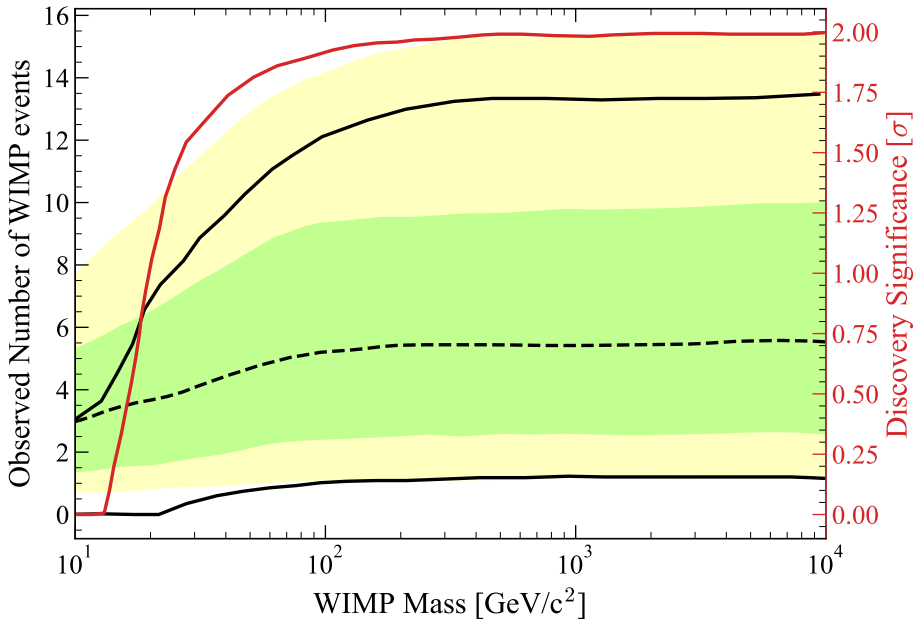


Figure 5.12: Inference results for the combined WS22 and WS24 likelihood with the solid black lines showing the upper and lower limits in terms of WIMP counts vs mass. The median expected upper limit, given the conditional best fit nuisance parameters to data, is the dashed line, with the 1 and 2σ Brazil bands given by the green and yellow shaded regions,respectively. There is an upper and lower limit for WIMP masses $\geq$ 20 GeV/c^2 as there was rejection of the background hypothesis, shown as a function of the tested WIMP mass in red and the right axis, of up to 2σ. Adapted from the statistical analysis note Ref.[160].

5.2.6 Un-salting WS24 and Including WS22

The removal of salt revealed 7 salt events shown in Figure 5.13, resulting in a best fit of zero WIMP counts shown in Table 5.4. This was more than was expected from previous fits, indicating either a strong under-fluctuation of the background model, allowing it to absorb some of the added salt events, or a mistake in the implementation of the model. It is evident from inspecting Figure 5.13 that there are not many suitable AC candidates present in the total WS24 data, and that some of the salt events were promising candidates. Despite this, the best fit AC events were mostly unchanged due to their strong constraint and a single event labelled anomalous at high S1c and low S2c. Despite the strong under-fluctuation, the goodness of fit in this space was still acceptable.

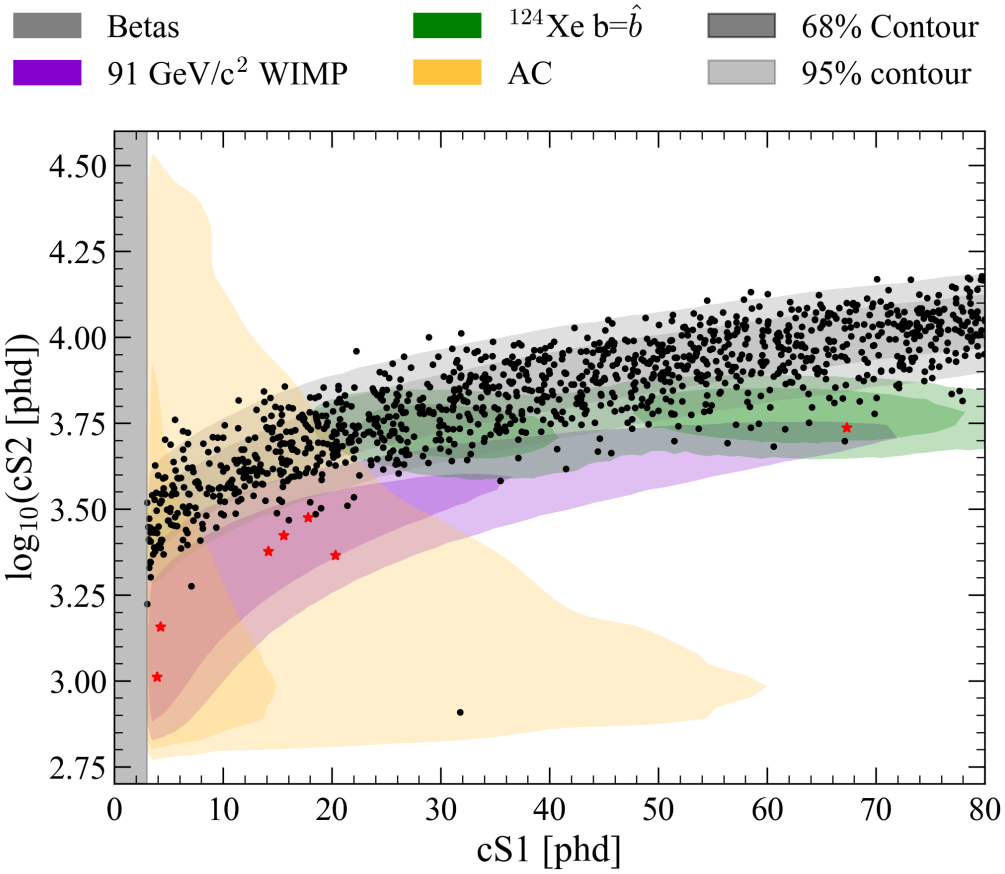


Figure 5.13: Scatter plot of the data post-unsalting, with the salt-events shown in red stars, for the entire WS24 data not including the veto. Several models are added to show their position relative to the salt. Note previously there are many salt events overlapping with the centre of the AC distribution.

Parameter	Predicted	Best fit	
		Pre Un-salting	Post Un-salting
^{127}Xe	3.15 ± 0.67	2.64 ± 0.64	2.66 ± 0.64
^{85}Kr ^{39}Ar Gammas	162 ± 22	162 ± 21	161 ± 21
Accidentals	2.75 ± 0.64	2.76 ± 0.61	2.61 ± 0.62
^{124}Xe	19.4 ± 2.5	20.5 ± 2.4	20.3 ± 2.4
^{124}Xe Box Param, b		$1.45^{+0.12}_{-0.11}$	$1.43^{+0.12}_{-0.12}$
^{214}Pb isolated	606 ± 86	577^{+32}_{-31}	575^{+32}_{-31}
Rn Tagging Efficiency, ϵ_{rn}	0.600 ± 0.040	0.623 ± 0.026	0.624 ± 0.026
WIMP 91GeV/c^2		$4.3^{+3.9}_{-2.7}$	$0.00^{+0.90}_{-0.00}$
Total counts	1211 ± 92	1208 ± 33	1202 ± 32
Observed		1228	1220
Det. NR		$0.0^{+2.6}_{-0.0}$	$0.0^{+2.4}_{-0.0}$
Total Counts Veto	198 ± 42	141 ± 13	141 ± 12
Observed Veto		138	138

Table 5.4: Table of parameters similar to Table 5.3 but only showing parameters which have changed between the two data sets. The Pre un-salting fit is shown again for the new WIMP model known to better match salt. The 10 TeV/c^2 WIMP model is the same except the best fit counts are $0^{+0.99}_{0}$, with a wider constraint likely due to some correlation with ^{124}Xe.

As in Figure 5.11, the 2D GOF does not improve with the box parameter, but the optimiser still prefers the higher density model. Varying fixed values of the box parameter in the unsalted data does not change the number of best fit 10 TeV/c^2 or 91 GeV/c^2 WIMPs. This indicates that the box parameter is being driven by an error in modelling of the ER band and not a result of salt or potential WIMPs. It should be noted that this is only true because the bulk of the WIMP distribution at all WIMP masses is at low S1c, this conclusion may be different if the signal were not constrained by this empty region. For example, if the signal were some mono-energetic NR source in the same region as the LL-peak. That is to say, these events could be attributed to some signal that is not ^{124}Xe, just not a spin-independent WIMP signal.

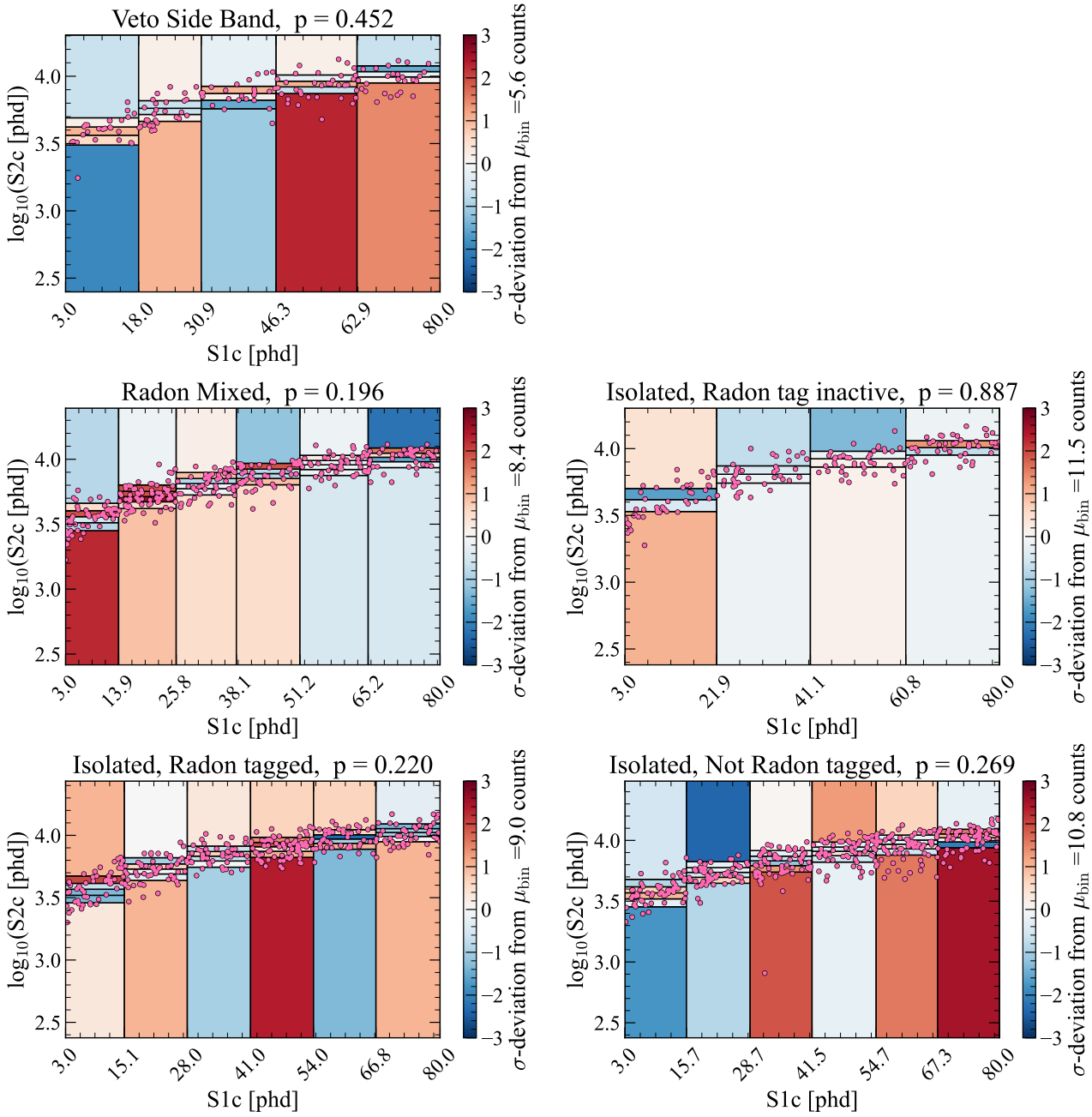


Figure 5.14: The 2D GOF tests and plots for all spaces with the un-salted data set with a 10 TeV/c^2 WIMP model. Generated using code developed by Leah Wolf and the statistics team, including author.

The limit curve with and without the WS22 data set is shown in Figure 5.15. It shows that the WS24 data set gives a -1 sigma under fluctuation across almost all WIMP masses. The inclusion of WS22 returns the strong -2 sigma under-fluctuation from an "arm-pit" effect caused by cosmogenic activation products in WS22, discussed in Ref.[95], where events were not observed in a strong overlap region expected from cosmogenic ^{127}Xe and ^{37}Ar. The power constraint limits the under fluctuation in both cases, and is dominated by the WS24 model and data.

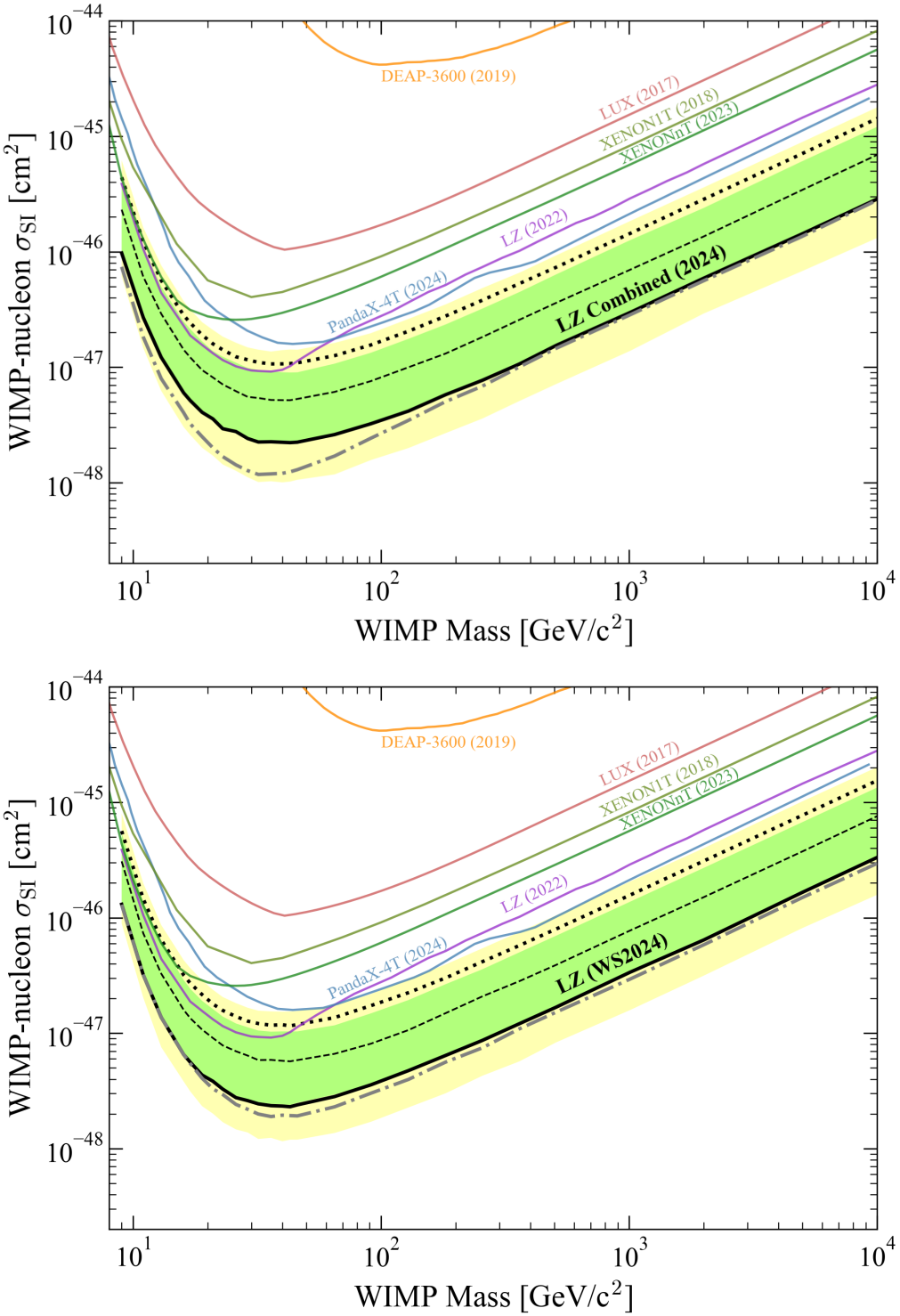


Figure 5.15: Derived upper limit and Brazil bands for WS24 with (top) and without (bottom) WS22 included in the combined likelihood. The definitions are identical to Figure 5.12, except there is only an upper limit which is also shown as unconstrained in the gey dot-data. The dotted line is the cross-section required to obtain at least 3σ evidence 50% of the time, calculated the same as the Brazil bands using nuisance parameters from the conditional best fit at each mass. It shows that there is a -1σ under fluctuation at all masses for the WS24 data and that this is enhanced for masses below 100 GeV/c^2 by the inclusion of the WS22 data. This was observed in WS22 and is discussed in the text. These plots were taken from the statistical analysis technical note in Ref.[160]. The curves from other experiments were obtained from Ref.[161, 162, 163, 164, 165, 166].

5.2.7 Cross-Checking Unsalted Result

As the dominant background, and with two of the salt events being near its peak, it is clear that the AC component is the cause of the under-fluctuation in WS24. If this is due to an error in modelling, it is conservative in the case of discovery, as it obscured several WIMP signal events. It is still important to investigate this, as in the case of reporting an upper limit, one could artificially improve their observed limit by overestimating their background. Intuitively, imagine if you expect a certain number of backgrounds and observe less than that - it is then even less likely that any signal is present. As this would worsen the expected sensitivity, the power constraint procedure discussed in Chapter 3 would correct for this, but there is a degree of freedom to game the result within the chosen power level. Therefore, it is important to check that the background model is not over-estimating the AC and that the observation is consistent.

First, to assert that this is not some mistake in the procedure, and that the under fluctuation is conservative in the case of discovery, the author created a pseudo-salt test where events are simulated from a custom salt model and added to the observed data. The custom salt model assumed 6 events from a 91 GeV/c^2 WIMP and 1 event from a flat pedestal, as the original model was. This was done for the nominal model, removing the accidental constraint, and removing the anomalous AC event. The results in Figure 5.16 show that even without a constraint the AC model absorbs most of the salt events and that removing the anomalous event restores 5 of the 6 expected 91 GeV/c^2 candidates, with the remaining lost to leakage with other backgrounds and overlap with ^{124}Xe in the tails. The Detector NRs showed negligible variation between each toy, as they are strongly constrained by the veto, but the ^{124}Xe box parameter did vary upward, suggesting it was absorbing at least the flat component of the salt model. Observing only 1 accidental event with 2.6 excepted is not unlikely, so this can be thought of as a shape-fluctuation whereby the AC event happens to not overlap with the WIMP model.

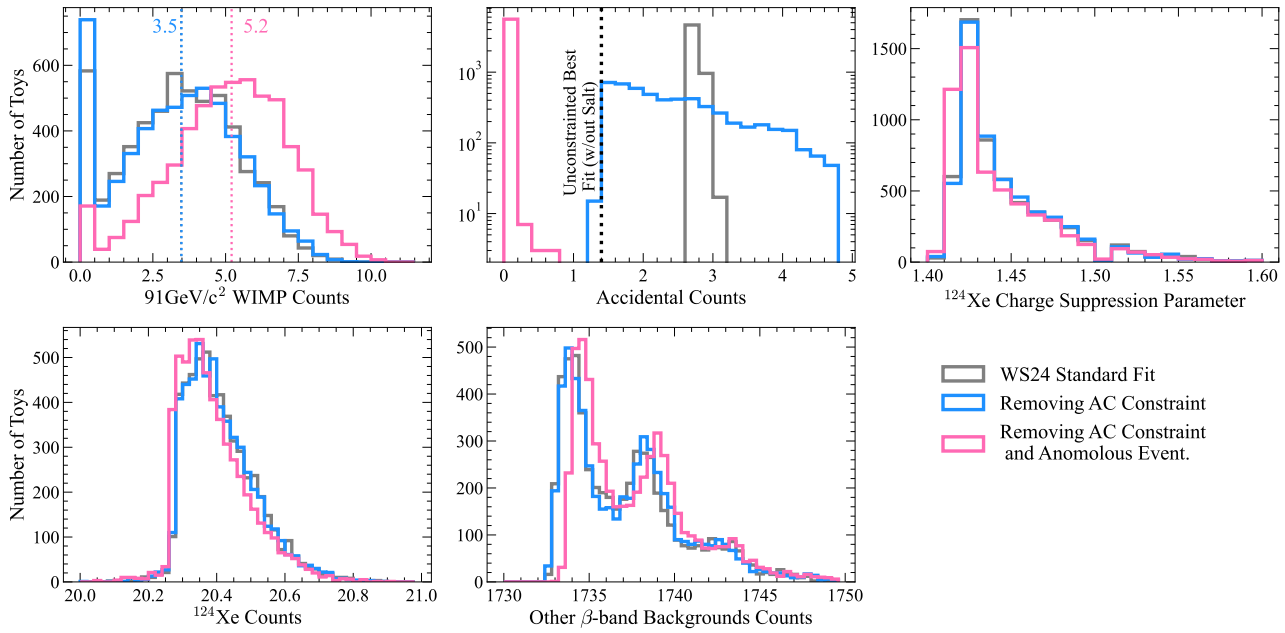


Figure 5.16: Pseudo salt study, injected a fixed number of 6 91GeV mass WIMPs and 1 Flat NR events into the unsalted observed data. The events are distributed randomly over the PDF and spaces appropriately without any Poisson fluctuations, so the number of salt events is fixed. Three configurations were chosen, the default WS24 model and dataset with all spaces, removing the AC constraint, and then removing the anomalous event as well. The left most extent of the histograms indicates the best fit to unsalted data. Caution should be exercised when interpreting a shift in the best fit counts of a background source as evidence for absorbed salt events. The best fit value for the beta component can increase by more than the total number of salt events, since adding events to the tail of the distribution can influence the likelihood function more strongly than the constraint on the total event count. The main result is that the recovered 91GeV mass WIMPs is unaffected by removing the AC constraint as the single event constrains the model such that WIMPs are still absorbed, removing it almost fully restores the recovered salt.

It should be noted that in the salted data-set the total number of WIMPs could be recovered by fixing the box parameter to a lower value, therefore more default-EC like, shown in Figure 5.11. This also allows the WIMP to increase above the number of salt events, and in the un-salted data does not contribute to any changing WIMP counts. Alongside the above study, this indicates that whilst one could resolve some of the under-fluctuation by fixing the box parameter, this would be arbitrarily assigning events to WIMPs that are better described by the DEC model. Further, in the case of discovery it could cause one to over-state significance or over-estimate the WIMP cross-section. Changing methodology based on what is conservative for a specific case, i.e a limit or discovery, is an arbitrary choice and can lead to being over or under conservative. The choice to allow the box parameter to float in the final model is therefore the best choice for accurately modelling the observed data and both cases.

Given that the main driver of the under-fluctuation in the WS24 data set is the anomalous event, it is prudent to quantify the degree to which it is expected. To do this a rudimentary

counting test was set up and shown in Figure 5.17, the frequency in which we observe the current configuration is 2%. This may fail a standard requirement of 5% but this is a rudimentary test. Further, as discussed in the statistics analysis note in Ref.[160], it was observed that there were two events barely failing the S1c lower threshold, one of which would enter region 1. If we consider one event in region 1, then the probability increases to 7%. Given these events, that this test is rudimentary, and that our GOF tests in Figure 5.14 show no model disagreement - it is reasonable to accept this outcome and not reject our AC model.

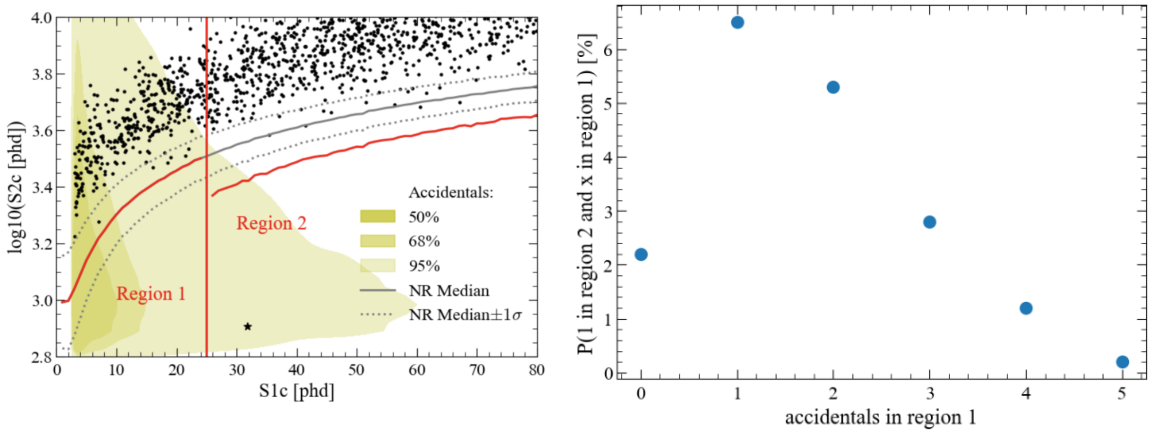


Figure 5.17: The regions for which the counting experiment was conducted (left). The frequency for which one event is observed in region 2 given a number of events in region 1.

Finally, the weaker under-fluctuation of the observed limit in Figure 5.15 to the bottom of the -1σ band is not surprising despite the low-probability from the above test, as the WIMP signal is not completely degenerate with the AC model. Most of the under-fluctuation at all WIMP masses can be attributed to the AC component. This may seem counter-intuitive, but the WIMP spectrum for all masses is an exponential in energy, so the near-threshold background of AC will impact every mass.

5.2.8 Constraint on Xe124 DEC Box Parameter

Despite its high efficiency, the radon tag does not significantly improve the WIMP sensitivity, as high discriminating power between ERs and NRs makes ^{214}Pb sub-dominant to AC background. However, since it qualitatively impacted the determination of the ^{124}Xe DEC model the author wanted to quantify this. For this a basic WS24 was defined, with total exposure combined and the Veto and WS split, and the full Split/Tagging model, with the WS split into mixed, isolated radon inactive, tagged, and untagged. A simple shift of the box parameter $b \rightarrow b - 1$ allows the same inference routine for WIMP counts to be used, with b=0 being the EC model, and anything greater is the DEC model. In Figure 5.19 the test statistic distribution is shown, and although it is not fully asymptotic, we can see that the approximation is conservative as it consistently overestimates the p-values and underestimates the significance. Using the asymptotic approximation, the discovery significance is 3.5(2.6)σ with(out) the radon tagging.

This radon-tagged asymptotic significance is expected from the Gaussian errors on the value in Table 5.4. Using non-asymptotics boundary effects raise this to be 3.9(2.9), but asymptotic is quoted for later comparison. To verify that this is expected, a sensitivity study is also run, shown in Figure 5.19, predicting expected discovery significance for each value of the box parameter in WS24, showing that an under-fluctuation has occurred in the background for this search, but it is well within expectation. The underfluctuation may have been caused by the ^{214}Pb in the isolated state, as seen in Figure 5.9. This means that the radon tag provided quantitative evidence to reject the ^{124}Xe EC model and accept the DEC model. With further exposure, we may find that ^{124}Xe is not the only novel ER, both the double beta decay of ^{136}Xe and inner-shell scattering of solar neutrinos may create denser and therefore more NR-like events. Using the radon tag to identify and study these sources in LZ will be essential to predicting the performance of XLZD, where these backgrounds will be dominant. This also emphasises the need for shape-varying parameters in our inference if sufficient precision can not be attained in their models.

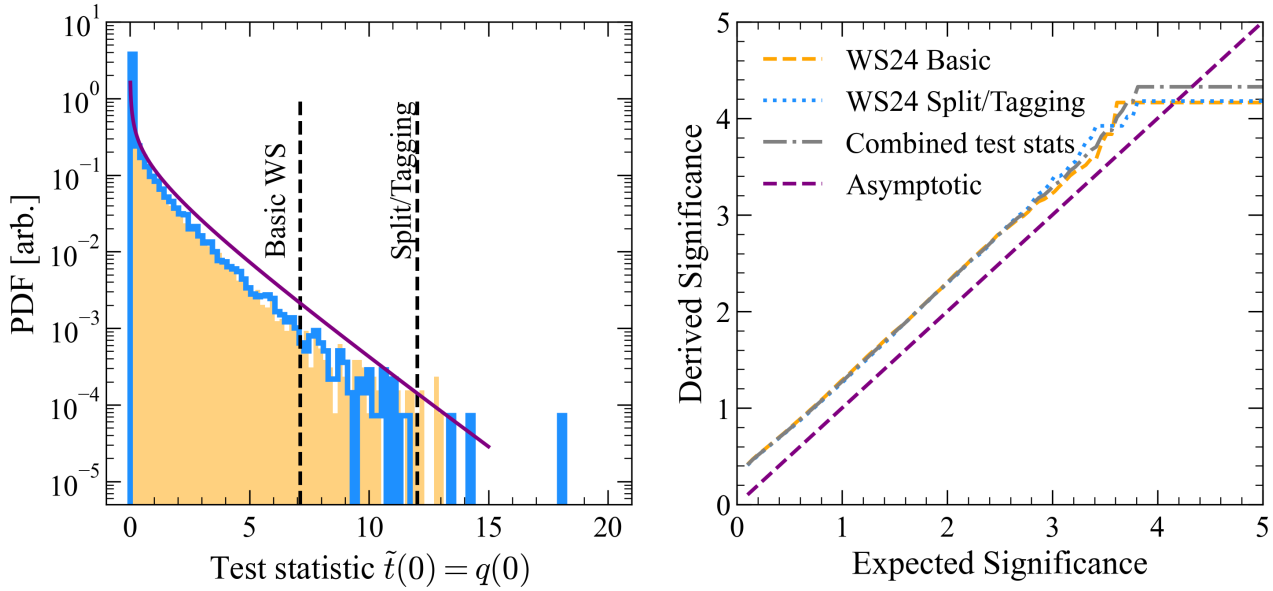


Figure 5.18: Evaluating the test-statistic distributions assuming the default EC model and conditional best fit to observed data of nuisance parameters, $f(\tilde{t}(b=1)|b=1, \vec{\theta} = \hat{\hat{\vec{\theta}}}(b=1.))$, for the box parameter discovery significance using 68k toys and ∼5k hours CPU time, 50-node hours with 100 tasks per node, with the observed test statistics of the two configurations in grey vertical lines (right). The colors are toys generated assuming a basic and split likelihood, following those in the legend of the plot on the left. The derived significance from the test-statistic distributions vs the expected from asymptotic (left). There is a constant shift in significance as the beta-band in many background toys may cause the maximum likelihood to prefer a box parameter $b < 1$, but the physical boundary forces this to 1. Therefore there are many toys for which the $-2ln\left(\frac{L(b=1.)}{L(b=\hat{b})}\right)$ would be greater than 0 but is forced to zero as $\hat{b}$ is forced to 1. Note that with and without the tagging and splitting of the WS, the distributions are identical but observations in the grey lines are distinct. Combining the distributions and doubling the test-statistics does not significantly increase accuracy, as one would need orders of magnitude more toys to do so. When the observed significance becomes greater than the end of the test statistic distribution, on the left plot, the derived significance plateaus, as seen on the right plot.

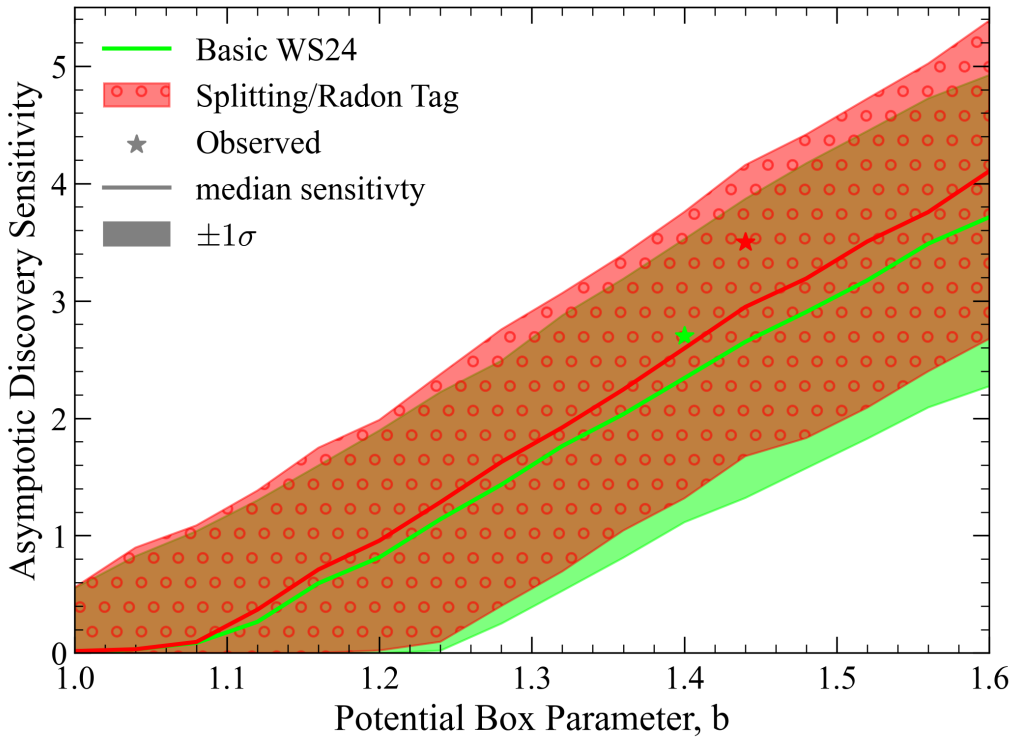


Figure 5.19: The median and $\pm 1\sigma$ band of the distribution of asymptotically derived significances under different box-parameter assumptions and conditional best fit to observed data nuisance parameters, $f(p(b = 1.)|b = b, \vec{\theta} = \hat{\vec{\theta}}(b = 1.))$. The observed significances both over-fluctuate above the median, suggesting a reduced background. This would be strongly impacted by the ^{214}Pb rate and distribution, and may be due under-fluctuation in the isolated state in Figure 5.9 which would also explain the increased over-fluctuation of significance when applying the tag here. No corrections are applied for the shift shown in Figure 5.18.

5.2.9 Summary of Template Based WS2024

The novel features of WS24, split between two periods and flow states with different rates, the inclusion of radon tagging, and the inclusion of a veto sideband necessitated the use of more adaptable and transparent tools. For this, `LZFlameFit` and `flamedisx` were used and the author was key to adapting `flamedisx` for this task. Namely, the author re-designed the template sources such that background counts could be used consistently as a parameter, designed the combined likelihood structure, and use of linear template interpolation that existed in the codebase. The use of salt allowed analysers to determine that there was greater leakage than expected at high S1c, this was determined to be from the DEC of ^{124}Xe giving increased density ER deposits, suppressing S2c. This DEC background causes increased overlap with the signal model, parametrised by a box parameter for the LL peak. Linear template interpolation of the templates has poorly defined gradients and Hessians, which does not interface well with `flamedisx`. This showed bias in the inference of this box parameter so the author resolved this by introducing B-spline and cubic spline interpolation, and after weighing their merits, B-spline was used in the final inference. The author was involved in preparing and processing

GOF checks on the salted data-set and without any major anomalies present the data set was unsalted. This revealed a greater number of salt events than was inferred in the salted fits. The author showed that this was likely due to the anomalous realisation of the AC model, which was reasonably likely and not a great cause for concern. As an extension, the author showed that the radon tagging quantitatively improved the determination of the apparent non-ER behaviour of the ^{124}Xe DEC like events. Finally, the author would like to thank the statistics team for a chance to work on this, and for being supportive, hard-working, colleagues.

Chapter 6

Towards Full `flameNEST` in LZ Inference

Recall from Chapter 4 that `flameNEST` is the implementation of explicit likelihoods in the NEST. That is, instead of sampling successive probability distributions to produce binned PDFs, it constructs ranges of possible parameters informed by the observed data and evaluates the probability distributions for those ranges. This scales far better than MC templates, the former method, for more observables and shape varying parameters. It did have some performance problems with the NEST models, but the author resolved those in Chapter 4. Note that, unlike MC templates, it does scale linearly with observed data points and number of sources. Upon resolving these problems, and implementing LZ's inference tools in a `flameNEST` compatible way in Section 5.1, the author had the choice to either integrate `flameNEST` into inference on science data or in calibrations[1]. As noted in Section 6.1, calibrations currently use an averaged response and remove all field, or other, position dependence from the model. This means that the tuned NEST model may not be naively extrapolated to other fields and detector effects have potentially been convolved with microphysics ones. However, in constructing a fit to science data the author can show that it fully repeats the simulation and background modelling chain. As well as allowing for immediate future adoption in novel searches with WS24 conditions, this proof-of-principle would increase the chance of wider collaboration adoption. Therefore, the author chose to implement a fit and evaluate how ready the tool was for non-asymptotic inference.

The use of the full `flameNEST` calculation in WS24 allows for improved background discrimination, model clarity, and multiple shape-varying parameters. Improved background discrimination occurs not only due to the spatial distribution of some background sources but also from that of the detector response itself, as position-based corrections create degeneracy by projecting a 6D space (S1,S2,x_{obs},y_{obs},drift time, event time) to a 2D one (cS1,cS2). Whilst the model response has been treated as uniform here, due to only minor variation over WS24, the event time still allows us to discriminate $^{14}CH_3T$ and xenon activation isotopes by their decay time. As noted in the previous chapter, MC template interpolation can introduce computational errors and becomes more complex with more parameters. For using shape varying

[1]Recall the challenge remaining for calibrations, the problem of many inseparable parameters in total rate estimation, readily resolved by the application of Markov chain MC as discussed in 4.

parameters in flameNEST, one only needs to consider the interpolation of a scalar MC estimated value, the total rate, which is simpler and requires less resources. Finally, the flameNEST object oriented structure allows for easy modification of the model, and interpretability, allowing for more transparent reporting. The strength of flameNEST would be to integrate calibrations, simulations, and inference into one tool. As carrying out inference is the most complex of these tasks the author focussed on this in order to prove flameNEST's functionality. Separately, after the author's work, flameNEST can be used to produce the same MC templates as the LZLAMA + BGSkimmer simulation chain. Whilst this does not take advantage of the full functionality of flameNEST, it benefits from centralising the modelling tools in a single place and making it easier to develop and modify different models and configurations.

In Chapter 5 of Ref.[111], Dr Robert James implemented spatial and time dependence by calculating the differential rates for each source once in flameNEST and used the result essentially the same way as a PDF in Section 5.1 of this work. This meant that shape varying parameters could not be included. Here, the author presents the full calculation as a proof of principle using the complex likelihood model of WS24. As WS24 is a large accumulation of work, the author used what was available for this analysis from various working groups and notes where improvements on the model would be needed for a formal result. The major challenges the author needed to overcome were; to update the flameNEST models to agree with the re-calibrated LZLAMA and NEST, refurbish flameNEST as previous approximations had lost functionality, and build the full likelihood model. In addition the author needed to integrate flameNEST into the current GOF tools.

The next section will cover the configuration of flameNEST and describe specific modelling choices for WS24. Then the spatial and temporal ingredients gathered by the author are presented. Finally, the first flameNEST fit in the combined likelihood framework of LZFlameFit and first full flameNEST fit to LZ WS data are presented. As a full inference result would then be premature and possibly biased, due to the immature state of some inputs, the author presents pull-tests of the resultant model to confirm it is self-consistent, quantifies the improvement to discrimination, and estimates the computing requirements for a full limit calculation. In addition, the author explores several technical aspects that could be used to reduce the computational resources required.

6.1 Configuration

There are three main aspects of the configuration of flameNEST, separate from the spatial and temporal dependencies of energy deposits. These are the detector response, physics modelling, and analysis cuts. As the WS22 configuration is well documented in Ref.[111] and covered in section 2.2 we only cover the changes to WS24 that the author ported from the MC templates

simulation chain into `flameNEST`. At the time of writing all parameters are contained within unmerged branches of `flamedisx` [2] including source energy spectra derived by Dr Robert James in Ref.[111]. Private LZ information, such as the efficiency curves, LCEs, and position distributions discussed here may be made available upon request.

6.1.1 Detector Response

The greatest difference between WS24 and WS22 is the field configuration, notably that the cathode was decreased from 34kV to 18kV, decreasing the average drift field from 192.9V/cm to 97.1V/cm. The Gate and Anode were also adjusted, changing the extraction field and therefore S2 properties. The author had contributed to the decision to lower the drift field, which will be discussed in Section 7.1.The extraction field also changed in this period, alongside it the measured values of g_1^{gas} and g_1 changed to 0.112±0.002 and 0.0764±0.0020 phd/photon, respectively. The electron lifetime improved to 9000 us and it was found that a non-zero S1 linear noise parameter provided a better fit to data.

With the weakened drift field and the accumulation wall-charge, modelling of electric fields and electron drift required updating. This is inferred in large part using the data-wall, the point radially at which data appears to stop before the physical TPC wall. The data-wall is caused by two factors, deflection of S2 electrons inwards and capture of S2 electrons on the wall. For WS24, a field and drift map were derived from first principle simulations tuned to observed data to account for wall-charge. These can predict: the strength and direction of a field at an event's location; its observed S2 (x_{obs},y_{obs}) position; its derived drift time; the probability of the S2 attaching to the wall; and the horizontal diffusion of the S2, i.e the uncertainty on its position. For WS24 the drift map was used only to map simulated, or true, position to observed, and field was approximated to be uniform. As discussed in the position dependence section, the author chose to use the observed position consistently throughout the models in `flameNEST`, not distinguishing between observed and true position. For future studies, and a cross-check in the following material, the author implemented the drift field in both simulation and the differential rate calculation for `flameNEST`. This was implemented for both true and observed position, the latter convolving the field and drift maps, both are shown in Figure 6.1. The drift map was also introduced in the simulate function of `flameNEST` to bring it closer to `LZLAMA`.

[2] All work in this Chapter was carreid out using `https://github.com/FlamTeam/flamedisx/tree/JRG_WS2024DD`.

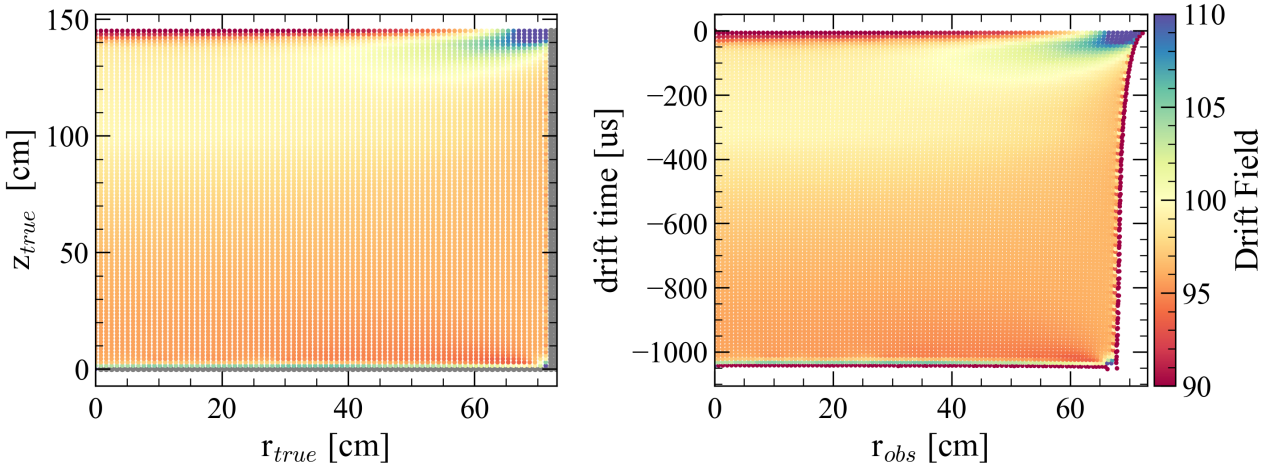


Figure 6.1: Field map before and after conversion into observed variables with the drift map. The field is assumed to be uniform azimuthally.

The final component of the detector model are the LCE maps, these are data-derived and therefore given in observed position and shown for WS24 in 6.2. As explained in Section 2.2.3, the LCEs were updated for WS24 using a source that should be less dependent on field; therefore, future analyses may use them in studies with a non-uniform field.

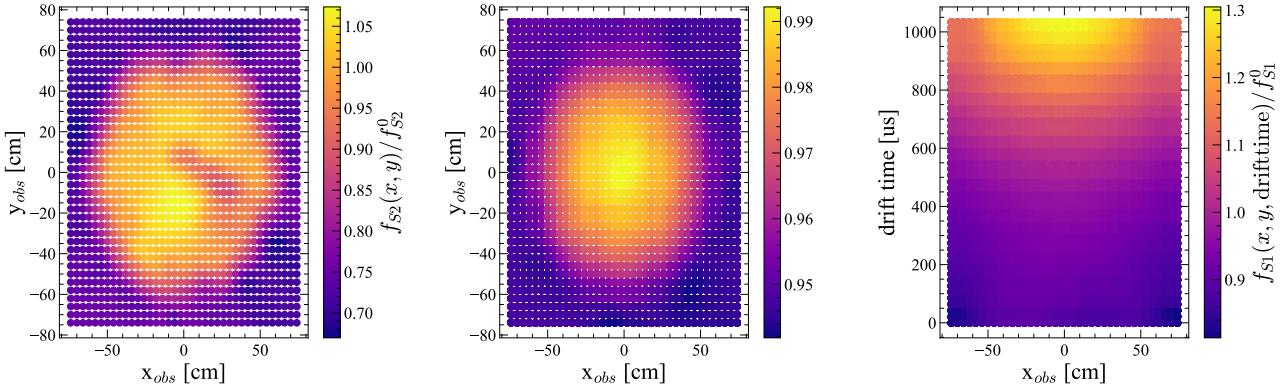


Figure 6.2: Light Collection Efficiency (LCE) maps evaluated for fixed points over the detector for S2 (left) and S1 (middle,right). The S1 LCE is in drift time as the position of an event impacts efficiency, with greatest efficiency at the bottom of the detector, whilst S2s are all generated in the thin extraction region at the top.

6.1.2 Physics Modelling

Here we discuss some aspects of the physics models that changed between WS22 and WS24. The author updated all of these in `flameNEST` with some assistance from Wei Zha in cross-checking the ^{124}Xe box model. The physics responses are grouped into ER, NR, and unique in Table 6.1 with a brief explanation of their spatial and temporal dependence to be discussed in the next section. Despite their potential for increased density deposits, ^{136}Xe double beta decay, solar neutrino electron scattering, and detector ERs in the WS ROI are assumed to follow the

NEST ER model. The latter because they are subdominant in the ROI so any departure from this model will have a negligible effect.

Source	Model	Dependence	Explanation
ER Sources : ER model tuned to ER $^{14}CH_3T$ calibration data. Grouped†.			
^{214}Pb Isolated		Space	Inferred with flow and field simulations.
^{212}Pb/^{218}Po Isolated		Space	Discussed.
^{214}Pb/^{212}Pb/^{218}Po Mixed		N/A.	Mixed state removes any spatial dependence.
$^{14}CH_3T$ Tritium		Time	Exponential decay from injection time due to removal by hot getter.
^{85}Kr		N/A	From contamination/slow ingress
^{39}Ar		N/A	Activation but negligible.
^{136}Xe		N/A	Long Lived Isotope.
Detector ERs		Space	Gammas from materials, depends strongly on source position, self-shielding, and vetoes.
Solar ν		Time, Ignored	Rate varies with distance from Earth to Sun.
NR Sources: Tuned to DD Calibration Data			
Detector NRs		Space	Neutrons from materials with distinct shape and spectra in the veto-sideband.
WIMPS		Time Ignored	Rate varies with the motion of Earth relative to the WIMP wind.
Atmospheric & ^{8}B CEνNS		N/A	negligible so removed
Unique Sources			
^{127}Xe + ^{125}Xe	EC	Space	Prompt Gamma ray gives strong self-shielding and veto effects.
^{124}Xe	DEC sec.5.2.3	b	Custom box-parameter controls deposit density.
Flat ER Veto	Gamma/ER	Space, ω_γ	Veto Tagged Detector ERs, uncorrelated with ROI. Unique weighted between Gamma and Beta.
Accidentals	Template	N/A	Data-driven template based approach. Assume random pairs have a flat spatial or temporal distribution.

Table 6.1: Table summarising model choices for this `flameNEST` analysis.
† - All these sources are grouped into a single calculation, see Section 6.6.1.

The NEST models for both NRs and ERs are modified for WS24 to be constant with drift field and have a simpler form than in Ref [113] and Section 2.2.1. Explicitly, the terms in Equations 2.2.1 and 2.7 become constants when they depended on field, along-side other components like the various Fano factors. The maximum variation of the drift field in the active and fiducial volume are ∼10% and 5%, respectively. Despite the impact of this variation on both ERs and

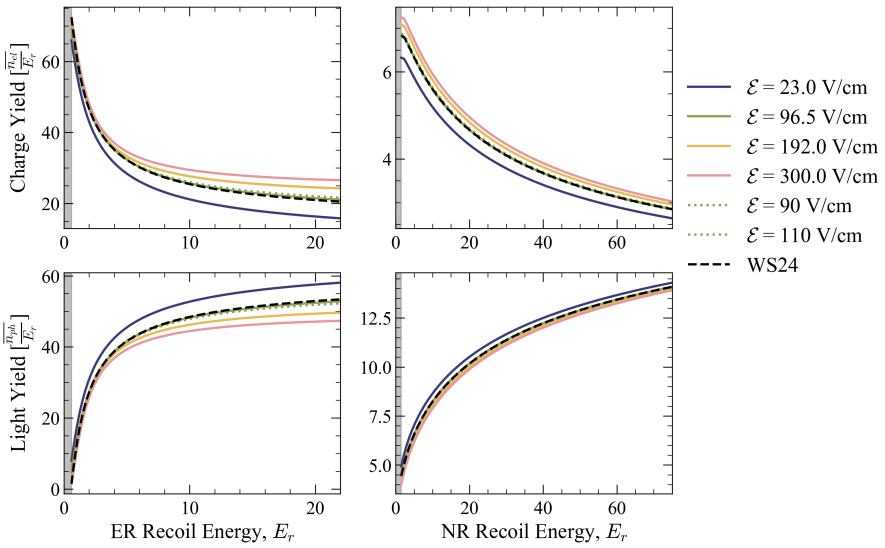


Figure 6.3: Default NEST models from Ref.[113] plotted with various fields, WS22 used these models and an average field of 192 V/cm. The Field for WS24, 96.5 V/cm, is plotted alongside the minimum and maximum drift fields from Figure 6.1. This causes a maximum light/charge yield variation of 0.2/1.1% in NRs and 0.6/4.3% in ERs.

NRs being relatively small, using an averaged response is ultimately a result of being unable to include position-dependent information in calibrations, and may prove to be limiting at higher energies where drift field impacts ER charge and light yields more. It may also impose a bias in the NR model, as the strongest variation in the field, shown in Figure 6.1, occurs near a DD conduit used for calibrations. The ^{127}Xe and ^{125}Xe EC follow the gamma-beta weighted model. The gamma model remains unchanged from default NEST but the weighting with beta is now a constant value, instead of being field dependent. This weighted charge yield gives the $Q_y^{\beta-\gamma}$ from Equation 5.3 used to calculate the charge yield of the ^{124}Xe DEC. To calculate this, the model parameter ζ^L alongside the other parameters discussed here.

The Flat ER component in the veto consists mainly of gamma ray scatters from Detector ERs and were assumed to be uncorrelated with the ROI Detector ERs. In the template analysis they were modelled as pure betas in NEST but as they are the dominant source in the veto, the author implemented a gamma-beta weight parameter in the charge yield like in the EC model. This was left unconstrained but bounded between totally beta and totally gamma, 0 and 1, respectively. In the future, a calibrated estimate of the veto gamma-efficiency could be used to correlate the ROI and Veto Detector ERs. The rate of this source includes a set of semi-naked ^{214}Pb decays where the ^{214}Bi daughter is left in an excited state, the relaxation gamma is then tagged by the OD. This relatively small contribution would have a different spectrum, position distribution, and would not be impacted by this charge yield. These considerations have been left for a future analysis.

The AC model is a data-driven distribution of S1c-$\log_{10}$S2c and was discussed in Chapter 2.2. It is implemented in the same way as Ref. [111]. This implementation assumes that the source has the same spatial distribution as a uniform ER source and that its S1c and $\log_{10}$S2c can be transformed using the LCE at a measured point to the S1 and S2 used in `flameNEST`. In the future, this model may need to be modified as these corrections depend on the (x_{obs},y_{obs}, drift time) of the observed events assuming a physical source, but for an AC event the S1 and S2 do not necessarily come from the same physical origin. This means that the position should have no impact on the S1 LCE nor would the drift time measured have an impact on the S2. In the future it will be possible to replace this with a more robust AC model, but for now assuming that they behave identical to physical sources is conservative.

6.1.3 Analysis Cuts

The analysis cuts discussed in Section 2.4.1 are mostly applied by the `BGSkimmer ALPACA` module on `LZLAMA` simulation outputs. Unfortunately, `flameNEST` currently does not model skin and OD responses and whether or not the incident particle has a veto tag can affect the energy spectra and position of deposits inside the TPC, which is explored in Section 6.2.2. Aside from this, `flameNEST` can model the same RQs as `LZLAMA` so the fiducial volume and ROI cuts in S1c-S2c can be directly ported from `BGSkimmer` to `flameNEST`. However, `LZLAMA` was not configured to produce pulse-based information, without which the impact of certain detector effects and of data-quality cuts cannot be evaluated. To resolve this efficiency curves are used.

The impact of data quality cuts on observed events was derived by applying these cuts to $^{14}CH_3T$ ER calibration data and cross-checked with NR chopstitch using a mixture of $^{14}CH_3T$ ER and AmLi NR calibration. This work was carried out by the authors of the cuts and efficiencies technical note Ref.[167], building on work from various other groups, and the author summarises the origin of each efficiency curve here. As there are two sets of cuts, S1 and S2 based, there are two derived curves that are functions of cS1 and Log_{10}cS2 but in WS24 were observed to be constant. The S2 trigger dominates data collection and has an associated efficiency that is a function of $S2_{raw}$ measured by comparing S2 trigger with GPS/Random trigger in $^{14}CH_3T$ calibration data. The ability of LZap to correctly identify a SS event is derived from DD and $^{14}CH_3T$ calibrations, cross checked by hand-scans, and is a constant value. Finally, as S2 electrons drift, they diffuse along their direction of travel. This can cause the resultant S2 wave-forms to split and LZap to miss a fraction of the S2 signal. Since LZap is tuned to calibration data this effect is more pronounced for small S2-signals, which $^{14}CH_3T$ does not produce, and near the bottom of the detector, where the DD does not have a conduit. As such, this effect is modelled by a splitting-efficiency derived from full-chain simulations by Dan Hunt. The author implemented these in `flameNEST`, and their values and functions are in Figure 6.4.

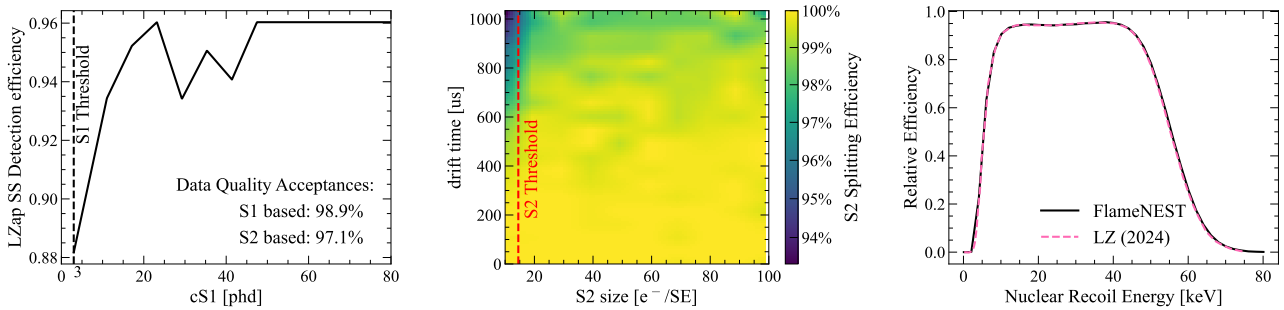


Figure 6.4: The acceptance curves for WS24 ported to `flameNEST` by the author, the S2 based trigger efficiency is 100% well-before the lower S2 threshold so is omitted here. These are summarised in the background modelling technical note in Ref.[156] using the trigger and cut efficiencies tunings from technical notes [167, 168, 169]. The final NR efficiency within the FV is also shown and compared to that of the WS24 paper in Ref.[90], here the uncertainties from calibrations are omitted.

The author re-derived the NR efficiency curve from Ref.[90] with `flameNEST` in Figure 6.4, which shows close agreement. These are derived within the fiducial volume, so are not significantly impacted by choices between observed and true position.

6.2 Including Position Dependencies

There are two types of position dependence to consider, where the position deposits are non-uniform and when they are uniform, but detector effects create a non-uniform observed position. Whether or not a source has uniform deposits is indicated in Table 6.1. The author derived the initial non-uniform position maps for WS22 from `BACCARAT`+`LZLAMA` simulations and worked with Dr Robert James to develop reasonable approximations for the distributions used in Ref.[111] and here.

6.2.1 Spatially Uniform Deposits

Although the signal is uniform, it is essential to correctly model the uniform deposits to implement spatial dependence for the spatially non-uniform deposits. There are several cases where detector effects not modelled in deriving the corrected variables (cS1, cS2) create an effective spatial dependence and improve discrimination. As discussed in Section 6.8, this is not the case here.

In addition to the physical location of the energy deposit, the observed position of an event is impacted by electrostatics, position reconstruction effects, and light collection efficiencies (LCEs). The impact of LCEs is well modelled by `flameNEST` with data-driven maps for S1s and S2s. The electrostatics can be modelled by the axisymmetric drift map and field map. However, this fails to capture an azimuthal dependence of the data wall or reconstruction uncertainty captured by additional corrections and S2-based position smearing in `LZLAMA`. The author implemented

the drift map in `flameNEST`, but unfortunately the addition of S2-dependent position smearing would require another random variable in the block model, increasing its dimensionality. Although this is not prohibitive, for WS24 it was found to be reasonable to ignore this detail under the assumptions that:

1. Observed position is an accurate measure of true position, i.e that $R(\vec{r}) \approx R(\vec{r} + \ \sigma_{\vec{r}})$.
2. S2 dependent smearing within the ROI is small such that $R_0(\vec{r}_{obs}, S2, E) \approx R_0(\vec{r}_{obs}, E)$.

The first ensures that we do not need to add another random variable to the block-model and can use observed position maps derived from `LZLAMA`. The second ensures that we need not bin our position-dependent energy spectra in an additional dimension. The derived position maps would capture the behaviour of the data-wall and modify the rate of the source accordingly.

To verify assumption 1, a sample of events was generated in the centre of the detector, the differential rate was then calculated in `flameNEST`, and then the observed positions in radius and drift time were changed. The gradient of the sum of the log differential rates, equivalent to the shape term of a likelihood, with respect to the radius is then approximated at each point. For this test, the original NEST models were re-introduced for ERs and field-map used to produce full position dependence. In figure 6.5, we see the results and predicted position uncertainty from the NEST models. For WS24, the gradient never exceeds 2.5%/cm and the smearing is less than 1 cm throughout the physical volume, so we are safe to ignore this $\mathcal{O}(1\%)$ impact on the differential rates. The strongest rate of change is at the data wall, meaning that the FV could be safely expanded in future analyses. This also shows that Assumption 2 will hold for the WS24 ROI, as the impact of ignoring any S2 dependence would be of the same order of magnitude as this gradient.

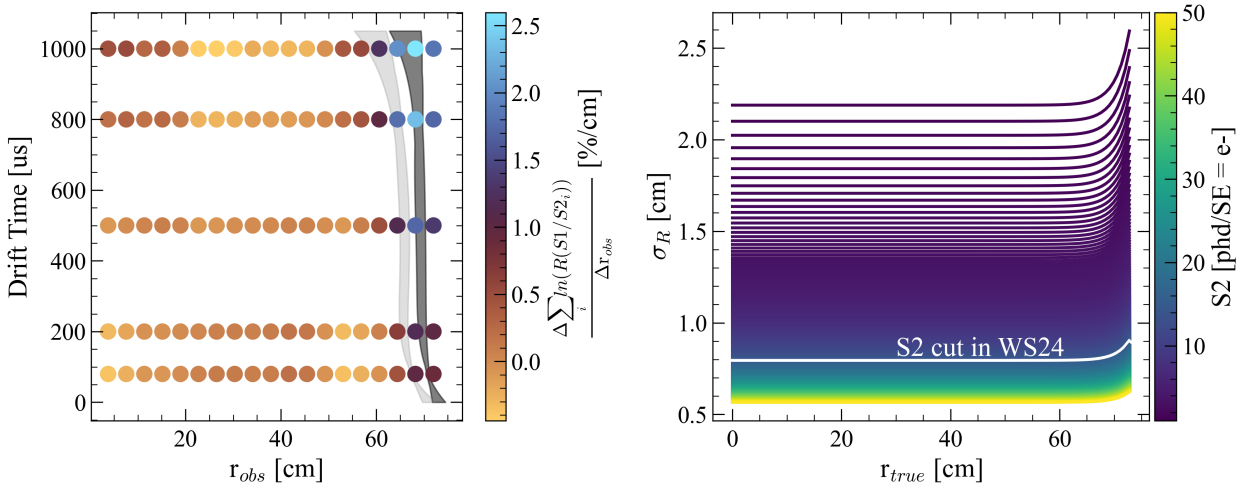


Figure 6.5: The approximated gradient of log sum differential rates for a 5 keV$_{ee}$ mono-energetic ER source using 50 events with respect to radius (left). The S2 and true radius dependent width of the Gaussian position smearing applied in `LZLAMA` to approximate reconstruction uncertainties for different S2 sizes in electrons (right).

Although the S1c-S2c distribution can strongly depend on the spatial distribution, the reverse is not true without applying cuts in that space. Using a uniform flat ER or NR spectrum to generate positions in `LZLAMA` allows us to create a position map (x$_{obs}$,y$_{obs}$,drift time) for uniform deposits that includes wall tuning and position smearing without implementing them in `flameNEST`. In figure 6.6 the derived position distributions for ER/NR are shown to agree, as expected. 6.6

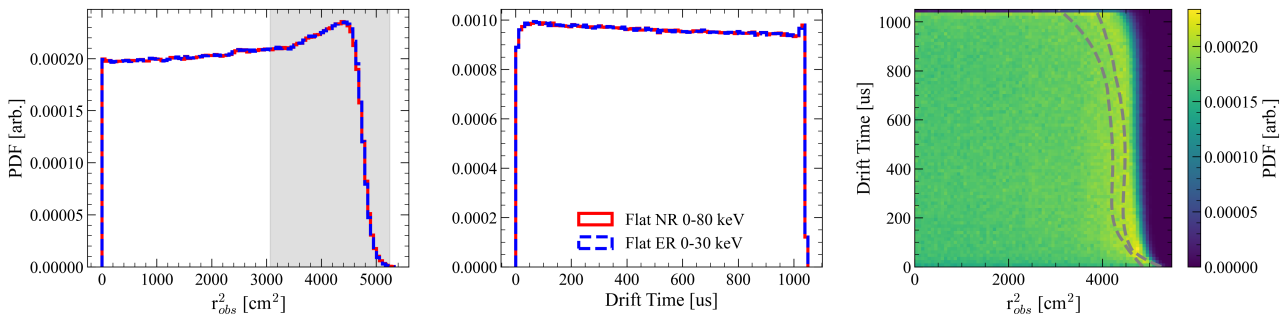


Figure 6.6: Spatial distribution of a uniform source from `LZLAMA` with no LCEs applied. The radial (left) and vertical distribution (right) compare ER and NR to show no distinction. The radial distribution is modified correctly to mimic the wall, and the minimum and maximum extent of the FV cut is shown by the grey band. The 2D PDF shows the minimum and maximum extent of the FV cut azimuthally (right).

6.2.2 Spatially Varying Deposits

Details of the position distributions and spectra for ^{214}Pb, Detector ERs and NRs, and ^{127}Xe are provided here. The physical source of this dependence is covered alongside the work by the

author to produce a reasonable model for this analysis.

For this analysis, it is assumed that radon daughter backgrounds are uniform throughout the TPC in the mixed state, with the more turbulent flow dominating any charge dynamics. The distribution of ^{214}Pb was inferred using the observed position of decays of its progenitors, ^{222}Rn and ^{218}Po, and that of its daughters, ^{214}Bi and ^{214}Po, which help inform charged and neutral flow as their pairings have distinct charge fractions. Once a flow model is constructed, the progenitor ^{218}Po alpha decay sites can be extrapolated to construct a position map of expected ^{214}Pb decay sites. This procedure was carried out for WS22 in Ref. [123] and details on its application for WS24 can be found in the radon backgrounds analysis note Ref.[170]. The radon tagging group used the updated simulations to produce deposit maps in corrected variables (r^2_{corr},z_{corr}) and (x^2_{corr},y_{corr}). These corrected variables are derived in `LZap` using the drift and field map, it is technically correct to just undo the procedure but due to limited statistics the author found that inputting these into `LZLAMA` to also apply smearing helped produce a (x_{obs}, y_{obs}, dt) map with no zero-bins, which can cause biases on observed data.

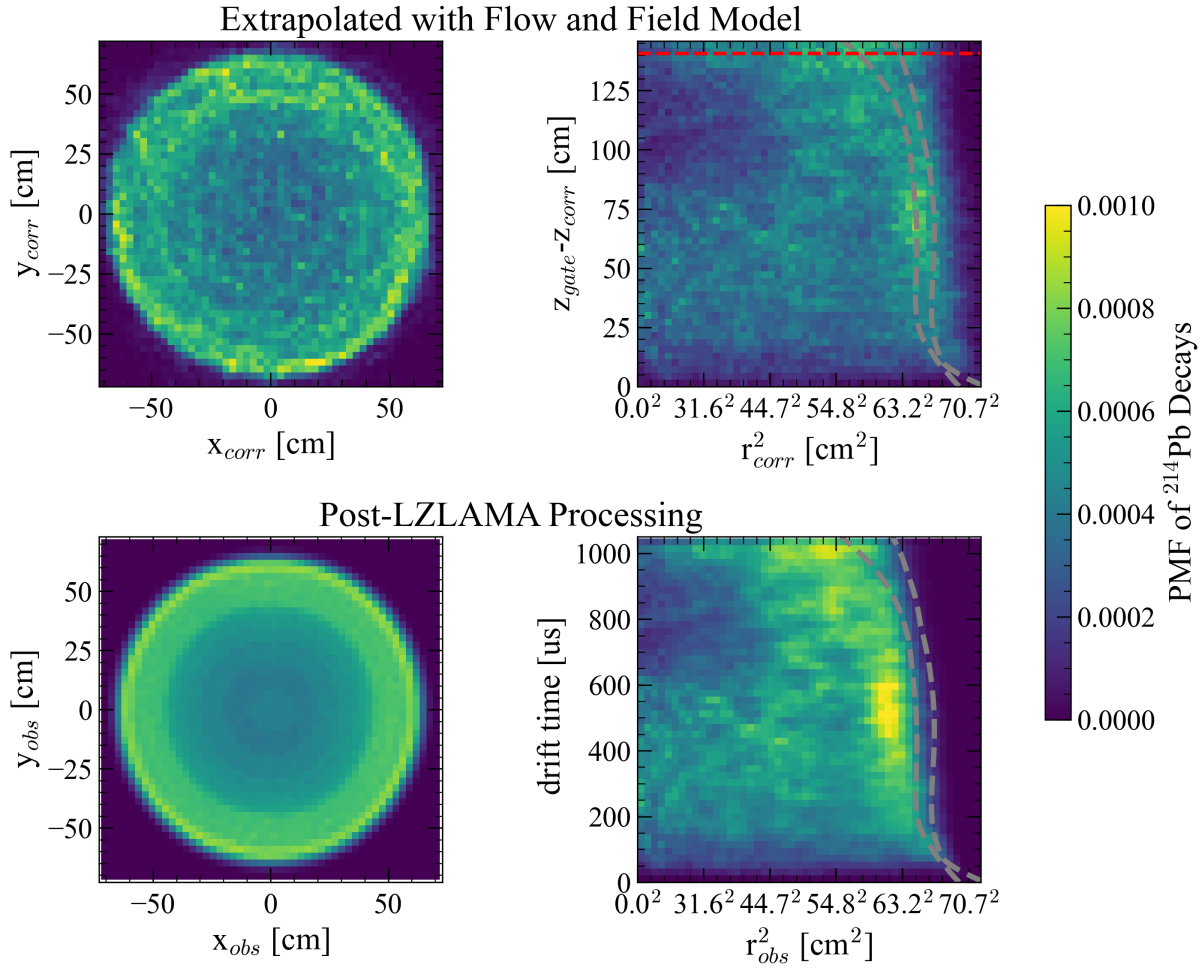


Figure 6.7: The discrete 2D Probability Mass Functions (PMF) of the ^{214}Pb decays extrapolating from ^{218}Po sites in corrected position, where z_{gate} is the maximum z position a SS can be detected in the TPC (top). Those positions after being processed through `LZLAMA` (bottom). Unfortunately, `LZLAMA` only supports (r^2_{true},z_{true}) inputs, so the 2D distribution on the bottom is more uniform than the top. Grey regions indicate the minimum and maximum FV boundaries, these are plotted in corrected co-ordinates with a linear approximation for drift time and assuming r_{obs}=r_{corr}. A red marker is where files were incomplete and the author had to copy the bins above, but the population is low in this region.

The final ^{214}Pb for this analysis is shown in Figure 6.7, the author notes that smearing and corrections in `LZLAMA` may be 'double-counting' shifting the expected distribution inward. Preliminary studies of ^{218}Po showed that it has a similar position distribution to its daughter ^{214}Pb, especially to the fidelity of the observed data, which is expected as it has a relatively short half-life. However, it is sub-dominant and assumed to be uniform in this study. For ^{212}Pb the progeny isotopes have significantly shorter half-lives and ^{212}Pb itself has a longer half-life than ^{214}Pb. This can enter the active volume and contribute to the background but is difficult to tag. There is no reason to assume this decay is uniformly distributed in the TPC or for it to follow the ^{214}Pb distribution, but without further input from a specialised analysis, it is

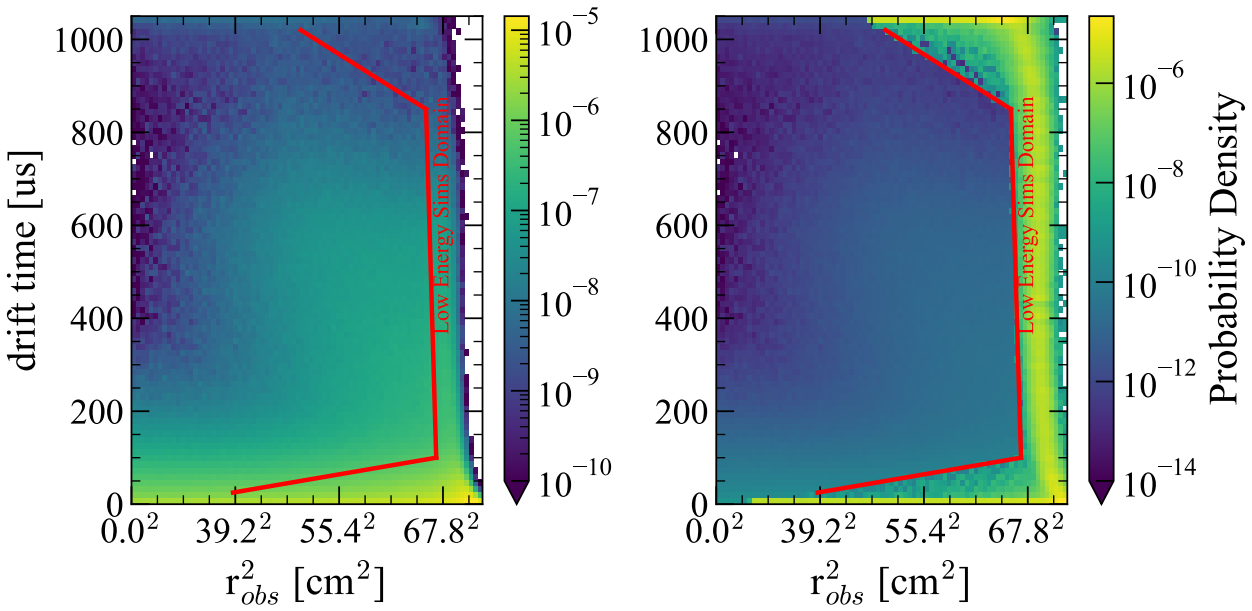


Figure 6.8: High energy simulations of detector ERs between ¿1 MeV (left) and low energy simulations (right). Within the volume marked in red, the high energy position map is used and normalised to the low energy counts in this region.

assumed that they are uniform.

The detector NR and ERs are composite distributions of neutrons and gammas from radio-isotopes in the detector components. To model them, many separate components are modelled as sources of each radio-isotope and propagated through the detector geometry in `BACCARAT`. The radio-assays of these components and background spectra fits carried out in the WS22 backgrounds paper Ref.[123] give the expected activity of each prominent isotope and component. Due to the large number of component and isotope pairings, the processing of these in `BACCARAT` and `LZLAMA` is carried out in large productions. The author used the WS24 production alongside an additional post-WS24 production for another near-identical analysis to boost statistics. This section covers the processing of these post-`LZLAMA` and `BGSkimmer` simulations to produce accurate position distributions for this analysis.

Despite larger simulations than in Ref. [111], the number of detector ER deposits in the FV were not sufficient to produce a complete PDF. For detector ERs, as higher-energy Gamma rays generally penetrate deeper in the LXe, we can produce a conservative background distribution in the ROI by using higher-energy deposits than the ROI. For deposits without a veto tag, an inner volume is chosen, within that volume the higher energy distribution is used and renormalised to the counts within that volume from the low-energy simulations. This procedure is illustrated and the results shown in Figure 6.8. This conservative choice for the ROI is not sufficient for the veto sideband, as the detector ER component dominates and this significant

modelling error prevents a reasonable fit. The raw simulations were not sufficient either, as low-to-zero value bins from fluctuations in otherwise high-density areas cause miss-attribution of events to other sources in the likelihood fit. These fits outright fail if no other source can describe that event. To resolve this, Gaussian Kernel Density Estimation (KDE) was used to smooth the histogram such that all bins had a defined value. The energy spectrum was derived for deposits up to 30 keV, as in the WS. First iterations gave poor fits to data, investigating this, it was observed that the rudimentary energy cut left position distributions contaminated with events that scatter within charge-insensitive volumes, such as near the electrodes and within the RFR, or by merged MSs. The former can clearly bias the position distribution to low and high drift times so should be removed, as shown in Figure 6.9.

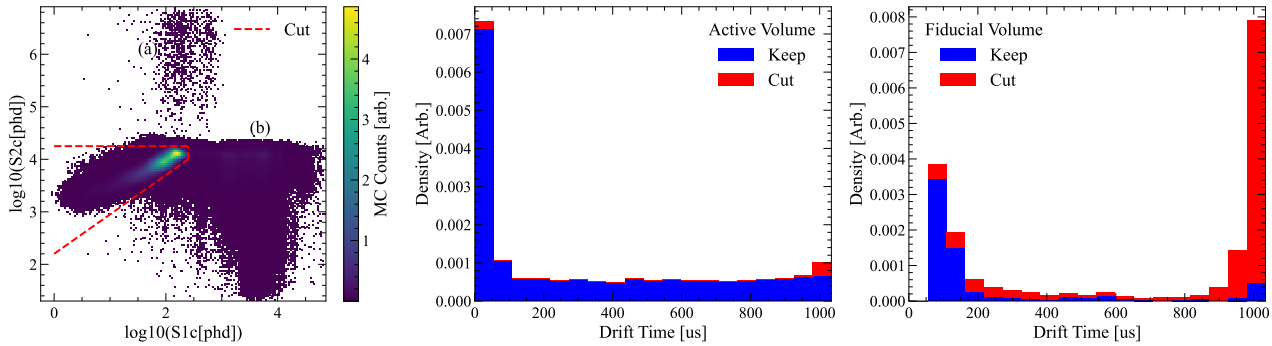


Figure 6.9: Cutting population of multiple-scintillation single-ionisation events (b) and merged MSs (b) that appear in the simulations (left). Impact of the cut over the total volume (middle). Impact of the cut within fiducial volume (right).

The statistics were still insufficient, so to resolve this in a biased manner the KDE factor, which governs how broad the smearing is, was adjusted until a qualitatively good fit to the observed veto sideband data was achieved. The results from the template fits in Table 5.3 were used to estimate the contribution from other, spatially flat, backgrounds and the gamma background to compare to data, showing reasonable agreement as seen in Figure 6.10. The low statistics also prevented a meaningful check of the spatial dependence of the spectrum; it is likely that it varies spatially but remaining consistent with the template analysis a flat-spectrum is chosen.

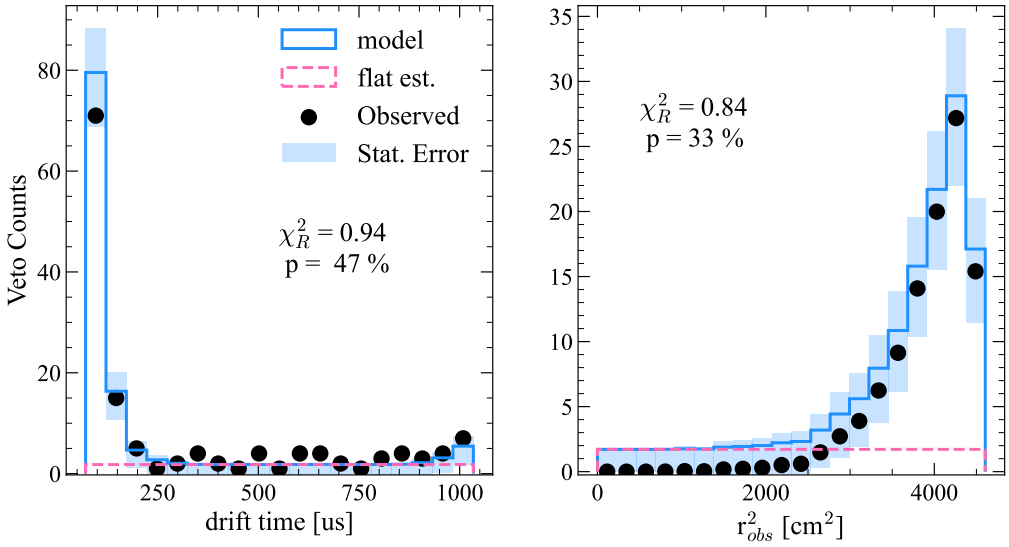


Figure 6.10: Comparing the KDE estimation to veto sideband with a flat background from template based fits.

The mean free path in LXe is greater for neutrons than gammas so there were enough detector NRs events surviving FV cuts to produce a smooth PDF for most of the detector except the centre. To resolve this, a Gaussian KDE was used to produce a smooth estimate of the PDF in (x_{obs},y_{obs},drift time), but this did not agree with the original simulation near the data wall as KDEs struggle with boundaries. Therefore, a cylinder in the centre of the TPC was chosen, outside it the original simulation would be used and inside it the KDE model renormalised to the counts from the original simulation was used. To ensure a smooth PDF for even small bins direct evaluation of the KDE at bin-centres was used rather than resampling.

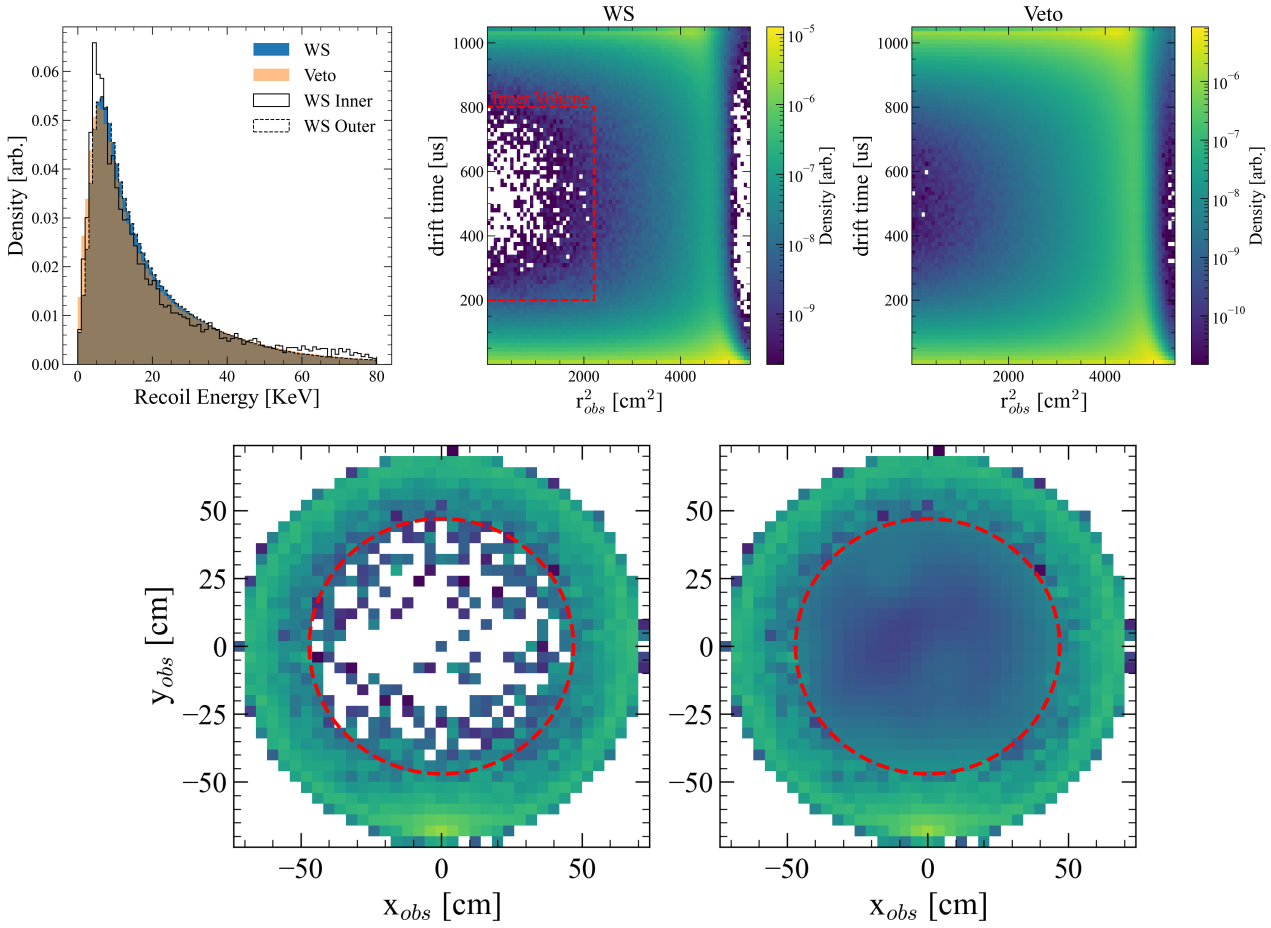


Figure 6.11: Combined simulations of the Detector NR showing change in spectra with an inner-volume cut and between Veto and WS (top). The impact of the veto tag and an inner radial cut indicates how the spectrum may change (top,left). The inner volume used for that check is indicated in the simulated data for the WS ROI (top,middle) and the simulation is also shown for the Veto sideband (top,right). An example of before (bottom,left) and after (bottom,right) the KDE smearing for a slice of the WS ROI over 450 < drift time < 550.

For both Detector ER and NRs, the author notes that in the future a parameter to control the level of extrapolation should be implemented. A conservative choice for discovery may artificially enhance the observed limits by overestimating the overlap of background with signal. As these sources are subdominant, this was left for future analysis. For detector ERs, improved simulations are required, as the outcomes relevant to the WS are exceedingly rare, event biasing techniques could be used. These tools are available in GEANT4, summarised in Ref.[171], but would require dedicated efforts to integrate into `BACCARAT`.

The activation products, ^{127}Xe and ^{125}Xe, position distributions are mainly determined by whether the prompt gamma escapes without forming a MS, allowing the event to pass the SS requirement. This can only occur near the outer-edges of the detector. These sources have the same atomic structure, producing identical EC deposits in the TPC and so are combined in the template analyses. However, they leave their daughter iodine nuclei in different excited states

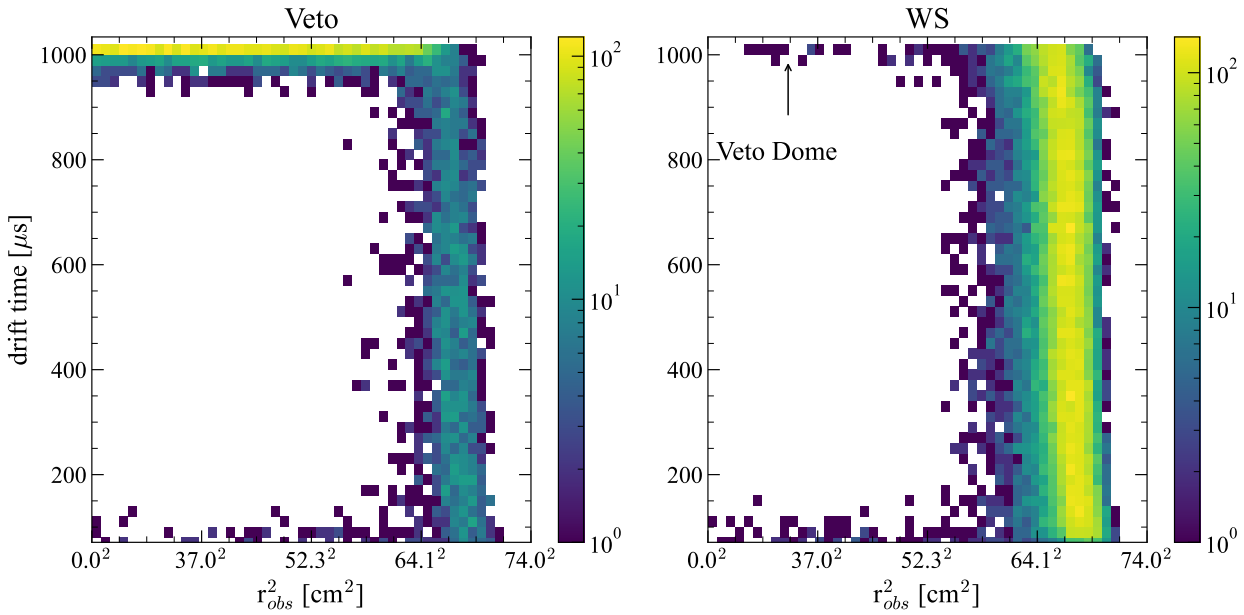


Figure 6.12: Position distribution of the xenon activation isotopes with and without the veto cut applied.

producing distinct prompt gamma rays, which may yield different position distributions. To produce their distributions, the author used the same method as Dr Robert James in Ref.[111] and constructed a new `BACCARAT` generator for the ^{125}I excitations shown in Section 2.3.1. For this analysis, ^{127}I/^{125}I excitations were simulated in a 15cm thick disk at the top and bottom of the detector and a 15cm thick sheath from the physical wall of the TPC. The likelihood of any events further inside the TPC than these volumes to escape the liquid Xenon without multiply scattering is negligible. As processing enough deposits through `LZLAMA` to produce a smooth PDF would not be feasible, the author removed any deposit with energy greater than 30 keV, guaranteeing that if the deposited gamma merged with the EC deposit, it would be outside of the energy range that could influence the ROI. The resultant spatial distributions were identical, so they were summed together. There was no attempt to smooth these distributions as they are negligible sources in the WS ROI and subdominant to detector ERs in the veto, therefore unlikely to cause any optimiser issues. It can be seen in Figure 6.12 that the position is highly localised to the edges of the TPC. Also, the veto preferentially tags events at high drift times near the bottom of the TPC, as this is where the skin-dome is present. Figure 6.12 could inform future implementation of semi-naked ^{214}Pb as the ^{214}Bi produces similar gammas energies as the ^{127}I/^{125}I. The fact that there is no difference between the former implies that the semi-naked ^{214}Pb position distribution in the veto would also be identical. The impact of the veto tag likely dominates the distribution of the ^{214}Pb isotopes from flow and charge dynamics.

6.3 Comparing flameNEST Simulations to Template Models

In order to simplify their use, the spectra for flameNEST were normalised to one, instead of their predicted activity as in LZLAMA. To confirm the simulation capabilities of flameNEST the author obtained the normalisation of the input energy spectra for a subset of sources in order to derive their total counts from,

$$\nu = \frac{N_{survive}}{N_{sim}} MT \int_0^{E_{max}} R_0(E) dE$$

where the first term is the survival fraction in flameNEST simulations, the second and third are the active mass and livetime, which constitutes the exposure. The final term is the integral of the input rate vs recoil energy in evts/tonne/year, this is often calculated using some isotope activity in the case of ^{214}Pb, or some flux and scattering calculation in the case of WIMPs, see Section 1.5. In Table 6.2 the rates predicted from input spectra into flameNEST and those from the Template analysis chain are compared. If the observed position map is correct, uniform sources should match in predicted counts. There is a discrepancy of approximately 1%. This may be due to the choice of energy bounds, which are slightly tighter in this flameNEST configuration, but as it is true for the mono-energetic ^{124}Xe spectra it is likely an impact of the position map. Still, this is well below the uncertainty of any of the source rates.

Model	^{124}Xe	^{136}Xe	^{214}Pb Mixed	^{214}Pb Isolated
Type	Mono-energetic	Spectrum		+ unique spatial
flameNEST w/spectra	15.54	44.89	135.96	706
Predicted Counts	15.79	45.33	134.45	606

Table 6.2: Table comparing the predicted counts for WS24 from flameNEST and those from LZLAMA and BGSkimmer. Both packages simulate necessary RQs and the acceptance of the ROI, FV, and data quality cuts. The active mass of the LXe target is 7 tonnes and the WS24 live time is 220 days. The ^{214}Pb Isolated is significantly higher due to greater FV acceptance from the position map.

Interestingly, the ^{214}Pb Isolated rates are over-predicted by flameNEST. The distribution is mapped for the entire active volume, but in the template analysis it was assumed to be uniform. It seems that this is showing the observed ^{214}Pb decays have a higher acceptance in the FV as they are more likely to occur in the FV than the mixed state.

The (cS1,log$_{10}$cS2) [phd] templates for WS24 are re-derived using the simulate function in flameNEST for each model using uniform observed position maps, with the exception of Detector NRs, allowing for direct comparison. The Jensen-Shannon Divergence (JSD) is often used as a comparison metric of two precise probability distributions and is bound between 0 and 1, being total agreement and total disagreement, respectively [172]. The threshold should be

such that changes between histograms can be greater than the random variance from MC and we must have sufficient simulation size to be more sensitive than the experiment. Generating multiple templates of a flat ER with $N_{MC} = 10^6$ and comparing to themselves gave a JSD of 0.052 ± 0.004, showing the variance is low and that this is the lower bound on agreement. Scanning over g^1_{gas}, which the flat ER band shape is highly sensitive to, showed that this JSD value corresponded to $\pm 0.1\sigma$ in g^1_{gas}, sufficiently beyond our experimental sensitivity. Therefore, a JSD of 0.06 was chosen as the threshold to test the agreement of the template shape.

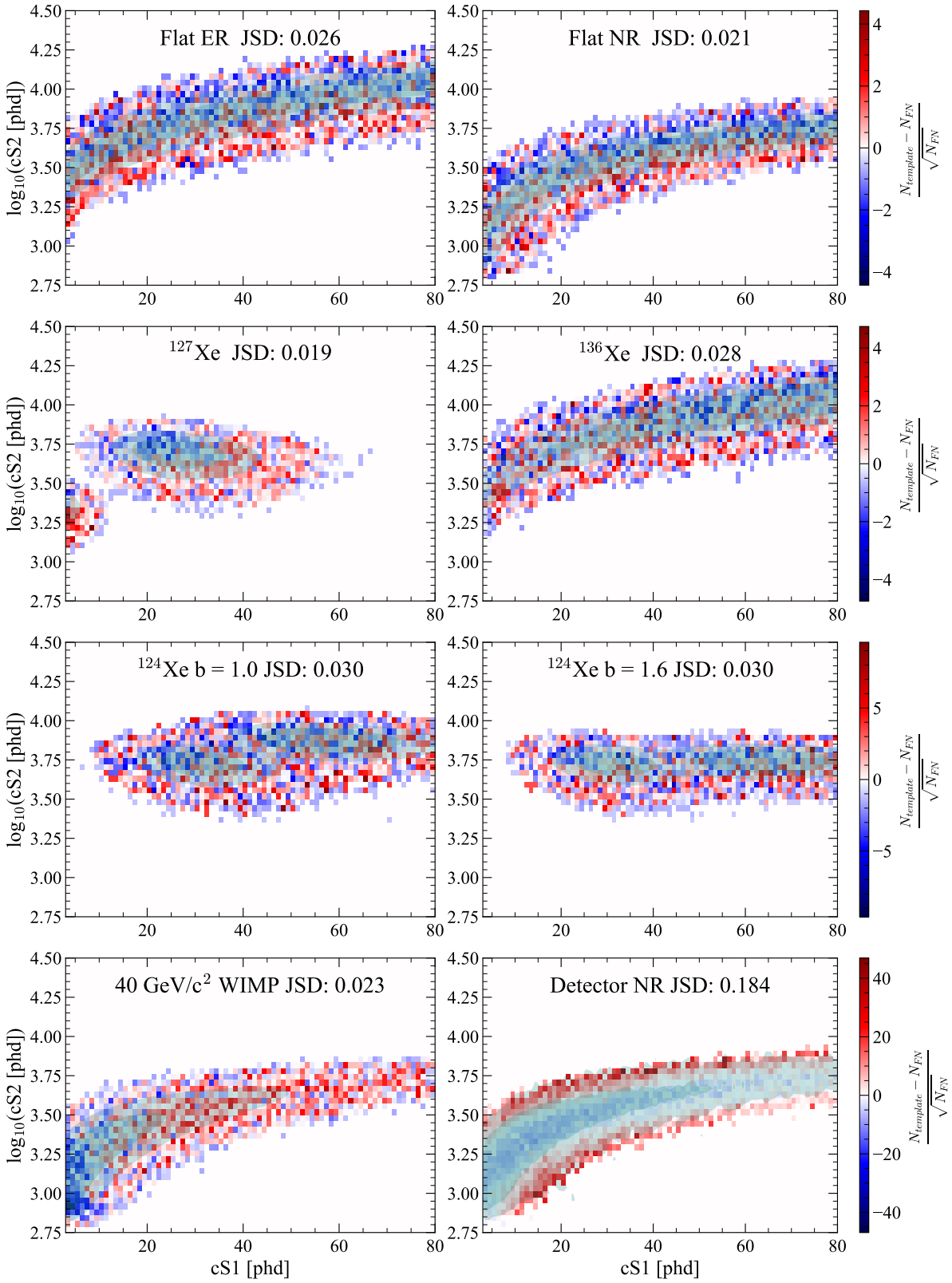


Figure 6.13: Comparing final background model from WS24 to `flameNEST` models configured by the author, with true position fixed to the centre of the detector. These are normalised PDFs, with the per-bin Poisson error in `flameNEST` propagated accordingly. The JSD metric does not rely on these uncertainties, it is just for inspection of any shape disagreement. The JSD is above the threshold chosen in text for Detector NRs and WIMPs have a visual disagreement near threshold.

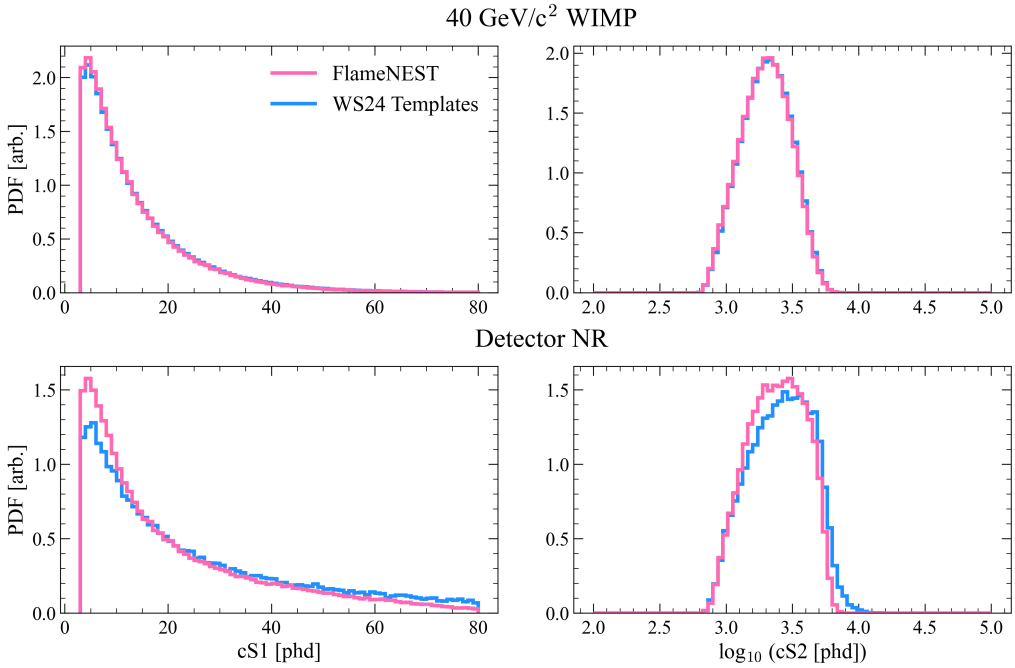


Figure 6.14: Projections of the WIMP and Detector NR models from Figure 6.13.

The result in Figure 6.13 shows some biases near the energy threshold for the WIMPs and Det NR is the only one in disagreement. As 2D histograms can be difficult to qualitatively understand, Figure 6.14 shows the 1D projections of two sources that disagree with templates. It is clear that this is only a problem with Detector NRs. For WIMPs, this is likely due to a different low-energy cut-off between the spectra in `LZLAMA` and `flameNEST`. For Detector NRs, the final models used the (cS1,log$_1$0cS2) distribution directly from the `LZLAMA+BACCARAT` simulation, which is why the `flameNEST` simulations were not set to be uniform. In contrast, the author derived the energy spectra from all events within the FV and a loose ROI cut. As seen in Figure 6.11 (top), the energy spectrum varies at low energies over the volume, which may require a (x$_o$*bs*,y$_o$*bs*,drift time, energy) histogram in the future. The simulations to derive the spectrum also have a coincidence cut applied, but whilst this should enhance the total acceptance, the resultant energy spectrum should peak less at lower energies, which is opposite to what is observed. To resolve this a reprocessing of all component `BACCARAT` outputs through `LZLAMA` without the coincidence cut would be required, which is not feasible or required for this work, as it is unlikely this bias will significantly impact the results.

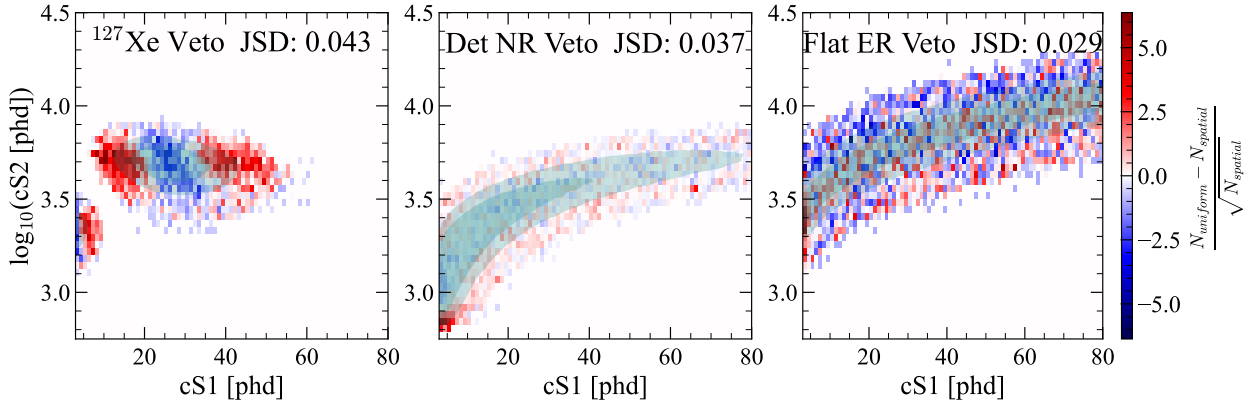


Figure 6.15: How the underlying spatial distribution of deposits changes the (cS1,log$_1$0cS2) when assuming the distributions in Section 6.2.2 with bin counts, $N_{spatial}$, compared to the derived observed uniform distribution, $N_{uniform}$.

To show the spatial dependency of (cS1,log$_1$0cS2), a subset of sources simulated in `flameNEST` with their derived spatial distribution are compared to those simulated with a uniform distribution in Figure 6.15. The largest impact is seen in the vetoed ^{127}Xe source, with an impact at small cS2 for NRs, whilst the Flat Veto ERs representing Detector ERs are virtually unaffected. Unlike in Ref.[145], the major factor is not electron lifetime, as changing it by two orders of magnitude smaller or larger does not impact the difference. Removing the LCEs reduces the JSD to 0.13 and the acceptances to 0.04. This implies that the coincidence and ROI cut, alongside the high drift bias of the vetoed ^{127}Xe, may be the largest impact of the position distributions.

6.4 Time Dependence

Whilst it is possible to model the time-dependence of S2c and S1c throughout WS24 in `flameNEST`, in order to remain consistent with the template analysis the spatial-temporal corrections to data are applied to flatten out this variation. Therefore, the model must be flat. The exception to this is the rates of two contaminant background sources from calibrations, Xe activation isotopes and $^{14}CH_3T$, have time-dependent rates. The Solar ER and WIMP counts have an annual variation of $\mathcal{O}(7\%)$ and $\mathcal{O}(5\%)$, respectively. Due to their low impact, these have been omitted for simplicity in this analysis, however they should be used in the future. The detector response does vary in time as the electron life-time changes over WS24, as seen in Figure 6.15, but as it is always many times the maximum drift time it is unlikely to impact the models, so is omitted here.

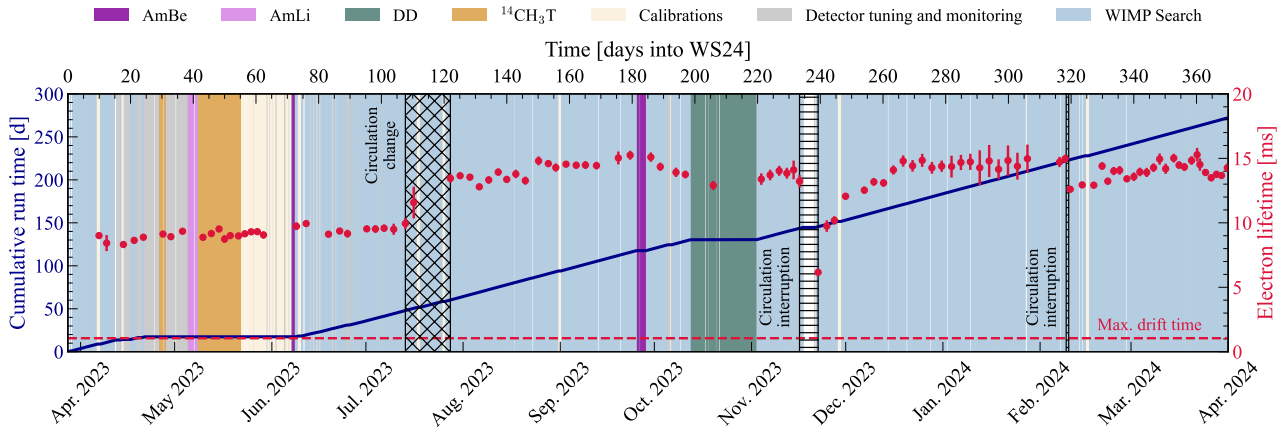


Figure 6.16: Plot showing the accumulated livetime of WS24 against calendar and days in. Shaded region shows the nature of data acquisition, for example when $^{14}CH_3T$ calibrations are taking place. Generated using code from Dr Aiham Al Musalhi, PDRA at University College London.

The removal half-life of methane, $^{14}CH_3T$, as it is removed by the getter was measured in LZ to be 3.154 $\pm$ 0.007 days. This uncertainty produces a negligible impact on the time-distribution during the mixed period, but alongside the initial activity measured contributes to the uncertainty on the counts. For simplicity, the derived counts for the mixed state were used as a parameter and the half-life was fixed. This means the start date of the decay may be taken as the first day of the mixed state.

There were three calibration campaigns that activated ^{125}Xe and ^{127}Xe during WS24, these were AmLi and AmBe in the mixed-state calibration periods and AmBe and DD during the isolated state neutron calibrations. To calculate their activity, Olivia Valentino selected double scatters from the prompt gammas in a similar analysis to that conducted in Ref.[124], cross-checked by Dr Tom Rushton's measurement using their peaks in the energy spectrum immediately following calibrations. The derived activity is plotted in Figure 6.17, as their half-lives are well known, the author made a simple qualitative fit for the exponential decays from after calibration periods using $t^{125}_{1/2} = 16.9$ hours and $t^{127}_{1/2} = 36.3$ days. The pre-calibration period during the mixed state has remnant ^{127}Xe contamination from pre-WS24 calibrations held sufficiently earlier that ^{125}Xe would have decayed. The analysers found that immediately post-calibration the data is contaminated with other activation products, as such the tails are used to constrain the fit. Due to its far shorter half-life, it is difficult to determine if there are contaminant activation products in the ^{125}Xe study, so this too represents an upper-limit on possible rate. As they produce identical EC decays, and their position distributions are identical, they are combined into a single source with their time-dependence being the sum of the two shown, weighted by their predicted counts. It is possible to float the proportions of the two components in the model, but this is left for future analysis.

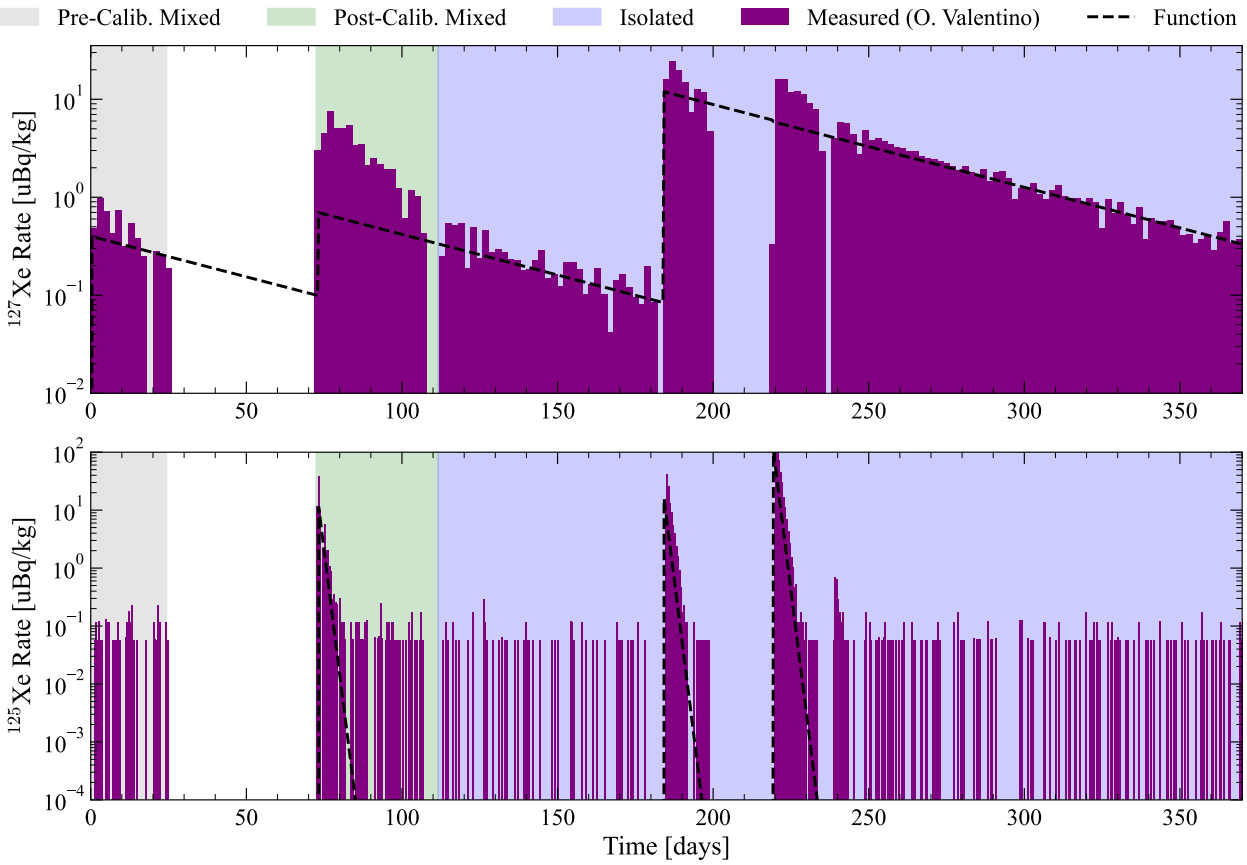


Figure 6.17: Activity of ^{127}Xe and ^{125}Xe plotted against days into WS24 derived by Olivia Valentino, PhD student at Imperial College London.

6.4.1 Including Live Time Losses

The buffer time for events in the DAQ, the application of cuts that remove periods of time with a high level of noise, and runs dedicated to calibrations means that the calculation of the time component of exposure to obtain expected background and signal counts is not straightforward. For this work, the livetime of interest is the physics livetime, that is the period of time the data-acquisition would be able to detect a WIMP signal. As this livetime acceptance is identical for all sources in the likelihood it will be a constant factor, thus is not necessary to include in the differential rate calculation on a given data set. To make this explicit, this assumption is fine so long as

$$\begin{aligned}\nu(\vec{\theta}) &= \int R(\vec{s},\vec{\theta})T(t)d^D\vec{s} \\ &= \int R(\vec{s}',\vec{\theta})d^{D-1}\vec{s}' \int T(t)r(t)dt\end{aligned}$$

where $r(t)$ is the time component of the model, like the decay of ^{127}Xe and $^{14}CH_3T$ rate. This only holds if the time dependence may be factorised out from the detector response, i.e is only in rate, and does not depend on any free parameters. For WS24, only rates depend on time so

this holds true, and the predicted counts from the backgrounds team already factor in this second integral via livetime calculations. In the future, it is likely better to multiply the livetime fraction by the temporal rate dependence of each source as a time dependence at the source level and avoid these restrictions on the model. This would not be difficult to implement but the author implemented the more limited method as, given the emphasis on spatial dependence and lack of temporal dependence of the detector response, this issue was not examined as fully as it might warrant at the time.

For comparing simulated to observed datasets, for example in the generation of toys for GOF or building a test-statistic distribution, it is necessary to include the livetime cut. This is because the time component that has been factored out is a constant in fitting parameters but function of the observables, namely time, of the data and will introduce bias in these distributions. The author extracted the livetime histogram and converted it to livetime-fraction. Using this, an additional Bernoulli trial could be applied to simulated data. The livetime fraction vs time is shown in Figure 6.18, alongside the impact on a uniform source and exponential decay source like Tritium. In simulating GOF toys, and for a future inference routine, the acceptance of this cut must be factored into the rates. In effect, this just scales up predicted counts in the simulate functions such that the appropriate counts are produced when the cut is applied. This may be applied at the total rate estimator step in the future to consistently predict counts in `flameNEST`.

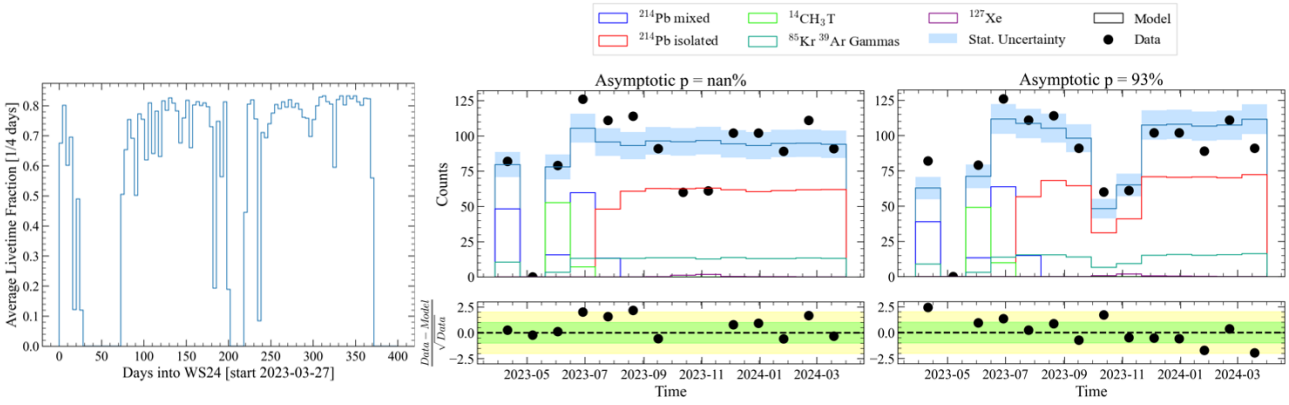


Figure 6.18: Plot of the averaged livetime fraction over WS24 (left). The expected model, i.e pre-fitting, before (middle) and after (right) applying the Bernoulli trial and scaling. There is little statistical power to discern a good-quality fit.

6.4.2 The Combined Likelihood for WS24 in `flameNEST`

The author initially aimed to use only two combined likelihoods, the veto and the WS ROI, and the various sections within this, e.g mixed and isolated, would be handled by the time dependence of each source. Due to various bugs, namely one that would set the space and time dependence of all events in a batch to that of the first event, this did not appear to work

in the first application. Instead the author constructed a Pre- and Post-Injection likelihood for the mixed state which are before and after the calibrations period, respectively, and an Isolated state likelihood. The mixed state likelihoods do not have the isolated ^{214}Pb source in them, and vice versa. Both Pre- and Post-Injection likelihoods contain ^{125}Xe+^{127}Xe, but only the Post-Injection likelihood contains the $^{14}CH_3T$ source. The issues in batching were resolved by the time the veto likelihood was implemented, and so those periods were included with an event time mask. This means that, in total, there are four likelihoods. In the future, returning to only the WS ROI and Veto likelihoods would be ideal for simplicity of the model. Radon tagging was ignored in this analysis, more work is required to impose the relevant exposure loss as a function of space and time on each of the sources and the relevant working groups were not able to provide this at the time of writing this thesis.

The input parameters for this likelihood are identical to those in Table 5.2, except that counts must now be transformed to rate multipliers via $\frac{C_j}{\nu_j(\vec{\theta}_0)}$ for source j. The interpretation of the rate multiplier changes somewhat depending on the source, but generally it is the predicted counts with the 6D observable based acceptance unfolded. This allows the impact of the rate of a source varying within uncertainty to be unfolded from the acceptance changing with detector parameters.

6.5 Multi-Dimensional Goodness of Fit

With the likelihood constructed, the author implemented a fit to observed salted data set to see if there were any features this preliminary model could identify. In order to validate the model, 2D and 1D GOF tests were developed. It should be re-iterated that the inputs are still preliminary, this section is to show the potential of 6D inference on LZ data with shape varying parameters, not to make definitive claims as more validation is required.

There are some disagreements with the template fit, namely in the detector NR counts and in ^{127}Xe +^{125}Xe, which can all be understood in the context of the updated model with more observables. The ^{127}Xe +^{125}Xe source was consistently fitting to low values of approximately 5 events. The author assumed that the ^{125}Xe component could be biasing the time distribution, so removed it. This had no effect on other parameters, but reduced the disagreement between predicted and best fit counts for ^{127}Xe. In the `flameNEST` fit, the detector NRs show a non-zero best fit as a result of the improved discrimination, which is discussed in this section. Note that, as the two works occurred at different times there is some slight change in naming conventions, particularly ^{212}Pb+^{218}Po is now Radon β and ^{85}Kr ^{39}Ar Gammas is now just labelled ^{85}Kr.

Parameter	Predicted	Best fit
Solar ν ER Counts	102.3 ± 6.1	104.9 ± 6.1
^{85}Kr Counts	162 ± 22	197 ± 21
^{136}Xe Counts	55.7 ± 8.4	60.2 ± 8.2
^{127}Xe Counts	2.15 ± 0.45	1.53 ± 0.43
^{124}Xe Counts	19.4 ± 3.9	20.9 ± 3.6
AC Counts	2.75 ± 0.64	2.67 ± 0.61
^{214}Pb Mixed Counts	137 ± 18	160 ± 12
Radon β Mixed Counts	22.2 ± 4.8	23.0 ± 4.7
$^{14}CH_3T$ Counts	58.3 ± 3.3	58.3 ± 3.1
^{214}Pb Isolated Counts	606 ± 86	536 ± 34
Radon β Isolated Counts	40.5 ± 5.8	40.1 ± 5.7
10 TeV/c^2 WIMP Counts		$4.7^{+4.2}_{-3.0}$
Detector NR Counts		$0.15^{+0.25}_{-0.12}$
^{124}Xe Box Parameter, b		$1.43^{+0.13}_{-0.12}$
g_1	0.1122 ± 0.0020	0.1123 ± 0.0013
g_1^{gas}	0.0764 ± 0.0020	0.0758 ± 0.0005
Total	1208.68 ± 108.03	1210 ± 55
Observed		1226
Veto		
None Unique Counts	35.6 ± 2.7	35.7 ± 1.3
^{127}Xe Counts	7.6 ± 1.6	5.5 ± 1.5
Veto Efficiency, ϵ_{veto}	0.92 ± 0.04	0.92 ± 0.04
Detector NR Counts		$1.81^{+2.32}_{-0.75}$
Flat ER Veto Counts	140 ± 42	85.3 ± 9.4
Veto ER β-γ Weight, ω_γ		$0.025^{+0.073}_{-0.025}$
Total	184.03 ± 42.03	129.4 ± 9.9
Observed		138

Table 6.3: Resulting fit counts, with best fit uncertainties combined appropriately using the correlation matrix from the fit. Uncertainty on parameters is not propagated to the individual uncertainties on predicted counts. As the template based equivalent in Table 5.3, the None Unique Counts is just the sum of those that scale with dead-time to the veto. The uncertainties for those have been combined in quadrature, but the combination for the total was handled correctly with the correlation matrix. The asymmetric uncertainties in parameters are derived by profiling over each parameter. For asymmetric uncertainties on counts the rate-multiplier is profiled over and the conditional best fit of the parameters of the total rate estimators are used in obtaining the uncertainties in terms of counts. As the Detector NR is non-zero now, the resultant counts from this procedure are reported in the WS ROI part of the table as well.

In table 6.3 the best fit counts and non-rate parameters for WS24 in `flameNEST` are shown and the pulls in terms of parameters of the model are shown in Figure 6.19. The detector NR counts in the veto are non-zero but consistent with the constraint from the salted template based fit of $0^{+2.6}$. The total counts for the WS24 ROI fit to nearly identically values as the template based of 1208±33 but the Veto sideband fits out lower than the 141±12. The ^{127}Xe strongly floats down now, and did so before the removal of the ^{125}Xe model, which may contribute to some, but not all, of the discrepancy. The primary driver of the discrepancy in veto counts is

the flat ER source, which may be due to a modelling error in the position distribution. The ^{214}Pb isolated counts are smaller than the the 576^{+32}_{-31} but the difference has been absorbed predominantly by ^{85}Kr, formerly 162±21, suggesting that this is an effect of the ^{214}Pb isolated position distribution. A final note is that the $\gamma - \beta$ weight for the detector ERs shows that their microphysics are nearly totally identical to betas.

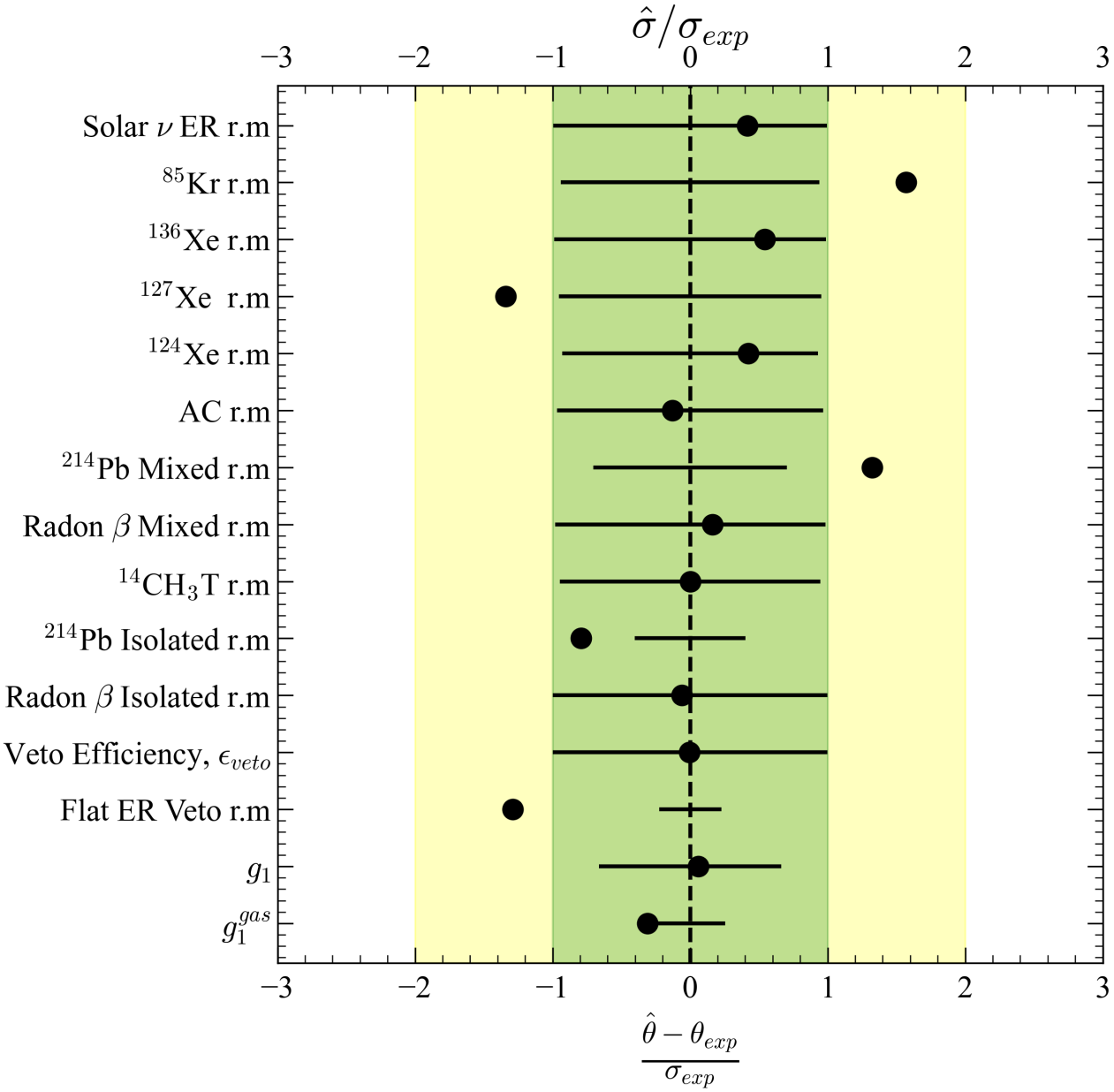


Figure 6.19: Pull plot of all the parameters in the likelihood model for WS24. Not that unlike in 5.9 these parameters are rate multipliers. These were derived using symmetrical uncertainties derived from the hessian of the likelihood function.

Using `flameNEST`'s simulation function, and the discussed modifications for livetime, the existing GOF framework from Section 5.1 can be used to produce 2D GOF tests comparing data to this model. There are, however, six dimensions now so three 2D projections are used,

(x_{obs},y_{obs}), (r^2_{obs},drift time), and (cS1, $\log_{10}$(cS2)). The results from each of the three WS ROI separate likelihoods are combined into a single model in each of the projections. As seen in Figure 6.20, all projections yield a reasonable p-value with the lowest being the radial and vertical positions in the Veto, which is to be expected with the heavy approximations made in Section 6.2.2. In fact, in that space strong over-fluctuation of data may be due to the omission of the $\sim$ 10 predicted semi-naked ^{214}Pb decays in the model.

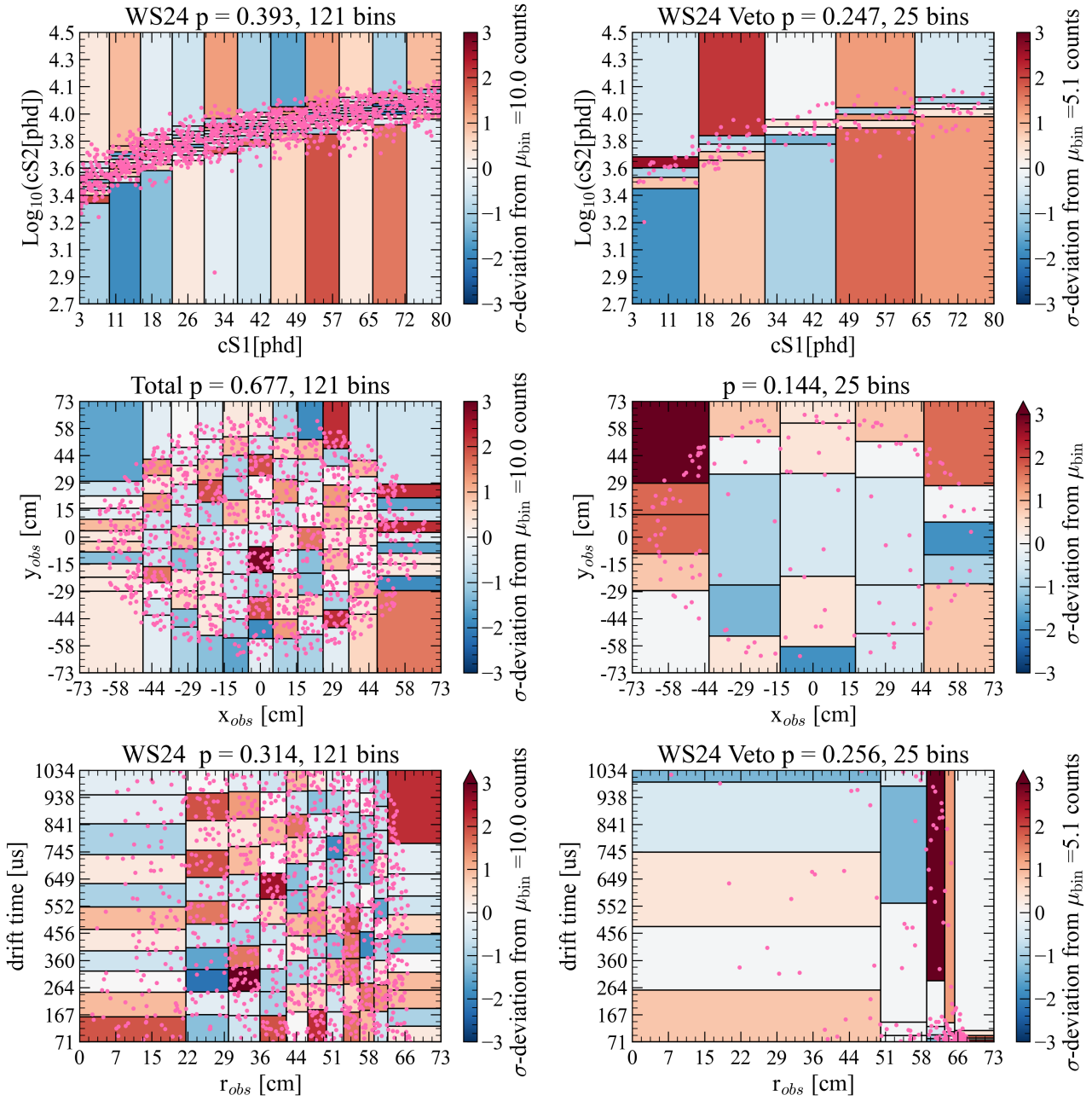


Figure 6.20: The 2D GOF as described in Section 5.1.2 but utilising templates for each space produced using `flameNEST` simulations.

As it is central to the assumptions made in Section 6.2.1, and the ^{214}Pb isolated distribution being a preliminary version, and the heavy assumptions made for the flat ER veto source, the

author included 1D GOF in the observed radial position. In Figure 6.21 it can be seen that whilst the p-value is good there is a systematic error in the position model at the outermost radii. In the WS ROI this is likely an issue with the isolated ^{214}Pb distribution and in the veto recall from Figure 6.10 that the radial peak in the Flat Veto ERs has likely been miss-modelled, driven by the exclusion of ^{127}Xe in that estimate and an over-estimate of the flat component.

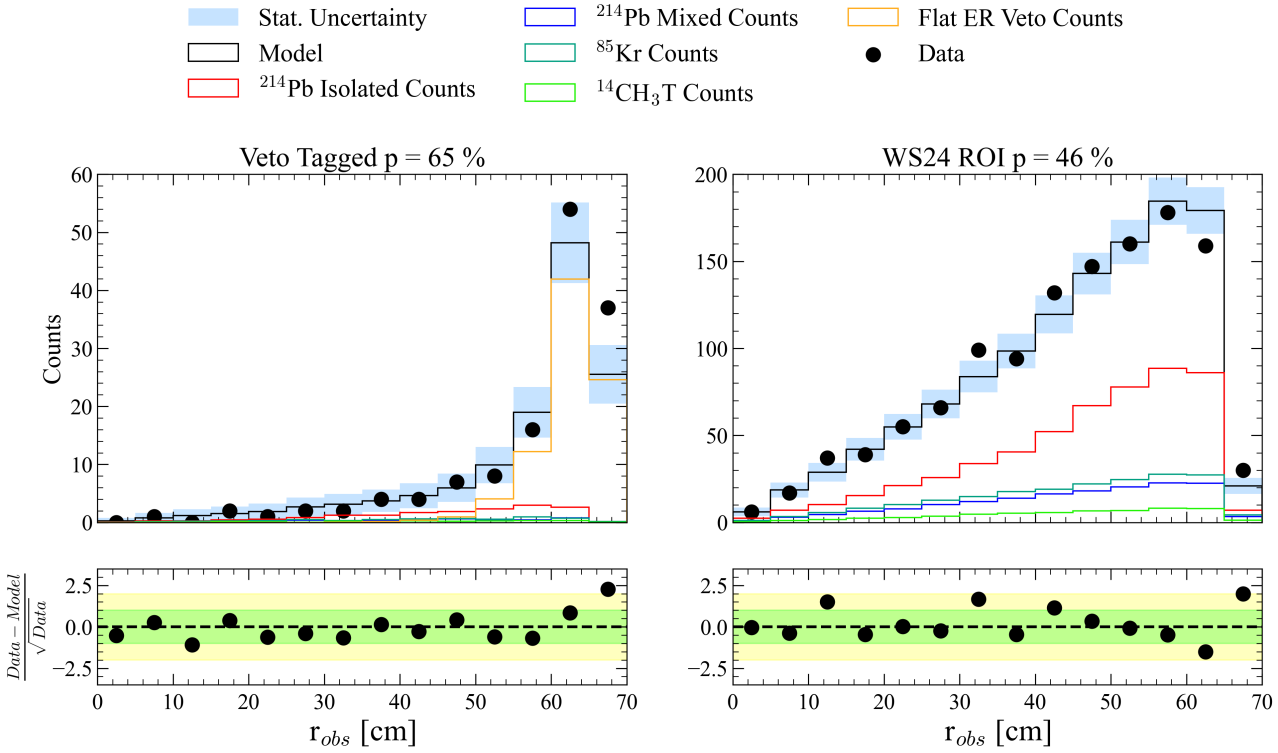


Figure 6.21: The 1D radial GOF using the asymptotic χ^2_m distribution where m is the number of bins.

The author also implemented a time-based GOF test in Figure 6.21. The post-fit time-based GOF is worse than the pre-fit example in Figure 6.18, as the fit is mostly driven by other factors, but it is still acceptable. Interestingly, it does seem that the post-injection period in the veto is over-fluctuating. Whilst this could be random, the residuals indicate a half-life of $\sim$ 1 month. It is possible that some additional unaccounted for activation sources contributed to the veto sideband, possibly the contaminants seen in Figure 6.17.

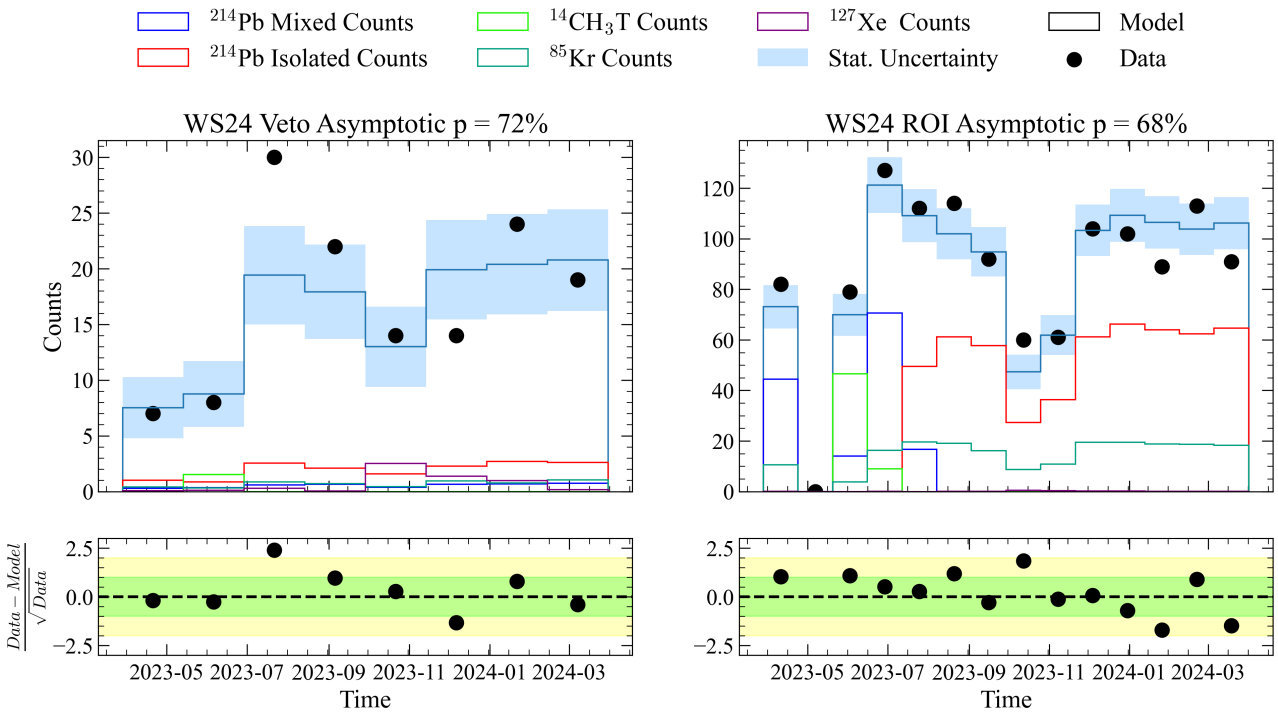


Figure 6.22: The 1D radial GOF using the asymptotic p-value calculation as in Figure 6.21.

To explore the detector NR events in the veto the author produced the so-called "pie-plots". Originally used by the XENON collaboration in Ref.[163], these show the fractional differential rate of each source evaluated at each event. Explicitly, the proportion of the pie attributed to each source is,

$$p_j = \frac{\nu^j_{r.m} R_j(\vec{s})}{\sum_k \nu^k_{r.m} R_k(\vec{s})} \tag{6.1}$$

for the j-th source. The size of the pies is scaled by the value of the differential rate of a chosen source to emphasise, in Figure 6.23 (top) that is the detector NR source. These pies help to track each event through a more than 2D observable space, and can be helpful when contours obscure the structure of the model. As can be seen in S1-S2 space there are three primary candidates that have strong overlap with the Detector NR model, these have event IDs of; 211225 at 5 phd, 257237 at 45 phd, and 158470 at 55 phd. There are two events at 15 phd of 193525 and 253840, but the former is highly likely to be a gamma event given its (r_{obs}, drift time) coordinates. The position of the veto signals for these events can be seen in Figure 6.23 (bottom), and whilst for event 211225 it is incredibly unlikely for a neutron to traverse the TPC like this, the probability of an accidental OD signal occurring in coincidence is more unlikely [173]. The introduction of time and spatial discrimination has played a strong role in identifying these events. For 253840 the (r_{obs}, drift time) is the greatest factor. For 257237, it does overlap with the skin-dome component of the ^{127}Xe in (r_{obs}, drift time) but is far from activation periods in event time. Whilst this is still only a proof of principle analysis, it is encouraging to see that extra dimensions have improved inference on the observed data.

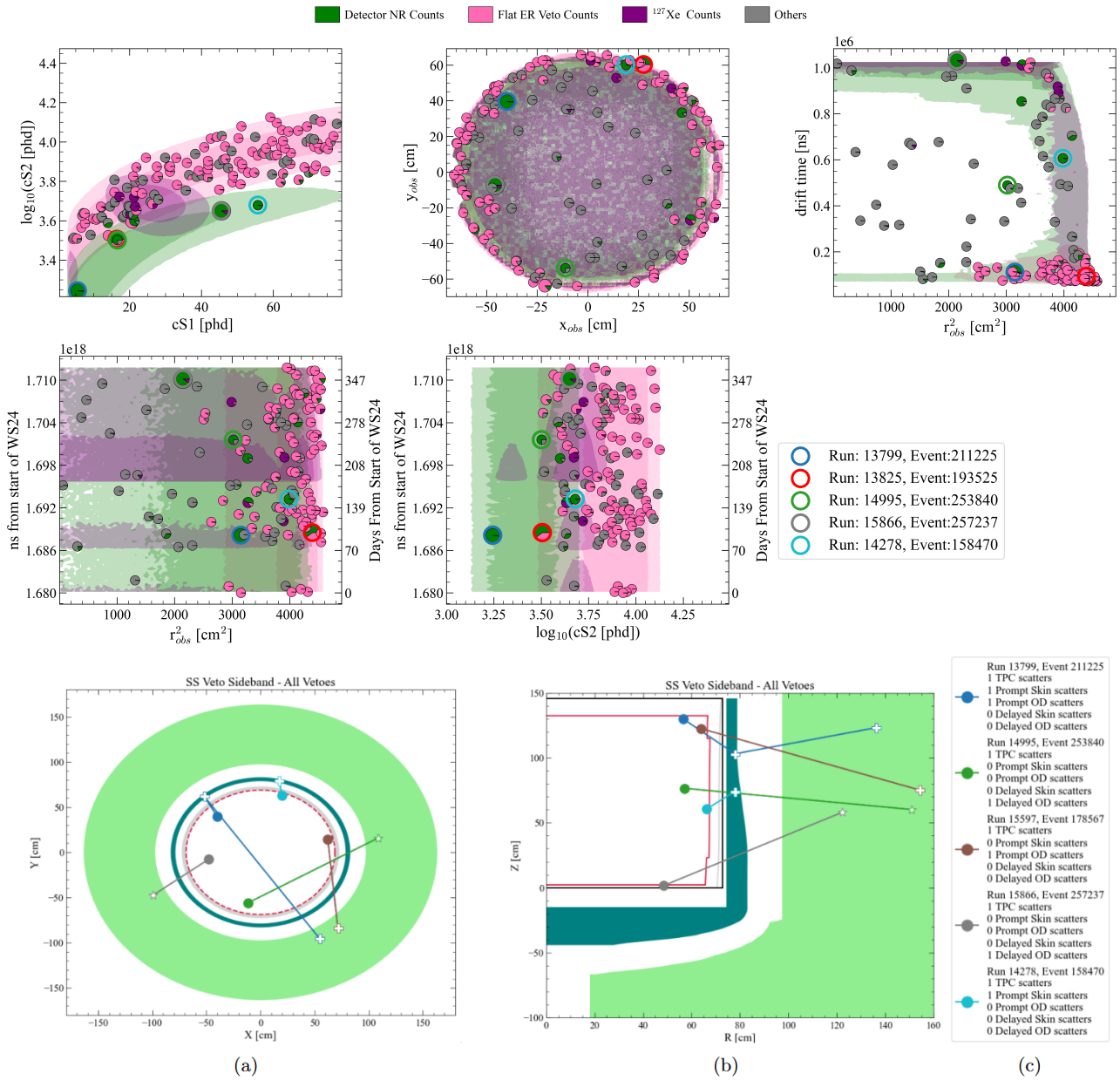


Figure 6.23: Pie plots of the veto sideband for this fit plotting in various key observables, contours of individual models are also shown (top). Diagram of location of scatters of the highlighted events in the various detectors, taken directly from the SR3 Neutron Efficiencies and Rates technical note in Ref.[173] (bottom). The radial position in the OD is a randomised value for visualisation purposes.

A note on the statistics here, one is often tempted to apply the Holm-Bonferroni test to combine p-values of multiple tests [174]. This takes the limiting value of an inequality, essentially assuming independence. This applies to the periods discussed in the template analysis, which are separated in time and therefore independent events, but is extremely over-conservative for these highly dependent tests. Therefore, it should be ignored in these tests and any failure is considered cause to scrutinise the model. Fortunately, none of them fail.

6.6 Technical Considerations

Among many bug fixes and model changes, the author developed three more technical aspects of `flameNEST` for this work that should be mentioned here for future analysers. These are source groupings using full differential rate calculations, polynomial 1D rate estimators for separable parameters, and adjustment of the optimisers conditions. These are introduced here and tested in Section 6.7.

6.6.1 Re-Introducing Source Groups

The author re-introduced a method used in Ref.[111] to calculate the most of the differential rate once for many sources that share the same microphysics, i.e ERs or NRs, called source grouping. However, in this work the author implemented it for the full tensor calculation rather than over cached values of the differential rate, allowing for shape-varying parameters to be included.

Source groups work by the fact that the calculation for the differential rate of different sources with the same microphysics, i.e all Beta ERs, only varies by the energy spectrum. Meaning, one can calculate the differential rate over a flat energy spectrum once, store the result as a tensor of differential rates and energy, then for each source rescale it by its energy spectrum that may depend on position or time. Due to the complexity of the central block (CB) in ref.[111] the results were saved once and rescaled later. This can be expressed as,

$$R^{ER/NR}_{source}(\vec{s},\vec{\theta}) = \sum_{E} \mathbf{P^{ER/NR}}(\vec{s},\vec{\theta}|E) R^{source}_{0}(\vec{s}|E) \tag{6.2}$$

where $\vec{s}$ is the observables that the detector response and the energy spectrum may depend. Previously, sources were implemented with pre-calculated differential rates, so the calculation in fitting was identical to template sources, just fetching the results. For a full tensor calculation, allowing parameters of the model to be floated, the author needed to modify the source group calculation to behave as a tensor source, and then create a custom likelihood function to manage the above calculation. A key aspect of this is that the granularity of the source groups energy spectra must be the same or finer than all the groups sources, and each source's energy spectra must be interpolated such that the rates are recalculated at every point in the group's energy spectrum.

The results of testing how the source group impacts likelihood calculations are shown in figure 6.24. There has been an error introduced in the gradients, especially for the shape varying parameters, but this appears to be random rather than a systematic bias. This also shows up in cross-terms of Hessians with a maximum of 5% variation, not shown here, indicating it is an error in the differential rate calculation. Further investigation showed that it is a result of missmatched energy bounds in the quanta tensors, and may be resolved by increasing the maximum

allowed energy dimension. The performance improvement induced in the WS24 likelihood is discussed in Section 6.7.

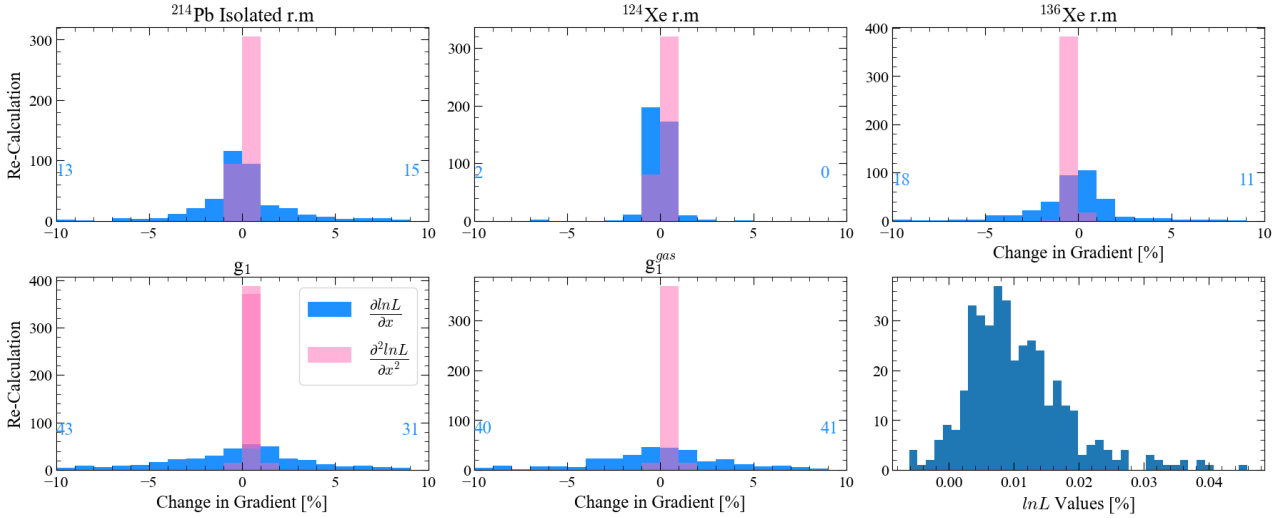


Figure 6.24: Comparing calculations between grouped and ungrouped likelihoods for two grouped sources, Isolated ^{214}Pb and ^{136}Xe, with an ungrouped nearby source, ^{124}Xe. Two shared shape varying parameters are floated to check any errors in the differential rate calculation. The Isolated ^{214}Pb was used in case there were any complications in the presence of a spatial source. A normal and grouped likelihood was used then for 400 toys, data was simulated from the normal likelihood then the value, gradient and Hessians were calculated on both at a guess. The log-likelihood values are precise, but the gradients and Hessians (not shown) do vary within 10%.

6.6.2 Rate Estimators

As found in Chapter 4, the linear interpolation of total rate estimators can lead to a bias as it impacts the first and second order derivatives of the likelihood, and therefore the optimiser. In fact, even if the function is linear, but with a large gradient, the discontinuity can cause bias. For this model, the author decided to use a set of one-dimensional 2nd order polynomials using the method of least squares. Some examples for a subset of the sources are shown in Figure 6.24. Note that the ^{214}Pb rate does display a weak dependence on g_1^{gas} which is ignored here.

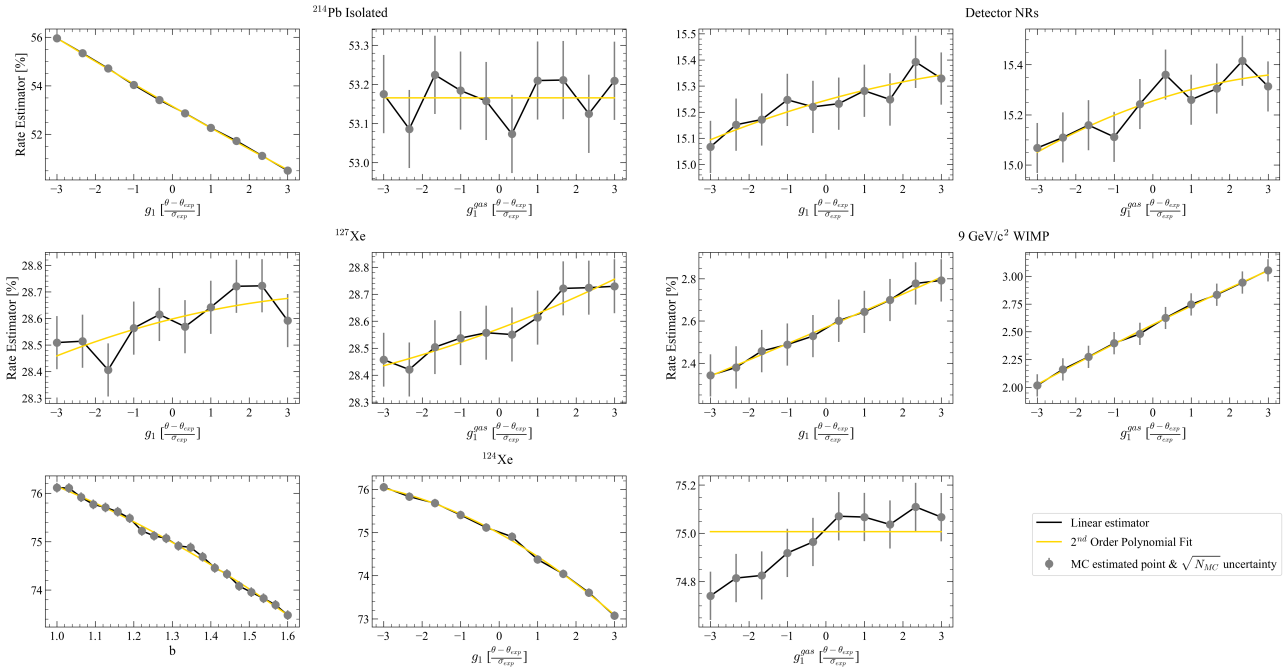


Figure 6.25: Example of the total rate estimators, as all source spectra are normalised to a total true rate of 1., this represents a fractional acceptance of the source. The Poisson uncertainty from MC simulation, $\sqrt{N_{MC}}$ where N_{MC} is the number of trials, is shown for the anchor points. The polynomial 1D fits used for this work are shown, where the change in rate is indiscernible due to the uncertainty on MC it is assumed to be flat.

The assumption that the impact on the total rate of each parameter considered is separable holds for most of the sources considered, but this should be tested for WIMPs and ^{124}Xe. As the WIMP source is near both the lower S1 and S2 thresholds, g_1 and g_1^{gas} can become correlated depending on its shape in those observables. This is certainly the case for b and g_1 in ^{124}Xe as the impact of the upper boundary S2 with g_1 will change with the value of b. We can see this correlation test by simulating a sparse grid and comparing the cross-product of the linear interpolators for each parameter, shown in Figure 6.26. The dependencies are weak, and impact on rate small enough, that this should not cause any bias in inference.

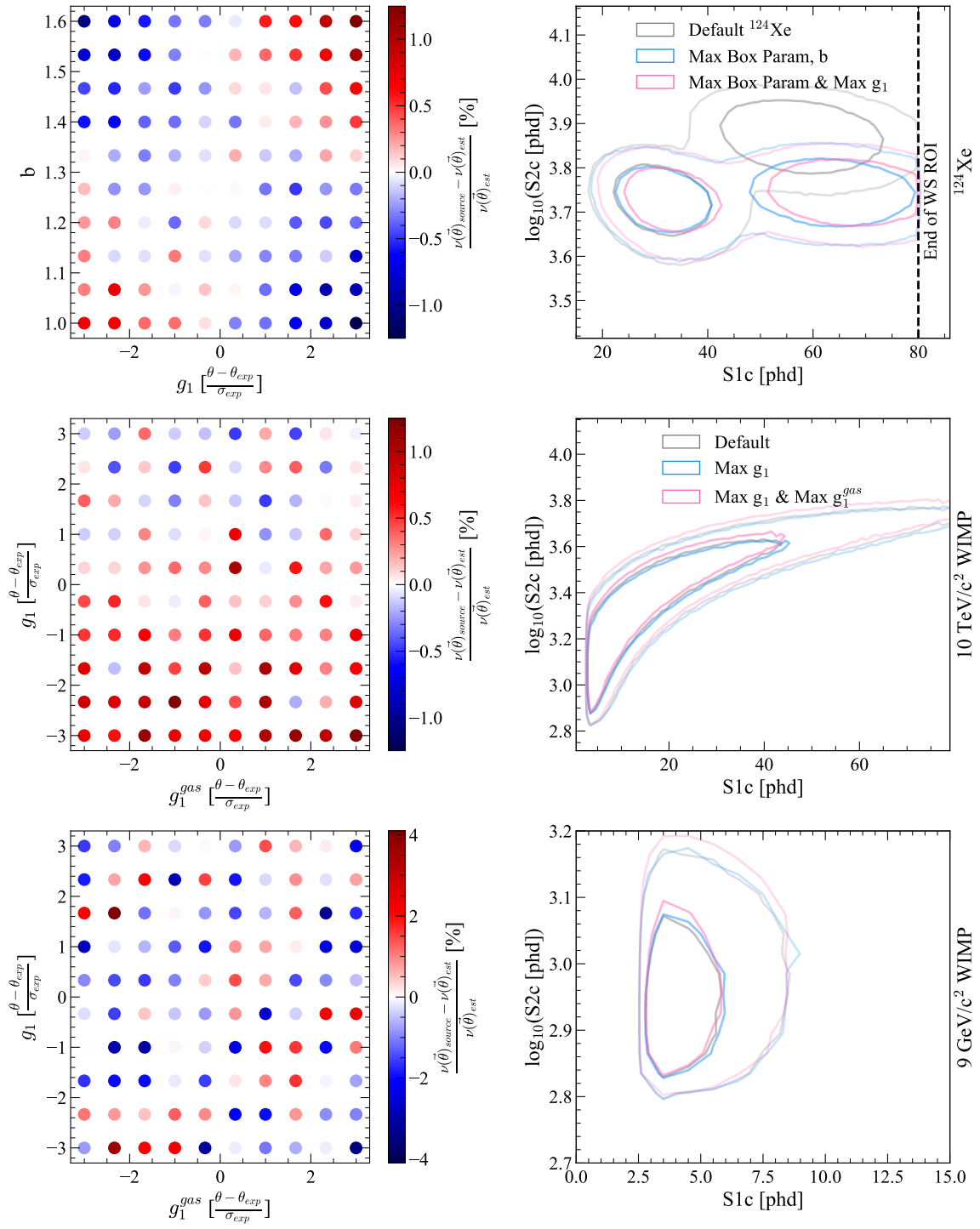


Figure 6.26: Checking the rate estimators for ^{124}Xe and two WIMPs, one low mass to represent a threshold dominated source and one at high mass. The left column is the difference of assuming separable parameters with linear cross-interpolation and estimating the rate with the source on a grid of points. The right column gives 2D contours of the models to interpret the rate. The bounds of the parameters is $\pm 3\sigma_{exp}$ and within this g_1 and g_1^{gas} are separable for both types of WIMPs, this may not hold for different thresholds or greater uncertainty. There is a dependence between g_1 and the box parameter, b, but this is small compared to the total variation in rate in Figure 6.25.

6.6.3 Optimiser Considerations

The optimizer is what is used to determine the parameters that maximise the likelihood, or minimise the log-likelihood. It is therefore one of the most crucial components of any inference routine. With auto-differentiation available the most efficient choice of optimiser is one that utilises the explicit gradients of the likelihood. The default optimizer used in `flamedisx` and `flameNEST` is `scipy`'s 'trust-constr' algorithm [144]. Details can be found in the source book in Ref.[175], but essentially this works by Taylor-expanding the likelihood function and solving a quadratic sub-problem at each step. How far that sub-problem is trusted, and therefore how large the steps can be, is defined by the trust-radius, which the algorithm has methods of obtaining. This means that the algorithm is sensitive to the precision of the first and second order gradients of the likelihood function. Importantly In order to include constraints, a barrier parameter is determined that quantifies the distance to the constraint. The algorithm attempts to find a minimum by finding a zero gradient and minimising the trust-radius and barrier parameter.

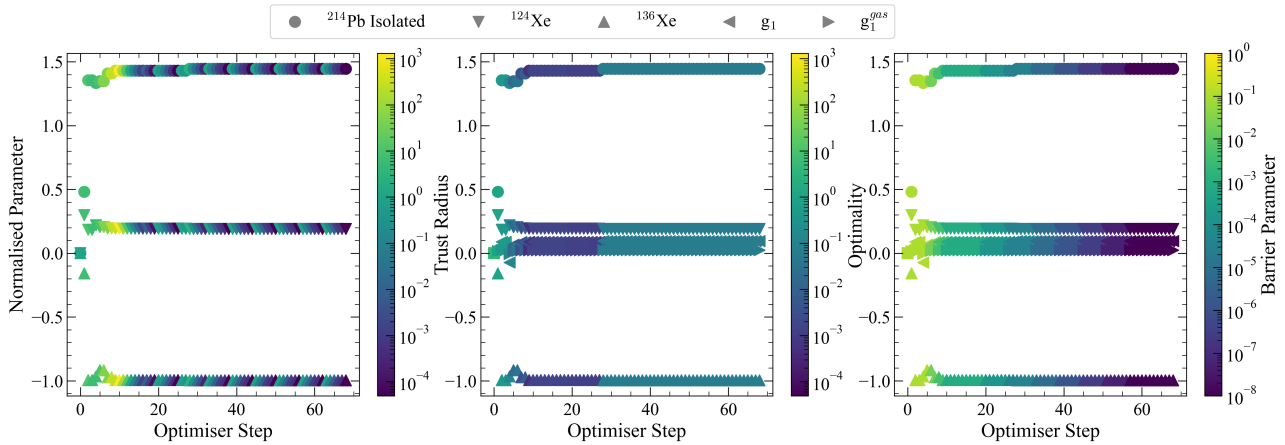


Figure 6.27: Fitting a simplified model with a handful of sources and saving the progress of the optimiser. Plotted are the normalised parameters, often to a guess or by their bounds, against the steps in the optimiser with the colour indicating the value of one of the optimiser's parameters. The optimiser wishes to minimise the trust radius, i.e stating it does not need to step far as it has reached a trusted optimum point, and this is seen to be initially minimised but then oscillating with no change in the parameters (left). The normalised gradient of the function with additional terms for constraints, the optimality, reaches a minimum then, alongside the parameters, never changes (middle). The only parameter that is changing after the first 10-20 steps is the barrier parameter, which quantifies how close parameters are to their physical boundaries (right). This may be due to the problem being tightly bound, especially in the shape varying parameters, so the tolerance for the barrier parameter should be adjusted.

The creators of `flamedisx` had adjusted the tolerances for the trust-radius and gradient to numerical accuracy but not the barrier parameter. The author found this can lead to excessive evaluations as seen in Figure 6.27 where the optimiser is making excessive steps for no improvement in the likelihood or change in parameters. Rather than quote a general

form for the tolerances, as it may be model specific, the author notes that a future analysers should reproduce the above plot on toy data to determine a reasonable configuration. In non-asymptotic routines, or in creating pull plots to test the model, it is necessary to carry out $\mathcal{O}(10^3)$ fits on toy data. For these, the true values of the toy data are known, such that one can start the optimiser at a guess close to the MLE. For both, the goal is to infer the variance for the MLEs around the true values, so tighter tolerances on all three of the above parameters may be acceptable to reduce computation time.

6.7 WS2024 Full `flameNEST` Pulls

Carrying out pull plots not only tests for biases in the model, but is a similar deployment of `flameNEST` on a computing cluster as non-asymptotic inference. Pull tests with and without grouping, as well as with and without shape parameters, are shown in Figure 6.28. There is no significant bias caused by the use of the ER source group and there is a bias in Detector NRs caused by the boundary effect for near-zero counts, identical to that in the template model in Section 5.2.4. It appears that there is a bias g_1 and g_1^{gas} which, as can be seen with their removal, causes the bias in the dominant ^{214}Pb and ^{85}Kr rates. Investigating the correlation between these two shape varying parameters in Figure 6.29 shows that when g_1^{gas} varies significantly from g_1 appears to absorb it. This may be caused by an anti-correlation not captured in the total rates, or an error in interpolation. It is possible this is caused by the choice of optimiser tolerances but it is unlikely as the two other shape parameters show no bias. Finally, in both scenarios, the box parameter does not exhibit the same bias seen in the template model in Section 5.2.4. This indicates that either the box-parameter bias in the template analysis was driven by the error in BSpline interpolation, and not boundary effects, or that the increased discrimination from ^{214}Pb isolated softens the boundary effect. The latter is unlikely, as one would expect radon tag to give better discrimination.

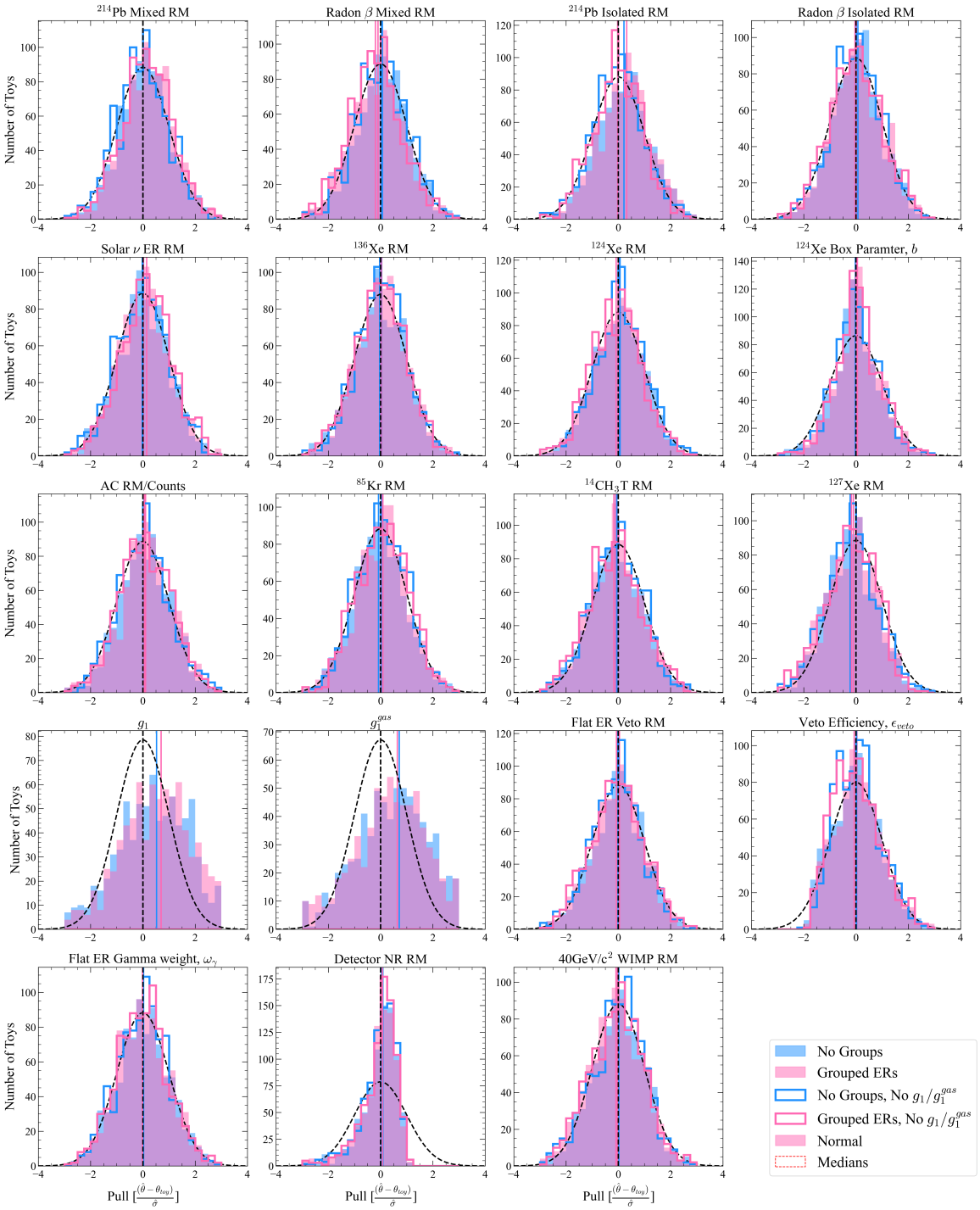


Figure 6.28: Pull plots with and without using grouping, it can be seen that there is a bias in the shape varying parameters g_1^{gas} and g_1 as well as in the dominant ER rates of ^{214}Pb mixed, ^{214}Pb isolated, and ^{85}Kr sources. If these shape parameters are removed the bias in the rates disappear. The tolerances for these toy pulls were 10^{-2}, 10^{-4} and 10^{-2} for the normalised gradient, trust radius, and barrier parameter, respectively. These were informed by fits to observed data and fits to toys on smaller likelihood models, such as those used in Figure 6.27. As in Section 5.2.4, in toys the box parameter and gamma-weight were uniformly sampled within their physical bounds. The mean rate for Detector NRs were uniformly sampled between [0,9] counts in the veto sideband, which corresponds to approximately [0,1.] in the WS ROI. The WIMPs were sampled with a mean rate of 50 counts over the entire ROI.

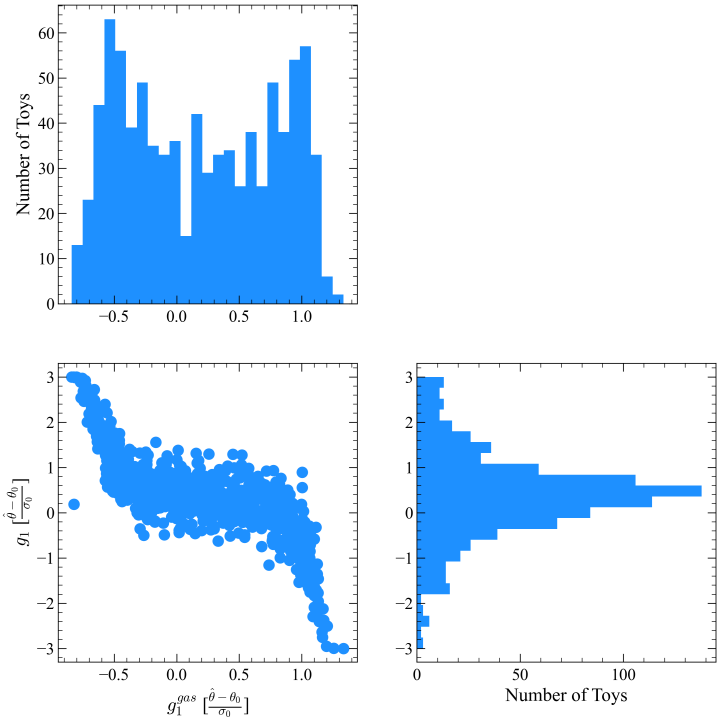


Figure 6.29: Maximum likelihood estimators for g_1^{gas} and g_1, $\hat{\theta}$ without using grouping. The parameters are normalised to the constraint centre, θ_0, and width, σ.

6.8 Improvements in Discrimination

This section quantifies the improvements in discrimination for dominant backgrounds of a WIMP search and solar neutrinos against standard backgrounds in LZ. This does explore one of the primary purposes of adding position dependence to LZ inference, which is expansion of the FV. The FV is chosen based on leakage of highly position dependent backgrounds that grow radially towards the data-wall. These are called wall backgrounds, and are gammas and other ERs that can have a modified response due to charge loss near the data-wall, as was seen when constructing the detector ER spatial distribution in Section 6.2.2. These modified ERs can pose a significant background to the WIMP search, so the FV is chosen to reject as many as possible. Now, with position as an observable to discriminate against them the FV can be expanded, but this would require some dedicated work to model the modified ER response. This is true to a lesser extent for Detector NRs, as the increase in rate is less significant because their mean free path is $\mathcal{O}(10)$ cm and the FV can only gain a few cm

As sensitivity routines have not been developed for this `flameNEST` model, the author chose to follow the original `flamedisx` paper in Ref.[145] and quantify discrimination with a ROC plot[3] which, in this case, shows the signal acceptance vs background leakage with some varying cut in the space. The parameter chosen is the ratio of likelihoods for an event to be from signal or background, which becomes the ratio of differential rates. By the Neyman Pearson Lemmma, this provides the maximum power single-event discriminator [1]. To produce an ROC curve the ratios are calculated on background data and signal data, then likelihood ratios for both. The maximum accepted value of the ratio is varied and accepted signal and background, being leakage, calculated from their datasets.

[3]Receiver Operator Characteristic (ROC) plots were originally developed by radar operators in WWII. Nowadays they are used extensively in particle physics when optimising cut parameters.

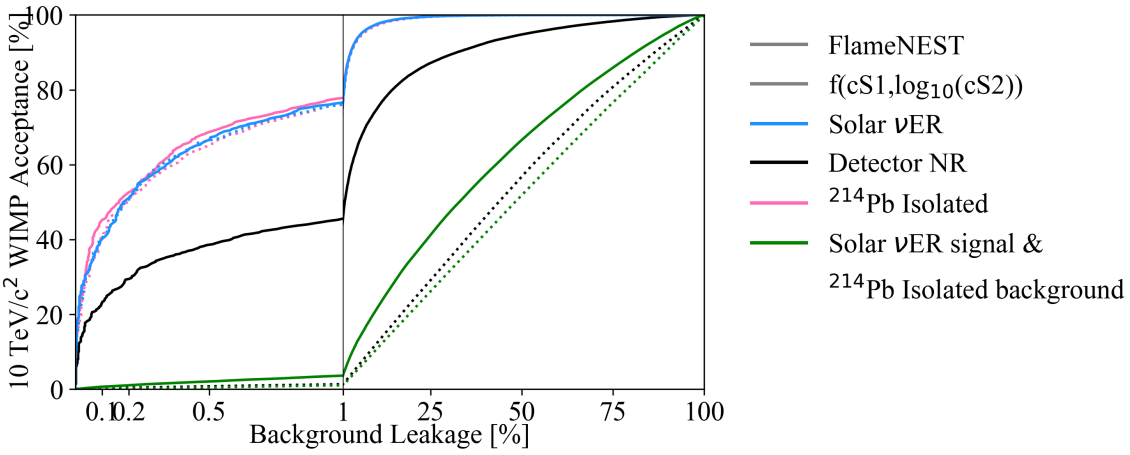


Figure 6.30: The calculated ROC curve, explained in text, comparing discrimination vs signal acceptance curves for various sources using the full `flameNEST` calculation (solid line) to a MC derived 2D template PDF (dotted line). The signal source is always a 10 TeV/c^2 mass WIMP, except for the green line where it is a Solar νER source. The templates were binned in (cS1,log$_{10}$ cS2) using 10^7 MC simulations and the same binning as the WS24 analysis. The Detector NRs and 10 TeV/c^2 WIMPs have a similar spectra and distinct position distribution. The ER sources can be noisy at low values of discrimination, in the future this could be resolved by importance sampling.

The author constructed this test for several sources, as shown in Figure 6.30. Unlike in the `flamedisx` paper Ref.[145], we do not see a factor of two improvement in discrimination between the uniform ERs and NRs of Solar νER and WIMP, respectively. This is because the ratio of LZ's electron lifetime to maximum drift time is far greater than assumed in that paper. Therefore, the resolution in S2 does not significantly change over the drift time of the detector. This may not be the case for other position dependence that has not been modelled here, like drift field, especially if the FV is expanded. It is also worth noting that just because these dimensions do not improve predicted discrimination of this source, they may improve a fit to any observed data as it was shown in Section 6.3 that the distributions have a measured impact on the (cS1,log_{10} cS2) space.

Finally, as another test of discrimination the author decided to use the box parameter inference study in Section 5.2.8 and calculate the asymptotic significance of the DEC recombination model being unique from the EC one. The asymptotic significance, rejecting $b = 1$, was 3.0 in the `flameNEST` model and 3.5(2.8) with(without) radon tagging in the template model. Note that these numbers have been re-calculated asymptotically on salted data with the 10 TeV/c^2 WIMP signal included, for close to one-to-one comparison. The two models still differ somewhat, as without radon tagging in templates merges the mixed and isolated state. This means that the spatial discrimination of the ^{214}Pb source, and potentially removal of interpolation, has improved the significance but is out-performed by radon tagging. This is to be expected as the radon tagging had an efficiency of $1-\lambda_{rn} = 85\%$ and leakage of ^{214}Pb isolated of $1-\epsilon_{rn} = 15\%$.

The implementation of radon tagging is possible in the future but the author omitted it for simplicity in this analysis.

6.9 Performance and Applicability

For the goal of implementing `flameNEST` in rare-event searches, non-asymptotic inference tools are often required in near background free regimes. Now that a preliminary `flameNEST` model has been constructed, and deployment tested, we can evaluate whether or not performance is suitable for non-asymptotic inference. In Table 6.4 the timing of each event likelihood configuration is presented. To project this for full-inference we can determine the node hours per confidence interval, i.e per mass in the WIMP search, using the formula

$$\text{Node Hours} = 2 \times (N_{POI} + 1) \times N_{toys} \times T_{fit}[hours]/n_{tasks},$$

where N_{POI} is the number of parameters of interest points scanned, N_{toys} is the number of test statistic samples used, T_{fit} is the time to fit to a single toy, and n_{tasks} is the deployable tasks on a node. The factor of two comes from the conditional and unconditional likelihood evaluations for the profile likelihood ratio. The author has only managed to deploy tasks on each of the four GPUs on PERLMUTTER[4]'s GPU[5] nodes, so $n_{tasks} = 4$. To derive a confidence interval for a mass in the WS24 analysis $N_{toys} = 8 \times 10^3$ and $N_{POI} = 16$ were used. From the performance in Table 6.4, and from the pull tests, a 6D `flameNEST` fit with multiple microphysics and detector parameters can take $T_{fit} = 20 - 30$ minutes. This would then require 22-30k node hours, or 40-60% of LZ's GPU allocation on PERLMUTTER in 2025. This is not feasible.

Configuration	Evaluate value and gradient	
Description (batch size)	With Hessian	Without Hessian
No Group, fix g_1/g_1^{gas} (25)	88.6±0.4	81.2±0.09
Grouped ERs, fix g_1/g_1^{gas} (25)	45.4±0.3	36.8± 0.4
Grouped ERs, fix g_1/g_1^{gas} (50)	40.8 ±0.6	34.7± 0.5
Grouped ERs, float g_1/g_1^{gas} (50)	63.2±0.5	36.4 ± 0.4

Table 6.4: Performance of `flameNEST` calculation for the WS24 log-likelihood, including a WIMP source. Optimisers can take somewhere between 20-30 steps to find the maximum likelihood in toy fits. Note that with groups the maximum batch size tested, and deployed, was 50 and without it was 25, due to memory requirements of the GPUs.

There are a range of ways to enhance this performance. Firstly, if one were to repeat the WS24 template analysis exactly, i.e fix the $\gamma - \beta$ weight, g_1, and g_1^{gas}, potentially to the unconditional fit values, then all sources but the ^{124}Xe box parameter could have their differential rate calculated once and fetched for each evaluation of the likelihood. Creating a likelihood with just

[4]

[5]Though more may be possible with a trade off in batch size and improved high-performance computing tools.

^{124}Xe and the box parameter defined yields an evaluation time of 2.44±0.01 s with Hessians. As fetching the differential rates for the other sources would be a negligible task this corresponds to a speed up of 26× on the grouped calculation. This factor is large as the ^{124}Xe energy spectrum consists of only two mono-energetic lines, meaning the size of the energy dimension for the central block is only 2 instead of 100 for sources with a continuous spectrum. This in itself would bring non-asymptotic inference with 6D observables and the box parameter into the realm of feasibility. A more practical aspect is that, on many computing clusters, there are discounts for jobs that can stop and start. This is a factor of four on PERLMUTTER and it is relatively easy to save the state of the `scipy` optimiser to restart later [149]. To keep g_1 and g_1^{gas}, or other detector parameters floating, one could cache the central block of the source groups, in preliminary tests by the author found that this could speed up differential rate calculations by up to a factor of 5, but further work is needed to see how this is integrated in a likelihood with many sources.

It was noted in Ref.[111] that higher-order terms in asymptotic approximations may take advantage of the exact first and second order gradients from auto-differentiation to obtain an accurate distribution of the test statistic with far fewer toys. Another optimisation at the inference routine level could be the use of importance sampling to bias the distribution, using a surrogate metric like leakage without evaluating the fit. The factor of improvement for importance sampling would be inversely proportional to the p-value being evaluated, so it would be a far larger improvement for discovery than for confidence intervals. The author believes that both of these warrant implementation regardless of `flameNEST`'s implementation as template inference routines can still be costly

It is worth noting that the likelihood evaluation times scale linearly with the number of sources and events to evaluate, for WS24 this is 13 sources and 1.3k events. For a near-threshold low mass WIMP search, where ERs are extremely distant, there are essentially only four sources: WIMPs, ^{8}B CEνNS, Detector NRs, and ACs. Further, there are far fewer events, with ~ 30 predicted events for an exposure larger than WS24 [151]. This means that there could be an approximately 300 times reduction in the node-hour requirements. This tool could be imminently ready for near threshold non-asymptotic searches after some additional considerations on position smearing, as discussed in Section 6.2.1 the position uncertainty grows with decreasing S2. Another area this tool is imminently ready for are low energy ER searches, for example for solar axions or neutrinos, that use the same data set and models as WS24. While this remains to be confirmed, the large background is expected to place these searches in the asymptotic regime, thereby alleviating the performance concerns discussed in [152]. In addition, the improved spatial discrimination seen in Figure 6.30, time dependence of ER signals and contaminants, and shape varying parameters will have a larger impact in this background limited regime.

6.10 Summary of Full `flameNEST` in WS2024

A `flameNEST` implementation for WS22 had been carried out by Dr Robert James in their thesis in Ref.[111]. However, this was a limited implementation not allowing for any shape varying parameters. The author implemented a preliminary full `flameNEST` model for WS24, requiring intimate knowledge of the physics and detector modelling, analysis techniques, and background modelling in order to configure the model. The WS24 configuration required implementation of the novel DEC model for ^{124}Xe, requiring a microphysics yield parameter which had previously been impossible before the authors work in Chapter 4. Since it was now possible, the author also implemented a $\beta-\gamma$ charge yield weighting for the detector ER gammas, as they are by default modelled as pure betas. Unlike the single-likelihood WS22 model, the WS24 model includes a veto sideband and the injection of sources at different times requiring the implementation of a combined likelihood with four spaces. The author had assisted in developing the simulation derived position maps for WS22 and developed new ones for WS24. The implementation of the veto sideband necessitated improvements in the detector ER distribution as purposefully conservative choices from WS22 now caused a bad model. The author also checked that the use of observed position is consistent, and quantified where position smearing would be relevant.

Having produced the preliminary model the author carried out a preliminary fit to salted observed data. The author used `flameNEST`'s simulate functionality to implement GOF tools developed for templates in Section 5.1.2, finding good overall agreement but some significant areas of poor modelling that require further work. Namely at the edges of the time distribution in the Veto and the radial position distribution in both spaces. Interestingly, the improved spatial and temporal discrimination appears to have identified detector NR events previously lost to background. The scatters in the veto detectors appear to be consistent with neutrons, and it is apparent which observables revealed which events. The detector ER charge yields were found to be almost purely beta. The addition of g_1 and g_1^{gas} showed best fits consistent with the pre-fit model, but both saw significant improvement of constraint, showing `flameNEST` can improve on calibrations with just the in science data.

The author explored several optimisations for the full `flameNEST` implementation, ranging from tuning of the optimiser parameters to improved methods of calculating the differential rates. The author implemented pull-tests to check these performance improvements and found that they introduced no bias. However, there was a model bias remaining in g_1 and g_1^{gas}, but this may be due to their degeneracy. Ultimately, the improvements to the full `flameNEST` calculation are insufficient for non-asymptotic inference to be used. Some simple approximations, like limiting the sources with shape-variation, will need to be applied. Alternatively, performance improvements at the inference routine level could be explored. As sensitivity routines where not implemented at the time of writing this thesis, improvements to discrimination were shown

in several metrics. It was noted that for low energy ER searches this model could be deployed imminently, with some validation of the inputs and implementation of radon tagging. With position as an observable, it is now possible to enhance the WIMP search reach not only by discrimination detector NRs but by expanding the FV and discrimination against wall backgrounds. However, this would take more substantial analysis to model the wall backgrounds. Finally, due to scaling with size of the data set and sources, the performance of `flameNEST` is sufficient for low mass WIMP searches, but some validation and additional modelling for position smearing would be required.

6.11 Outlook

The author has shown that `flameNEST` is computationally feasible for inference with a few in-separable shape varying parameters, any number of separable ones, and many observables with $\mathcal{O}(10^3)$ events the dataset and $\mathcal{O}(10)$ sources in the model. Performing inference on WS24 for low energy NR searches, such as WIMPs, with the enhanced observable space and the ^{124}Xe box parameter, and no other shape varying parameters, is feasible. This may improve WIMP sensitivity through the Detector NR constraint, but this would be subdominant to the expansion of the FV. Therefore, expansion of the FV and a suitable wall model should be prioritised. Without expanding the FV, it is possible to leverage position and time dependence to improve sensitivity in low energy ER searches. Some work on the application of the radon tag would be required for this. Due to the scaling with events and sources, `flameNEST` could be fully leveraged for threshold NR searches like 8B and low mass WIMPs. These are dominated by systematics, so `flameNEST` can improve constraints by introducing calibration data directly as a sideband in the likelihood model [151]. In addition, interactions of the threshold and LCEs induce a position dependence in the signal. Some extra work to ensure the increased position uncertainty at these low energies, and therefore S2s, is handled properly.

As discussed at the start of this chapter, the alternative pursuit with `flameNEST` would have been to integrate it into calibrations, yielding several advantages, such as greatly reduced MC uncertainty, no need for bands, and easy model customisation. This would increase adoption across the collaboration and allow calibrations, simulations, and inference to use the same framework and increase efficiency. As discussed in Chapter 4, this would require further work to manage many in-separable parameters, such as integrating `flameNEST` into MCMC.

Chapter 7

LZ Projections

During the running of LZ the author has carried out two major sensitivity studies; one to evaluate the impact of changing the electric field configuration and one to assess extensions of operational time required for specific science goals. The purpose of these studies was to extrapolate the performance of LZ with the best knowledge at the time for collaboration decision making in this chapter. Therefore, results are reported in live days, and fiducial mass estimates are omitted to avoid direct comparison with other experiments.

`LZStats`

The `LZStats` package was the statistical inference software based in ROOT and C++, developed for use by the LZ collaboration by Ibles Olcina in Ref.[176]. It is a robust package used on previous LZ results and by the author in the following sections. However, for WS24 and beyond, requirement for integration of future tools and the opacity of the framework necessitated the move to `LZFlameFit`. `LZStats` functions by using a RooWorkSpace which specifies the likelihood function and a yaml configuration file which provides options for inference, such as the range of the parameter of interest, confidence level, etc. One would need to create a new workspace for a different configuration of the model, such as accumulated live days, WIMP mass, or a different signal altogether. The parameter of interest is often the counts and here a unique yaml configuration would be required for scanning a greater range in the presence of more background, such as with increased live time.

7.1 Changing LZ Drift Field

The configurations of the LZ detector for science data acquisition is broken into three distinct Science Runs (SR), SR1 for WS22, SR2 was an intermediate state, and SR3 for WS24. The change in SR1 to SR2 was mainly the different Gate and Anode configuration to attempt to lower the spontaneous grid emission rate that contributed to isolated S1/S2 rates. In this section, the change between SR2 and SR3 is discussed.

In December-January 2022 LZ started to observe a period of strong light emission in the skin at the 9 o'clock position with some increased rate observed in the TPC. Increased rates in the skin could lower the skin's efficiency to tag gammas and neutrons in a spatially dependent way, making it difficult to correct for, especially with its poor spatial resolution. Light leaking into the TPC could lead to an increased isolated S1 rate, and therefore AC rate, or if mishandled a false detection of a physical event. The hypothesis was that this light was from the RFR field-cage resistors or the HV feed-through, as it was likely originating from the bottom of the detector. It was observed that lowering the cathode voltage , therefore the fields around these components, decreased the skin light emission, but doing so could compromise the ability to differentiate ERs and NRs in the TPC. The author was part of a task force to investigate the impact on WIMP sensitivity due to changing conditions. The author was responsible for creating the necessary `LZLAMA` simulations, processing them in standard tools, and assisting Makayla Trask[1] in learning to use `LZStats` to create the final limits.

Four field configurations were considered, as shown in Table 7.1. The altering field did not impact the FV cut used as `LZLAMA` used a drift map from SR1. Although not technically correct, it was sufficient to understand the impact on sensitivity from a changing drift field solely caused by the microphysics response to the drift field. From the NEST studies in Ref.[120], it

Cathode voltage (kV)	**Drift Field (V/cm) [configuration]**	**Cuts (Skimmer)**
-9	23 [SR2]	G/A = -6/1.5 kV SE_{size} = 50 phd $S1_c > 3$ phd
-15	61.8 [SR2]	
-20	96.2 [SR2]	
-32	192 [SR1]	$S2_{\mathrm{raw}} > 10 \times SE_{size}$ G/A = -4/4 kV SE_{size} = 60 phd $S1_c > 3$ phd $S2_{\mathrm{raw}} > 10 \times SE_{size}$

Table 7.1: Summary of grid voltages, G/A meaning Gate/Anode, drift fields and cuts in Skimmer alongside assumed SE size. There were no efficiencies applied for this study. The field is presented with a tag for the model configuration.

was expected that the leakage would not increase significantly for drift fields above 60 V/cm, but this may depend on certain detector parameters like extraction efficiency and S1 LCE. Therefore, the author reconfirmed this by calculating the leakage for the different proposed configurations, shown in Figure 7.1. Here, leakage is defined as the fraction of background events below the signal median in $\log_{10}$S2c vs S1c. The set of LZ performance requirements from the submission of the Technical Design Review (TDR), in Ref.[96], to the Department of Energy (DOE) is met as long as the drift field remains above 60 V/cm. In addition, the suspected drift field of XENONnT at the time was tested to compare against the competition.

[1]PhD student at University of California Santa Barbara

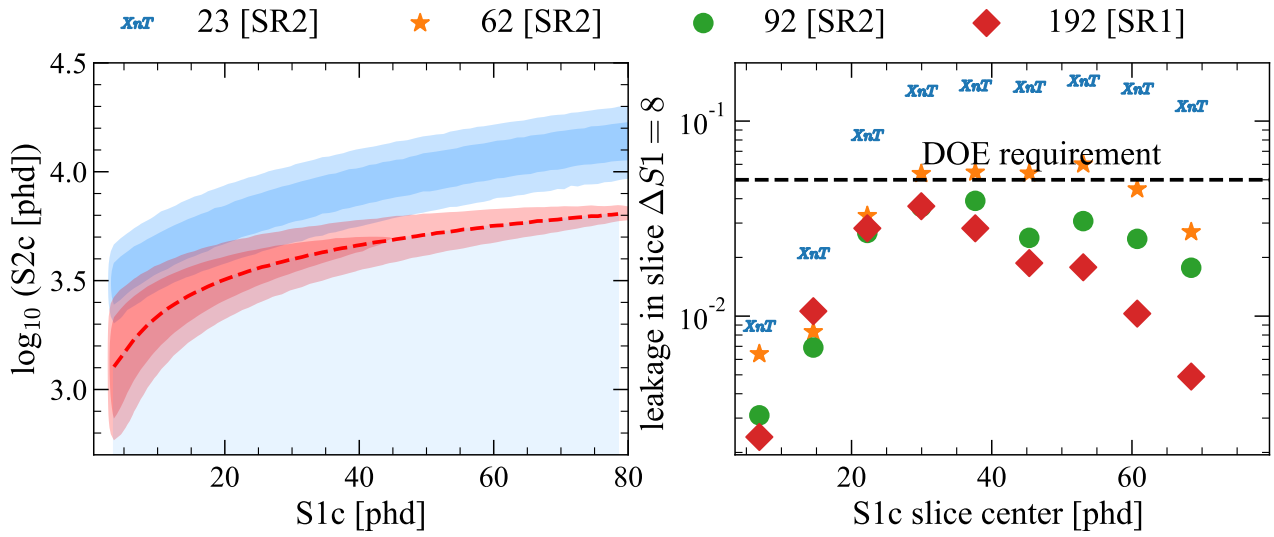


Figure 7.1: Diagram of the flat beta ER background into a 10 TeV/c^2 WIMP model and light-blue shaded region for which an event is determined to be leakage (Left). The leakage is then broken into slices of S1c to see how it evolves and to evaluate the maximum (Right).

To quantify the impact on LZ sensitivity with a full background model, the WS22 background model was used and its rates were extrapolated to 300 live days. This excluded the AC model and activation products. The AC model was excluded as it was uncertain how the AC background would respond to the field. Although lowering the field increased the maximum drift time and thus the potential for an increased rate of AC events, studies by Jose Soria and Amy Cottle showed a significant decrease in the rate of unphysical drift time events at the lowered cathode voltage, suggesting a lower rate in the isolated S1 and/or S2. The activation products of ^{37}Ar and ^{127}Xe decay over months and days, respectively, so they are safe to neglect for long-term studies. The full WS22 background model is described in the WIMP search paper in Ref.[95], the counts of all included backgrounds and sources are assumed to scale linearly with live time. Note, the latter means that the rate of ^{85}Kr ingress was assumed to be negligible.

The results are shown in figure 7.2, as expected from the leakage study, the median sensitivity and the median required number of WIMP events to claim 3 σ or more significance after a 300-day exposure remain similar under each field configuration $>$ 60 V/cm An additional configuration was implemented for this check of the SR2 model that was a fit to preliminary calibrations performed by Elm Zweig[2] at a field of 192 V/cm. This configuration was indistinguishable from SR2 configuration with 92 V/cm on the cathode. This implied that the difference between SR1 at 192 V/cm and SR2 at 92 V/cm was dominated by the changing G/A configuration.

[2]Former PhD student at University of California Los Angeles

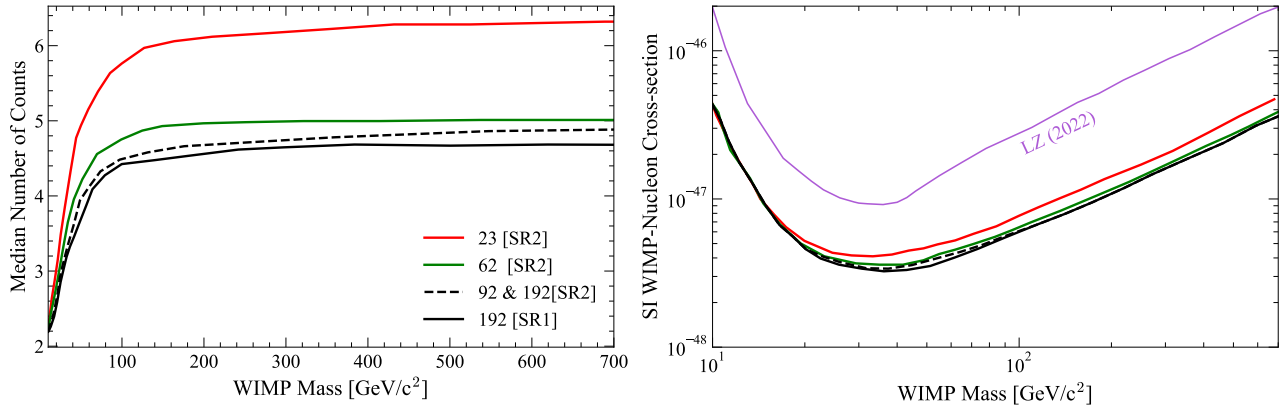


Figure 7.2: Results of `LZStats` inference carried out by Makayla Trask with guidance from the author using field configurations from Table 7.1 assuming 300 live days. The median number of WIMP events required to obtain 3 σ significance, or greater (Left). The median WIMP sensitivity (Right). One line represents 92V/cm and 192V/cm at the SR2 Gate and Anode configurations as they are virtually degenerate.

An additional benefit to the decreased field was a small increase in recombination, and therefore in S1 light yield, for NRs which drove an increase in ^{8}B neutrino CEvNS, or low-mass WIMP, events passing coincidence cuts. The change in yields is small, as can be seen in Figure 7.3, but this led to a 15% increase in rate if only applying the 3-coincidence cut, because the gradient in rate across this cut is large. It is worth noting this was with default NEST parameters not tuned at these energies or drift field and that the impact would change with different analysis cuts like S2 threshold. Given the negligible impact on WIMP sensitivity, potential complexities if the skin light emission remained at a high rate, and benefits for threshold NRs the field was lowered to 92 V/cm for the SR3 configuration used in WS24.

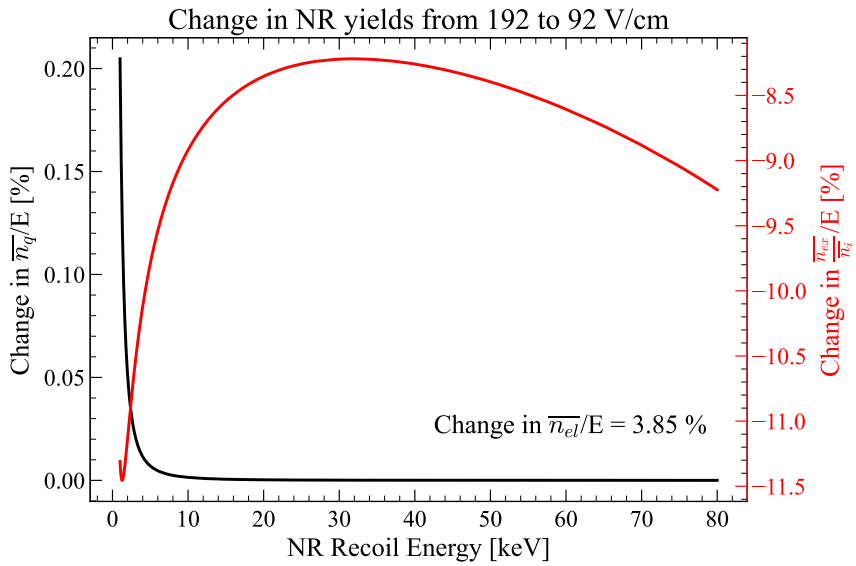


Figure 7.3: Change in NR mean yields between drift fields of 192 V/cm and 92 V/cm. The charge yield is shifted by a constant for changing fields.

7.2 LZ Run Duration

Due to the new field configuration, and loss of calendar time during the COVID-19 pandemic, it was necessary to determine if LZ required an extension of operations to 2028 in order to achieve its science goal. The author worked with several senior collaborators in a runs-duration task force to determine the required run time to achieve a median expected upper limit of 3×10^{-48} cm^{-2} on a 40 GeV/c^2 WIMP and 3σ median discovery sensitivity on ^{8}B CEvNS. This study required generating all the necessary models, making fundamental assumptions about the ROI, acceptances, and AC model at a time with only preliminary information. This also posed a technical challenge in the `LZStats` framework, which designed for a particular result or sensitivity at a given live time, so the author designed a routine to construct necessary workspaces and submit a live-time scan to the NERSC[3] computing cluster.

The background counts and sources were identical to WS22, again extrapolating the counts determined there and omitting activation sources, with the exception of the predicted rates for AC and ^{214}Pb. The `LZLAMA` model was updated with a drift map and field reflecting SR3 conditions, with the analysis cuts adjusted accordingly. For the AC model, three possible rates were considered for rate of events in the ROI, informed by a preliminary investigation of the data by the AC team. Initial radon tagging studies indicated that a 70% reduction of the ^{214}Pb rate was feasible at the cost of 10% of the exposure. This was implemented as a raw loss on exposure, as the side-band method had not yet been conceived. For the ^{8}B CEvNS search,

[3]National Energy Research Scientific Computing Center (NERSC) is the standard computing cluster for LZ.

the ^{214}Pb distribution has a small impact on sensitivity, so only the varying AC rate with no exposure loss was considered. For the WIMP search two scenarios were considered: goal and conservative. The former would have the best accidental rate and radon tagging, whilst the latter the worst and no tagging.

	Conservative	Goal
Acceptances	WS22 f_{accept}^{S1c} linearly extrapolated from S1c>3 to > 0 phd & $f_{accept}^{S2c} \rightarrow f_{accept}^{S2c}(S2c \times \frac{SE^{SR1}}{SE^{SR3}})$	
Cuts	S1c>0 S2> $5 \times SE^{SR3}$ 3 coincidence	
^{8}B CEνNS Rate†	4.13 events/yr	
^{214}Pb Activity [μBq/kg]	3.26	0.98
AC Rate [μHz]	0.500	0.125
AC Rate* [events/yr]	40	11
AC Rate uncertainty [%]	21 as in WS22	
Exposure loss [%]	0	10

Table 7.2: Table of cuts and rates for the runs duration study. The SR1 and SR3 single-electron sizes are indicated in $SE^{SR1/SR3}$. All other rates were assumed to be identical to WS22.
* - The overlap between the ^{8}B and AC model was around ∼20%.
† - The ^{8}B CEνNS rate was underestimated by approximately a factor of 2, see text.

For modelling physics signals and backgrounds, the author applied a baseline ROI shown in Table 7.2. In order to better represent data with the best knowledge at the time, the author created modifications of the acceptances from WS22. The cS2 based acceptances were scaled to the appropriate SE size, essentially claiming that the efficiency for the same number of electrons was the same. The S1 acceptance was linearly extrapolated down as it terminated at 3phd for WS22. The modelling chain follows;

- `LZLAMA` Produces RQs like S1,S2$_{raw}$, cS2, drift time, etc. and MCTruth information about generated events, such as the event weight,ϵ, which is normalised to an assumed rate in events/tonne/year.

- `BGSkimmer` At the time, the `BGSkimmer` applied analysis cuts, removing all events that failed, and modified the event weight to contain chosen acceptances that are some function of the RQs , $\epsilon \rightarrow \epsilon \times f_{accept}(RQ)$.

- `BGAssembler` Merges all necessary sources into a single file and creates histograms normalised to total rate.

- `LZStats` The `BGAssembler` and AC file is used to create a RooWorkspace with custom code for varying the WIMP mass and live days. `LZStats` processes this through its inference routines.

For this analysis, the `BGSkimmer` acceptances were turned off and the `BGAssembler` histograms were modified with the appropriate efficiencies in a custom script.

To produce the AC model with the interim data, the author obtained the interim AC chopstitch from the AC team [4]. As there was an additional feature of a peak in the isolated S2 rate at low S2s in the data previously not observed, the author created a new KDE smoothing, with a variable resolution to capture the low-cS2 peak, shown in Figure 7.4.

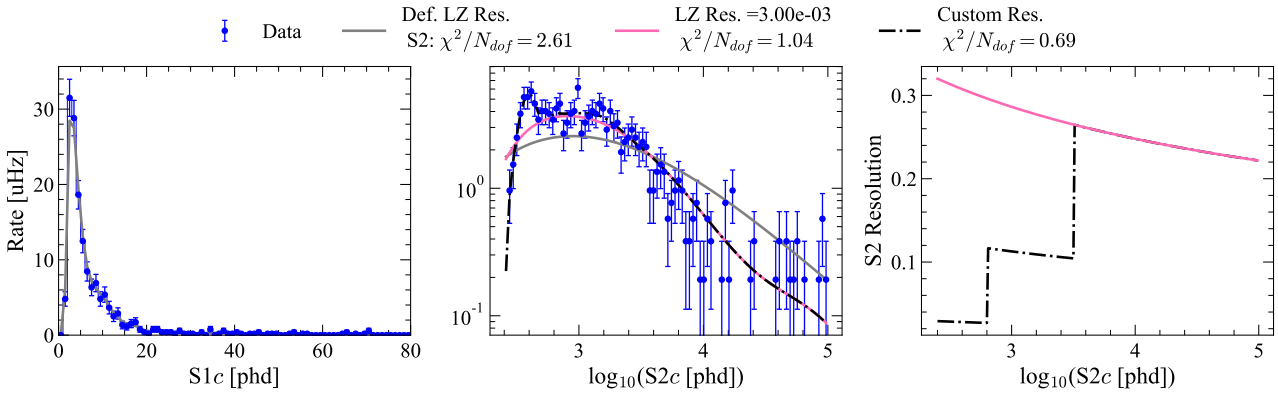


Figure 7.4: Fitting the isolated cS1 (left) and isolated cS2 (middle) distributions for a preliminary AC model. The cS2 resolution function used is also shown (right).

The author then generated an ^{8}B model using `LZLAMA` and a pre-WS24 tuning. The results in Figure 7.5 show the overlap of the two models, 11% and 20% of the AC rate was found to lie within the 68% and 95% contours of the ^{8}B CEνNS model, respectively.

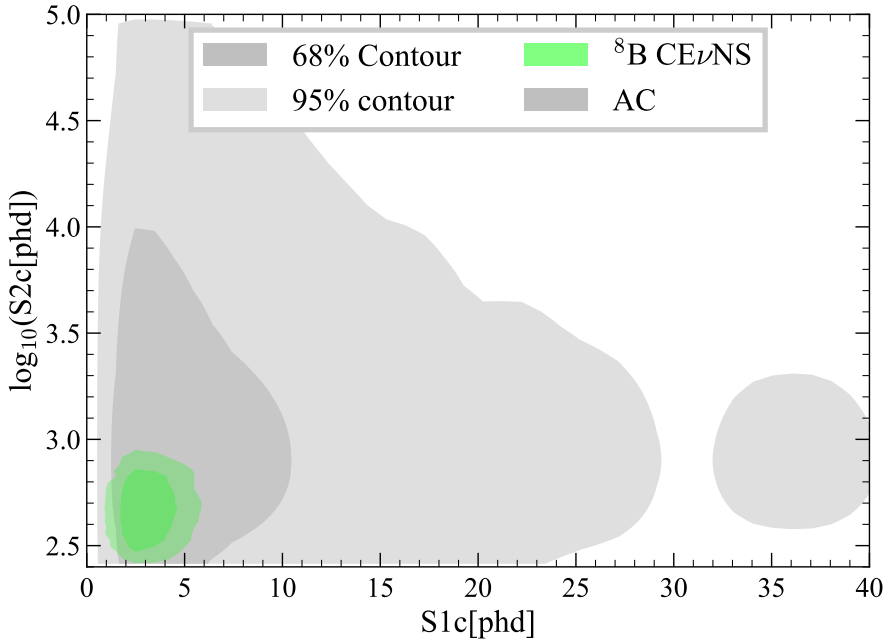


Figure 7.5: The AC and ^{8}B CEνNS model for this sensitivity study. The ER and NR models for the WIMP search have been omitted. The 95% contour of the ^{8}B model contains approximately 20% of the accidentals rate.

[4] Ibles,...

The background and signal rates, as well as the leakage considered, are low, so non-asymptotics would be the most general approach to evaluate the sensitivity of LZ. However, evaluating these distributions out to large significances can be costly. However, since only 30k toys were used to evaluate the ^{8}B CEνNS discovery significance at each exposure considered. In Figure 7.6, it can be seen that at $\sim$300 days the distribution is almost fully asymptotic. At all exposures, there are insufficient toys in the background only distribution to evaluate the significances of the full signal and background distribution. As the median significance approaches 3σ the background only distribution becomes asymptotic, prior to that it would underestimate the significance[5]. Therefore, it is safe to use the asymptotic distribution for all points. Note that the 90-day test would have approximately 10 AC events and 3 ^{8}B events, shown in Ref.[140] to be sufficient for the use of the asymptotic approximation if this was a single-bin test. However, a decent choice of ROI would have essentially 2 AC events for 95% signal acceptance. This means that the discrimination of the observables is great enough to drive it into the non-asymptotic regime.

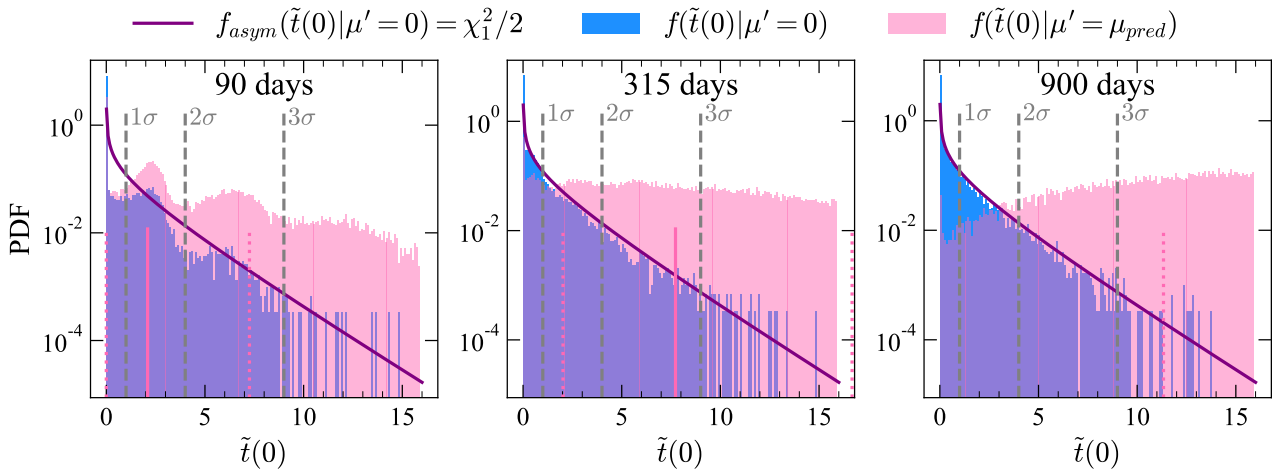


Figure 7.6: Grey vertical lines indicate the test-statistic values for asymptotic 1,2,3 σ significances. The blue histogram shows the test statistic evaluated on background only toys, this should approach the asymptotic purple line. The pink histogram is the test statistic evaluated on the signal and background toys, and pink lines give the median and $\pm1\sigma$ of the distribution.

The final results are shown in Figure 7.7, requiring at least 800-live days to achieve the DOE WIMP sensitivity requirement, and in that time we could be confident of gaining a 3σ observation of ^{8}B CEνNS. The latter statement is true despite the conservative scenario only reaching 3σ 50% of the time, as this is a rudimentary analysis that is not tuned for a near-threshold signal-like ^{8}B CEνNS, making it inherently conservative.

[5]over estimating the p-value.

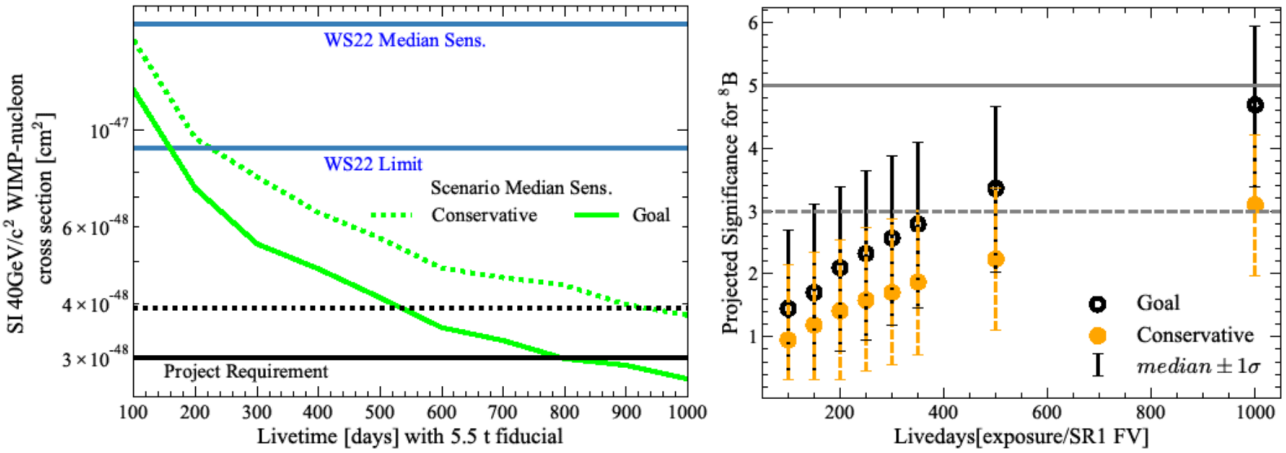


Figure 7.7: Median WIMP sensitivity vs live days, the additional dotted line above the DOE requirement indicates a 30% variation in rate, see text (Left). Median discovery significance for a ^{8}B CEνNS search vs live days (Right).

Unfortunately, after this study was carried out, it was found that the WIMP rates for >40 GeV/c^2 were under-estimated by up to 30% and this grows for lower masses to a factor of 2 for 6 GeV/c^2 and for ^{8}B CEνNS. The author spent some time discovering the cause of this error. Essentially, the acceptances had been accidentally applied by the `BGSkimmer` step in the modelling chain. It affects all sources, but as efficiencies drop at low S1 and S2 values, it disproportionately impacts ^{8}B CEνNS and low-mass WIMPS. This disproportionate impact can be understood by the overlay of models and efficiency curves shown in Figure 7.8. A factor of 30% does not greatly impact the findings of the WIMP search goal, and the two scenarios span this space, so conclusions of the report were unchanged. However, since it underestimated the time frame in which evidence for measurements of ^{8}B CEνNS was possible, it may have affected the focus of the analysers if left undiscovered internally.

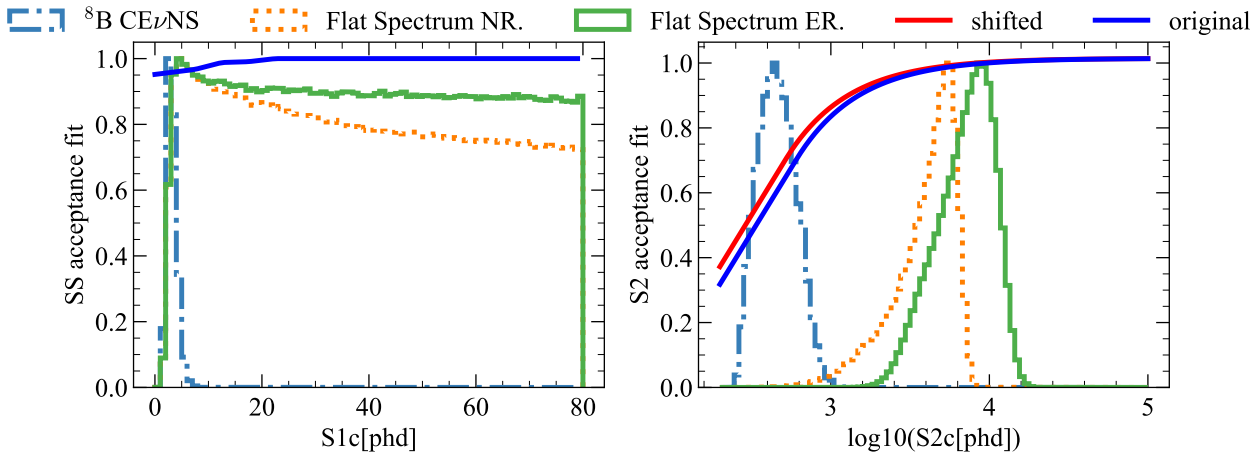


Figure 7.8: Plot of the S1c (left) and S2c (right) acceptance curves and projections of three models to show the relative impacts.

This error may have been avoided by checking between each step of the modelling chain and

recreating the WS22 analysis with the modified chain, but there was no directly comparable analysis for this work to indicate something was wrong. Ultimately, unifying tools would increase transparency, ease of use, and the chance of this type of error going forward. The developer of the `BGSkimmer` and Assembler tools, Dr. Dan Kodorrof[6], has made steps toward this, factorising the application of weights and cuts and moving to more explicit Assembler functionality in Python. This was extremely useful for WS24 and also motivated use of separately defined counts as an input for inference tools. In the future, `flameNEST` offers the functionality of `LZLAMA`, `BGSkimmer`, and Assembler in the same toolset as building likelihood and carrying out fits. The object-oriented nature and Python implementation make modifications transparent and easy, reducing the chance of error.

7.3 Mid-WS2024 ^{8}B CEvNS and WIMPs Prediction

Given the incorrect calculation above and early grids task force studies before defining the WS24 ROI to be blind to the ^{8}B CEvNS region, it was decided to use the preliminary information available to determine how much data would be required to obtain a 3σ observation of ^{8}B CEvNS at least 50% of the time. The author used the most up-to-date background model at the time and extrapolated the counts with live days. Two additional WIMP sensitivities were calculated for WS22 + 140 and 740 live days to check that the design requirements were met. The ROI considered, with the proposed ^{8}B blinding region, is shown in Figure 7.9.

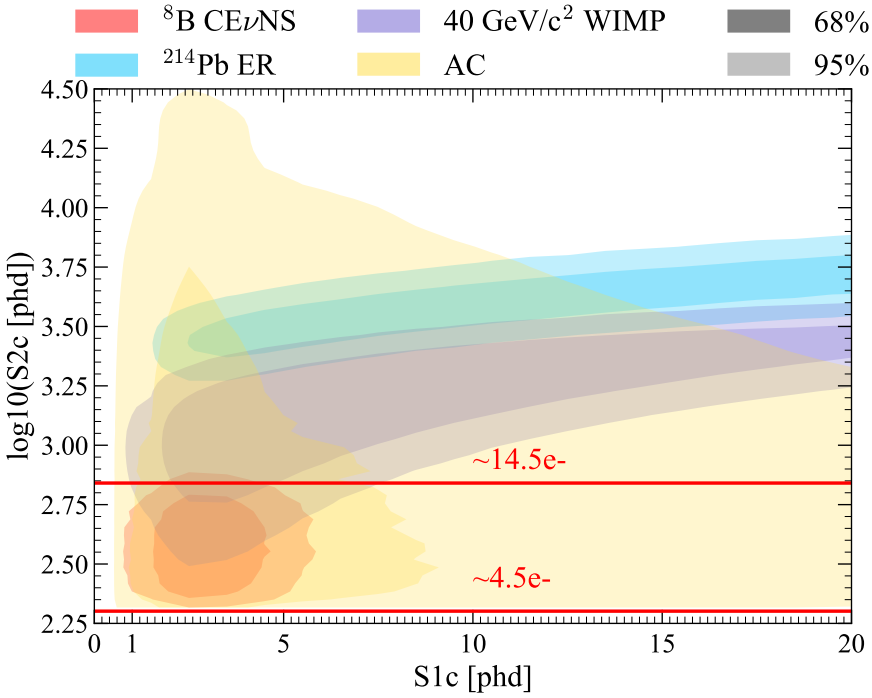


Figure 7.9: Plot of the models considered for this mid-WS2024 study.

[6]PDRA at Lawrence Berkeley National Laboratory

The ROI for this search was different from that for the previous study. Despite the 0<S1c<1 phd region having a physical explanation, events detected in a region of low S1 LCE, the threshold was raised to 1 phd. The $S2_{raw}$ threshold was defined as $>4.5e^-$ to capture as much of the $S2_{raw}$ distribution of $>5e^-$ as possible. The acceptances were updated to the best known, and the backgrounds team generated the models to avoid any error. The final ^{8}B CEνNS rate was 9.1 events/year and the AC rate was $(0.61\pm 0.09)\mu$Hz. Uncertainties on the ^{8}B CEνNS rate from neutrino flux uncertainties were ignored in this analysis. The ^{124}Xe model was fixed to its initial best fit to the salted data of $b = 1.25$ due to the interim state of template morphing. The radon tagging sideband from Section 5.2.9 was not yet implemented, so the radon tagging was omitted here with no assumption in reduction of rate or associated live time loss. The rest of the background rates were the same as those assumed in the WS24 analysis in Ref.[90], with interim acceptances, cuts, and fiducial volume estimates the actual rates would differ but are omitted here.

This study was performed in the `LZFlameFit` code developed by the author and the statistics team, described in Section 5.1. Here Rob James implemented the initial non-asymptotic method, and the author explicitly checked the asymptotic approximation and integrated it into the code, noting from the previous section that the non-asymptotic method would underestimate the median significance due to the limited number of toys, see Figure 7.10. For this test, the interim `LZFlameFit` code was not at full performance, which limited the number of toys to 5k. Fortunately, as a result of the increased background and signal rate, the asymptotic approximation was sufficient for all live days tested.

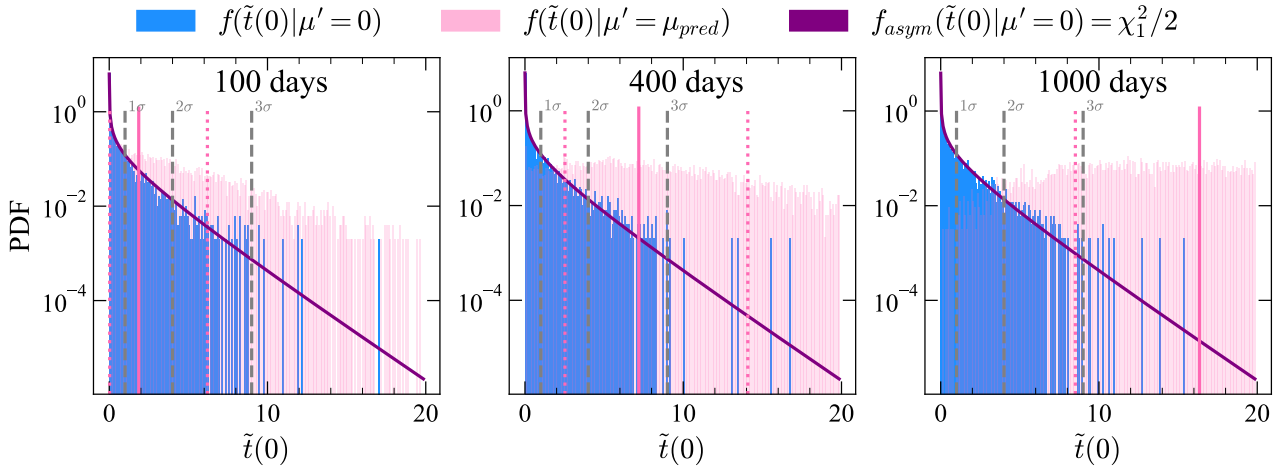


Figure 7.10: Results from the interim WS24 sensitivity study. The expected WIMP upper limits are derived using a combined likelihood with the WS22 models, not observation. The annotations and colors are identical to those in Figure 7.6 .

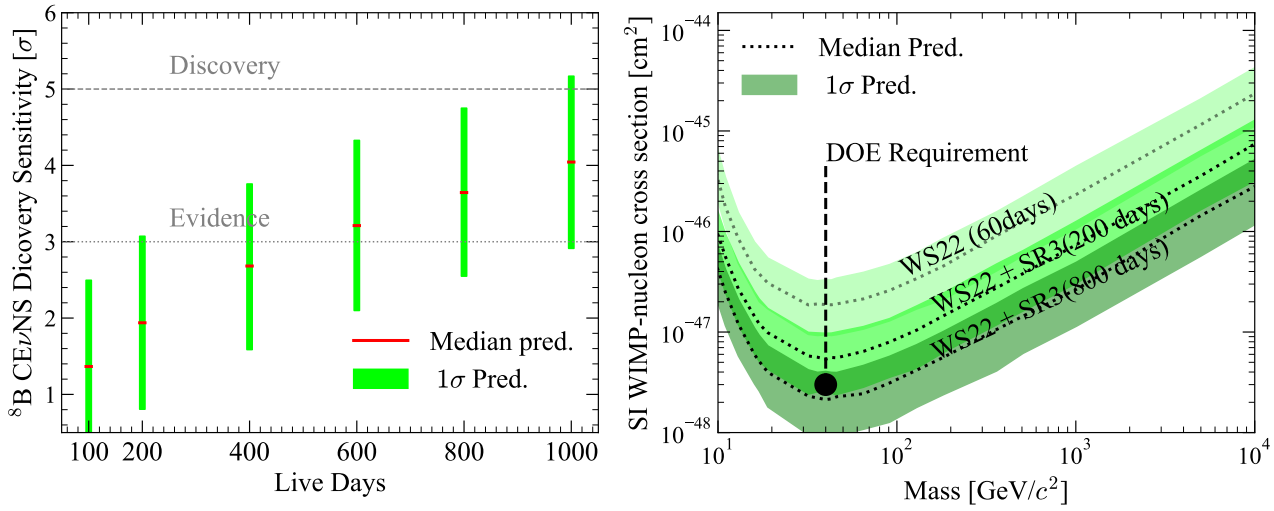


Figure 7.11: Predicted discovery significance of an ^{8}B CEνNS search vs live days accumulated, toys are generated for the quantiles but the significance of each toy is calculated asymptotically (Left). Sensitivity of a WS22 + set number of live days, evaluated non-asymptotically.

The interim predictions are shown in Figure 7.11. The inability to achieve discovery significance in ^{8}B CEvNS with the standard WIMP search ROI and techniques meant that blinding was the best option until a better analysis could be developed. The performance of the WIMP search was well within the target requirements and the 14.5e$^-$ S2$_{raw}$ threshold had a $< 5\%$ impact for WIMPs of mass $\geq$40 GeV/c^2. The author notes that this study uses an ROI where dedicated tuning for ^{8}B CEνNS was not carried out, and was preliminary for the total ROI. Further, dedicated calibrations for such a search would be required, which were not complete, which may increase the ^{8}B rate as indicated in the XENONnT study in Ref[177]. This study should be considered as the worst-case scenario for LZ low-mass WIMP and ^{8}B CEνNS searches given the detector and analysis performance at the time.

7.4 Summary

The author performed sensitivity studies investigating the impact of varying drift fields and determining the LZ run duration required to meet project requirements with legacy and current analysis tools. In studying the varying drift field and the first iteration of the runs duration study, the author created the necessary models in the simulation chain developed for LZ modifying it where necessary. In the first run-duration study, the author created a custom data-driven AC model and extrapolated information from the first WIMP search, such as acceptances. In the second runs duration study, the author used their experience to implement the required

features into the new `LZFlameFit` package for inference.

Chapter 8

Conclusion

Inference tools are essential for interpreting results and determining the expected performance of the LZ detector. With more complex and opaque tools the likelihood of human error increases as more time is spent configuring them; this was seen in the first runs duration study the author performed. The modernisation of the inference routines, factorisation of modelling chain, and strict definition of hand-offs between different working groups all help to reduce the chance of human error. Further unification of the modelling and inference tools into a transparent Object-Oriented Python package, like `flameNEST`, would help by allowing for flexible customisation and transparency. The author made significant contributions by utilising legacy tools, such as `LZStats`, and developing, as part of the stats team, new python based inference tools in `LZFlameFit`. The new tools were based on `flamedisx` utilising its template based inference infrastructure, derived from the standard MC simulation chain. The template based tools are a small part of the `flamedisx` package's functionality, originally designed to carry out explicit inference on models. That is, instead of inverse cumulative distribution sampling underlying probability distributions to produce all possible outcomes, the observed data is used to create an explicit evaluation of each probability distribution and combine them in a sum. The implementation of the `NEST` models used in LZ into `flamedisx`, called `flameNEST`, required an impractically large amount of memory as the complexity of recombination models meant that explicit inference on microphysics models, referred in the work as central block. The author modified the implementation of this calculation to improve performance and make it possible to microphysics parameters without changing the form of the `NEST` calculation. In doing a test calibration implementation the author noted that accurate determination, and interpolation, of the MC derived total rate term limits the number of possible parameters. This may be resolved through better calculation or different fitting methods, such as Markov chain MC. As a robust test of multiple models, and to confirm that a `flameNEST` can reproduce every aspect of an LZ science search, the author developed and performed a preliminary fit to data. Whilst this fit used rudimentary inputs from other analysis teams, it did show several interesting features such as better identification of radiogenic neutron background and identification of the novel ^{124}Xe DEC response. Aside from the case of Detector NRs, the added observables are

unlikely to improve the near background free low energy NR searches like WIMP searches. However, they may yet improve the WIMP search in future analyses if used to discriminate against backgrounds close to the wall allowing the expansion of the FV and increasing exposure. Although the fit times for WS24 are impractical, some very minor approximations could bring it into the realm of feasibility. Notably, fixing detector parameters to unconditional post-fit values will likely have a small impact on the result but allow for significant optimisation. After further validations, this implementation would immediately benefit the low energy ER searches as they are almost certainly asymptotic and, being background limited, the ^{214}Pb discrimination and implementation of Solar νER time dependence would greatly improve their performance. In the case of near-threshold searches, such as a low mass WIMP or ^{8}B CEνNS, the number of predicted events are significantly lower than WS24 and there are far less sources, both of which reduce computational requirements to a feasible level. An aspect of `flameNEST` is that it can now reproduce the template from MC simulation and is easy to configure which may be useful in carrying out future template based sensitivity studies as well as making it easier to report these analyses publicly.

Appendix A

FlameNEST Upgrades

A.1 Detailed Calculation

To reduce the degeneracy, one needs to understand how the quanta tensor is constructed. Firstly, the photon and electron dimensions are given from guesses propagated from the detector responses given the observables of an event. Then the quanta dimension is constructed from repeating the photon 1D tensor along the electron dimension size and vice versa, summing the two to form a quanta tensor which contains all possible combinations of photons and electrons. This means that above the explicit dimension of d_q is actually $d_{ph} \times d_{el}$. The ion dimension is constructed for each energy guess informed by the photons, electrons, and quanta.

The first step in optimisation was to produce a unique quanta dimension,d'_q calculated by `TensorFlow`'s unique function, contractions between models in the unique quanta dimensions and others were performed by an index and mapping to the original $n_{events} \times d_{el} \times d_{ph} \times d^i$ shape with the gather function. Unique ions, n_{ions}, tensors were already present for every energy, but were kept as 1D instead of expanding to the 4D shape. Alongside this, the values of number

of electrons produced in each batch were similar enough to merit using unique electrons dimension, d'_{el}, with its index as well. This means that d'_q is a reduction from $n_{events} \times d_{el} \times d_{ph}$ and n'_{el} one from $d_{events} \times d_{el}$. The latter has no physical reasoning other than events are often batched by similar reconstructed energy and therefore have similar electrons produced. Here, the limitations of 2 dimensional representation become apparent so writing the algorithm is better;

1. calculate unique n'_q and n'_{el}, and their indexing back to $n_{events} \times d_{el} \times d_{ph} \times d_i$.
2. For each energy calculate d_{ions}
3. calculate one dimensional $P^{NR}_{ions}(n_{ion}|E_i)$
4. Create a tensor $n'_q \times n_{ion}$ and calculate $P^{NR}_{quanta}(n'_q|E_i, n_{ion})$.
5. using the index for d'_q to map back to $n_{events} \times d_{el} \times d_{ph} \times d_{ions}$
6. Create tensor $d'_{el} \times d_{ion}$ and calculate $P_{electrons}(n'_{el}|E_i, n_{ion})$.
7. Using index for d'_{el}, map back to $n_{events} \times d_{el} \times d_{ph} \times d_{ions}$
8. Calculate the product $p_{mul} = P_{electrons}(n_{el}|E_i, n_{ion}) \times P^{NR}_{quanta}(n_q|E_i, n_{ion})$
9. Avoid unnecessary repeats by automatically calculate contraction over ions with tensor dot $R(n_{el}, n_{ph}, E_i) = p_{mul} \overset{\text{ions}}{\circ} P^{NR}_{ions}(n_{ion}|E_i)$

These steps would replace the orange blocks in Figure 4.5 (right). This methodology ensures that functions are only evaluated on unique elements of the quanta tensors that represent the true dimensionality of the model. Then the output values themselves are repeated. This means, for example, instead of $n_{events} \times d_{el} \times d_{ph}$ sub-graphs for $P^{NR}_{ions}(n_{ion}|E_i)$ there are only n_q plus a function for mapping the values. There is still a significant contribution to memory and time from the manipulation of these tensors; however, this is less than that of the original models and independent of the number of parameters used.

The structure of the computation for ERs is different, and follows a more intuitive path. The total quanta component of the model, $P^{ER}_{quanta}(n_q)$, is independent from the number of ions produced and the ions only comes in for the ion production, $P^{ER}_{ions}(n_{ions}|n_q)$, and in electron production as before. This means that the calculation diverges at step 3 and follows;

4. Calculate the 1D probability $P^{ER}_{quanta}(n'_q)$
5. Index P^{ER}_{quanta} from d'_q to $n_{events} \times d_{el} \times d_{ph}$.
6. Create the 2D quanta tensors of $d'_q \times d_{ions}$ and calculate $P^{ER}_{ions}(n_{ions}|n'_q)$.
7. Index P^{ER}_{ions} from $d'_q \times d_{ions}$ to $n_{events} \times d_{el} \times d_{ph} \times d_{ions}$.

8. Steps 6 & 7 from NR.
9. Calculate $p_{mul} = P_{electrons}(n_{el}|E_i, n_{ion}) \overset{\text{ions}}{\circ} P^{ER}_{ions}(n_{ions}|n'_q)$
10. Calculate $R(n_{el}, n_{ph}, E_i) = p_{mul} \times P^{ER}_{quanta}(n'_q)$.

The comparison between the performance of the ER and NR calculation is shown in Table A.1. The total peak memory usage is the same, but in the tracing window the NRs use considerably more memory. This may be due to the ERs having a simpler calculation.

Type	Batch Compute Time [iterations/s]	Peak Memory [GiBs]	
		Pre-Window	In-Window
ER	2.44±0.02	20.0	4.44
NR	3.03±0.08	20.0	8.33

Table A.1: Comparing NR and ER performance in current FlameNEST version for a batch size of 50 events. The peak memory is prior to the profiling window, likely due to memory.

A.2 Non-Uniform Microphysics Case

The covered performance is for a microphysics response, therefore, the CB functions are the same for all events in a batch, only varying in parameters and domain values. When a variable drift field is present, the CB response models depend on the position of each event, and so some modifications need to be made. That is, unique quanta must be expressed in $n_{events} \times d_{n'_q}$ and unique electrons no longer have meaning. After calculating the quanta, the dimensionality of the electrons can still be reduced by removing the photon dimension that does not appear in the subsequent calculations. The calculation is similar to the above, except that the n_{events} dimension is not reduced, no indexing is required on d_{el}, and some parameters of the models now have dimensionality of n_{events}, which requires some special handling.

Type With Field	Batch Compute Time [iterations/s]	Peak Memory [GiBs]	
		Pre-Window	In-Window
ER	1.71±0.03	19.1	19.1
NR	2.34±0.05	20.4	9.33

Table A.2: Comparing NR and ER performance in FlameNEST with field dependence in the CB for a batch size of 50 events. Though there is no change in peak memory, there is a significant increase of in-window memory for ERs. The author observed this was due to transient spikes.

It appears that this has a greater impact on ER models than on NR models. The computation time actually decreases for a non-uniform field, this is because there is no unique or gather functions operating on the electron quanta tensor. The memory consumption of the ERs increases in-window significantly. This still out-performs original FlameNEST, especially for

with CB parameters. This indicates that there were many degenerate values of guess electrons produced for ERs, which may be caused by their large charge yields.

Appendix B

LZFlameFit Package

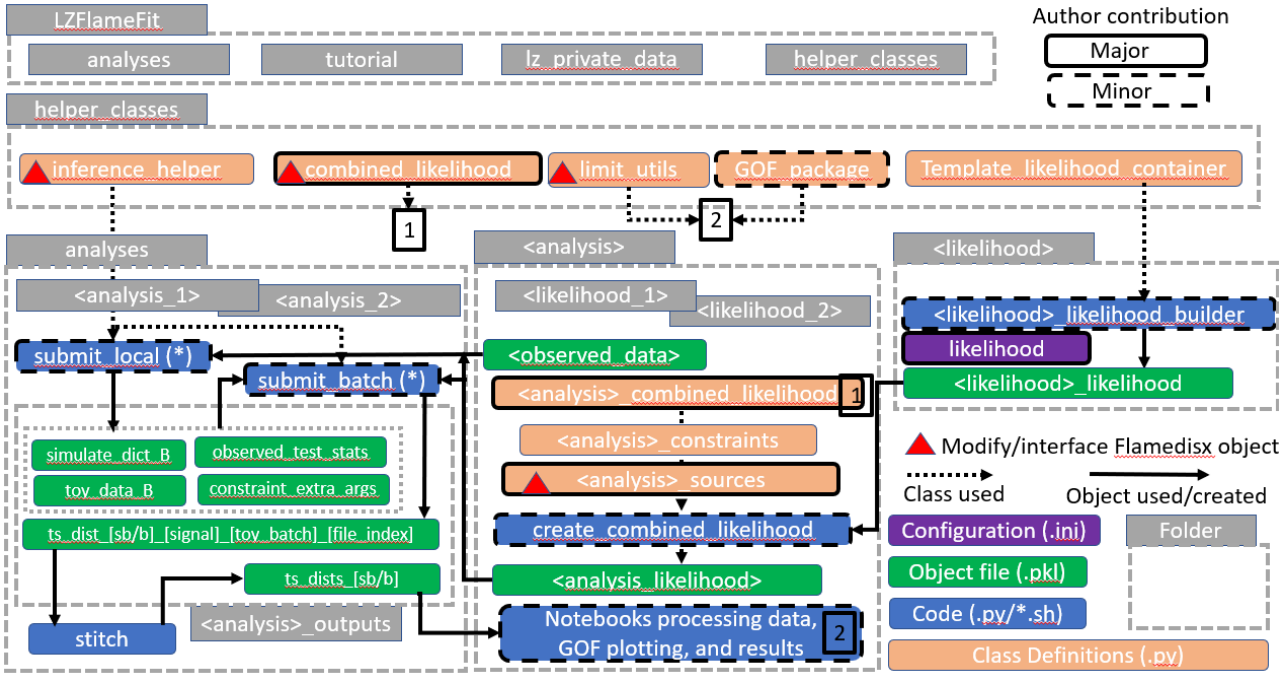


Figure B.1: The LZFlameFit package, the creation of a likelihood flows from right to left and the use of classes are shown with dotted arrows. The borders indicate where the author made contributions beyond quality assurance, debugging and use.

A diagram of the LZFlameFit package is given in Figure B.1. The author attempts to indicate their contributions. In general an analysis requires first, the construction of the relevant sub-likelihoods with the contributing counts and rates. Next the combined likelihood class and sources for the analysis give which likelihoods are combined, how their counts are combined, and how theor parameters are shared between each sub-likelihood. This is based

upon a combined likelihood base class the author designed and Dr Robert James generalised, which manages multiple `flamedisx` likelihood objects and manages distributing parameters and combining gradients and Hessians. Within the analysis there is also the custom constraints class that can hold special sampling routines for each parameter as well as their constraint terms, whose centres may be shifted in the inference routine. There are also custom notebooks for carrying out GOF tests, based on the GOF package built by Leah Wolf and Dr Joe McLaughlin. The local submission script uses `flamedisx` and the inference helper to derive the observed test statistics given data, or parses the priors in the case of sensitivity, and generates the necessary background only data sets. Then the batch submission script uses these outputs to generate the necessary test-statistic distributions for discovery or confidence intervals. The limit utils package is then used to process the results.

Appendix C

Rates in Sensitivity Studies

For future reference the rates assumed in the sensitivity studies of Chapter 7 are noted here.

Changing LZ Drift Field

These numbers were based on rates from the WS22 analysis, but as the NEST models were drift field dependent for WS22 the rates in Table C.1 marginally change over the changing configurations. The SR1 configuration is consistent with the values quoted [95], which can be checked by scaling these values by a factor of $\frac{60}{365}$.

	Rates for configurations [evts/year]			
	SR1	-9 kV	-15 kV	-20 kV
^{8}B CEνNS	0.87 ± 0.05	0.97 ± 0.06	1.05 ± 0.06	1.08 ± 0.06
Solar νER	165 ± 9.9	158 ± 9.5	166 ± 10	171 ± 10
^{136}Xe	92 ± 15	79 ± 13	86.6 ± 14	91.8 ± 15
Detector NR.	0.01 ± 0.24	0.01 ± 0.24	0.01 ± 0.24	0.01 ± 0.24
^{124}Xe	30.4 ± 8.5	32.2 ± 9.0	32.5 ± 9.1	32.5 ± 9.1
Acccidentals	7.4 ± 1.6			
^{85}Kr	196 ± 31	189 ± 30	198 ± 32	204 ± 33
^{214}Pb	997 ± 210	955 ± 200	1003 ± 210	1037 ± 220
^{212}Pb	107 ± 29	102 ± 28	108 ± 29	111 ± 30
Detector ER	1.46 ± 0.39	1.46 ± 0.39	1.46 ± 0.39	1.46 ± 0.39
Total	1597 ± 214	1604 ± 206	1658 ± 216	1701 ± 223
β-Decays + Det ER.	1301 ± 210	1248 ± 200	1311 ± 220	1354 ± 220

Table C.1: Rates for the sensitivity study evaluating LZ's field configurations. Note that the accidentals model is data-driven and was excluded from the study as it was not possible to construct a model for each configuration.

LZ Run Duration

The rates for various sources considered in the run duration study are shown in Table C.2. This contains the error in the efficiencies discussed in Section 7.2. By comparing the Goal and Conservative scenarios, one may note that ^{212}Pb was assumed to be reduced by the radon tagging. As discussed in Section 5.2.1.3, this was an incorrect assumption. Due to ^{212}Pb being subdominant, this mistake likely has no impact on the results.

Source	Runs Duration Rates [evts/year]	
	Conservative	Goal
Accidentals	14.6 ± 3.1	3.66 ± 0.77
Beta	1240 ± 200	504 ± 67
^{212}Pb	102 ± 27	30.5 ± 8.2
^{214}Pb	950 ± 200	285 ± 60
Detector ER	1.38 ± 0.37	
Detector NR	0.36 ± 0.21	
^{85}Kr	187 ± 30	
Solar νER	156.6 ± 9.4	
^{124}Xe	31.0 ± 8.7	
^{136}Xe	83 ± 13	
^{8}B CEνNS	4.13	
^{8}B CEνNS	4.13	
WIMP 11 GeV/c^2	89.38	
WIMP 40 GeV/c^2	970	
WIMP 1008 GeV/c^2	85	

Table C.2: Rates for the first runs duration study. In large part the ER backgrounds are inconsequential to the ^{8}B CEν search, those that are of importance are in bold. The WIMP rates are quoted for a standard cross-section of $\sigma_N^{SI} = 10^{-45}$ cm^2. As events are quoted in evts/year the application of a 10% livetime hit in the goal scenario is not applied here.

Mid-WS2024 ^{8}B CEvNS and WIMPs Prediction

The assumed rates for the Mid-WS2024 study are given in Table C.3. Whilst there is an uncertainty of the ^{8}B flux this is not included in frequentist statistics, as the signal is unconstrained. In the future it may be ideal to implement this in sampling the background with signal toys for predicted discovery sensitivity, as it folds in the prior knowledge. Either way, the fractional uncertainty in rate will approximately translate to a shift in the live time required to achieve a specific significance. Further, an improved study quantifying uncertainties on efficiencies would be required to properly quantify the uncertainty on the near-threshold rates.

Row	rate [evts/year]
^{214}Pb	1179 ± 61
^{212}Pb	65.3 ± 8.2
^{85}Kr	306 ± 49
^{136}Xe	93 ± 15
^{124}Xe	32 ± 5.4
Solar νER	172 ± 12
^{8}B CEνNS	9.11 ± 0.55
Accidentals	19.5 ± 2.9
Detector NR	0.44 ± 0.99
WIMP 11 GeV/c^2	77.52
WIMP 40 GeV/c^2	892.31
WIMP 1008 GeV/c^2	89.24

Table C.3: Assumed Preliminary WS2024 rates. The WIMP rates assume a cross-section of $\sigma_N^{SI} = 10^{-45}$ cm^2.

Bibliography

[1] J. Neyman, E. S. Pearson, and K. Pearson, "Ix. on the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, no. 694-706, pp. 289–337, 1933. DOI: 10.1098/rsta.1933.0009. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1933.0009. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1933.0009.

[2] G. Bertone and D. Hooper, "History of dark matter," *Rev. Mod. Phys.*, vol. 90, p. 045 002, 4 Oct. 2018. DOI: 10.1103/RevModPhys.90.045002. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.90.045002.

[3] F. Zwicky, "Republication of: The redshift of extragalactic nebulae," *General Relativity and Gravitation*, vol. 41, no. 1, pp. 207–224, 2009, ISSN: 1572-9532. DOI: 10.1007/s10714-008-0707-4. [Online]. Available: https://doi.org/10.1007/s10714-008-0707-4.

[4] F. Zwicky, "On the Masses of Nebulae and of Clusters of Nebulae," *ApJ*, vol. 86, p. 217, Oct. 1937. DOI: 10.1086/143864.

[5] V. C. Rubin and W. K. Ford Jr., "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions," *ApJ*, vol. 159, p. 379, Feb. 1970. DOI: 10.1086/150317.

[6] M. S. Roberts, "A High-Resolution 21-CM Hydrogen-Line Survey of the Andromeda Nebula," *ApJ*, vol. 144, p. 639, May 1966. DOI: 10.1086/148645.

[7] K. G. Begeman, A. H. Broeils, and R. H. Sanders, "Extended rotation curves of spiral galaxies : dark haloes and modified dynamics.," *MNRAS*, vol. 249, p. 523, Apr. 1991. DOI: 10.1093/mnras/249.3.523.

[8] N. Aghanim and others., "Planck2018 results: Vi. cosmological parameters," *Astronomy & Astrophysics*, vol. 641, A6, Sep. 2020, ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833910. [Online]. Available: http://dx.doi.org/10.1051/0004-6361/201833910.

[9] S. S. Kashyap, "CMBAnalysis: A Modern Framework for High-Precision Cosmic Microwave Background Analysis," *arXiv preprint arXiv:2411.12207 [astro-ph.CO]*, 2024, Accessed August 28, 2025.

[10] A. Lewis, A. Challinor, and A. Lasenby, *CAMB: Code for anisotropies in the microwave background*, Astrophysics Source Code Library, ascl:1102.026, 2000. [Online]. Available: https://camb.info.

[11] M. Tanabashi and P. D. Group, "Review of Particle Physics (2019 update)," *Physical Review D*, vol. 98, no. 3, p. 030 001, 2018, 2019 update; as listed by the Particle Data Group.

[12] T. M. C. Abbott et al., "First cosmology results using type ia superNovae from the dark energy survey: Constraints on cosmological parameters," *The Astrophysical Journal Letters*, vol. 872, no. 2, p. L30, Feb. 2019, ISSN: 2041-8213. DOI: 10.3847/2041-8213/ab04fa. [Online]. Available: http://dx.doi.org/10.3847/2041-8213/ab04fa.

[13] S. Chabanier, M. Millea, and N. Palanque-Delabrouille, "Matter power spectrum: From lyα forest to cmb scales," *Monthly Notices of the Royal Astronomical Society*, vol. 489, no. 2, pp. 2247–2253, Aug. 2019, ISSN: 1365-2966. DOI: 10.1093/mnras/stz2310. [Online]. Available: http://dx.doi.org/10.1093/mnras/stz2310.

[14] R. E. Angulo and O. Hahn, "Large-scale dark matter simulations," *Living Reviews in Computational Astrophysics*, vol. 8, no. 1, p. 1, Feb. 2022, ISSN: 2365-0524. DOI: 10.1007/s41115-021-00013-z. [Online]. Available: https://doi.org/10.1007/s41115-021-00013-z.

[15] M. Bradač et al., "Revealing the properties of dark matter in the merging cluster macs j0025.4–1222*," *The Astrophysical Journal*, vol. 687, no. 2, p. 959, Nov. 2008. DOI: 10.1086/591246. [Online]. Available: https://dx.doi.org/10.1086/591246.

[16] D. Clowe et al., "A direct empirical proof of the existence of dark matter," *The Astrophysical Journal*, vol. 648, no. 2, pp. L109–L113, Aug. 2006, ISSN: 1538-4357. DOI: 10.1086/508162. [Online]. Available: http://dx.doi.org/10.1086/508162.

[17] D. Walsh*, R. F. Carswell†, and R. J. Weymann‡, "0957 + 561 a, b: Twin quasistellar objects or gravitational lens?" *Nature*, vol. 279, no. 5712, pp. 381–384, May 1979, ISSN: 1476-4687. DOI: 10.1038/279381a0. [Online]. Available: https://doi.org/10.1038/279381a0.

[18] Kleinheinrich, M. et al., "Weak lensing measurements of dark matter halos of galaxies from combo-17," *A&A*, vol. 455, no. 2, pp. 441–451, 2006. DOI: 10.1051/0004-6361:20042606. [Online]. Available: https://doi.org/10.1051/0004-6361:20042606.

[19] H. Niikura et al., "Microlensing constraints on primordial black holes with subaru/hsc andromeda observations," *Nature Astronomy*, vol. 3, no. 6, pp. 524–534, 2019, ISSN: 2397-3366. DOI: 10.1038/s41550-019-0723-1. [Online]. Available: https://doi.org/10.1038/s41550-019-0723-1.

[20] D. Croon, D. McKeen, N. Raj, and Z. Wang, "Subaru-hsc through a different lens: Microlensing by extended dark matter structures," *Phys. Rev. D*, vol. 102, p. 083 021, 8 Oct. 2020. DOI: 10.1103/PhysRevD.102.083021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.102.083021.

[21] K. G. Begeman, A. H. Broeils, and R. H. Sanders, "Extended rotation curves of spiral galaxies : dark haloes and modified dynamics.," *MNRAS*, vol. 249, p. 523, Apr. 1991. DOI: 10.1093/mnras/249.3.523.

[22] K.-H. Chae, "Robust evidence for the breakdown of standard gravity at low acceleration from statistically pure binaries free of hidden companions," *The Astrophysical Journal*, vol. 960, no. 2, p. 114, Jan. 2024. DOI: 10.3847/1538-4357/ad0ed5. [Online]. Available: https://dx.doi.org/10.3847/1538-4357/ad0ed5.

[23] I. Banik et al., "Strong constraints on the gravitational law from gaia dr3 wide binaries," *Monthly Notices of the Royal Astronomical Society*, vol. 527, no. 3, pp. 4573–4615, Nov. 2023, ISSN: 1365-2966. DOI: 10.1093/mnras/stad3393. [Online]. Available: http://dx.doi.org/10.1093/mnras/stad3393.

[24] X. Hernandez, K.-H. Chae, and A. Aguayo-Ortiz, "A critical review of recent gaia wide binary gravity tests," *Monthly Notices of the Royal Astronomical Society*, vol. 533, no. 1, pp. 729–742, 2024, https://doi.org/10.1093/mnras/stae1823. DOI: 10.1093/mnras/stae1823. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1174196609.

[25] European Space Agency – Gaia Data Processing and Analysis Consortium, *Gaia data release scenario*, Data release, Retrieved September 9, 2025, from https://www.cosmos.esa.int/web/gaia/release, 2022.

[26] P. A. Cano, "Inconsistency of modified gravity in cosmology," *Physics Letters B*, vol. 858, p. 139 059, 2024, ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2024.139059. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370269324006178.

[27] S. R. Green and R. M. Wald, "How well is our universe described by an flrw model?" *Classical and Quantum Gravity*, vol. 31, no. 23, p. 234 003, Nov. 2014, ISSN: 1361-6382. DOI: 10.1088/0264-9381/31/23/234003. [Online]. Available: http://dx.doi.org/10.1088/0264-9381/31/23/234003.

[28] B. Famaey and S. S. McGaugh, "Modified newtonian dynamics (mond): Observational phenomenology and relativistic extensions," *Living Reviews in Relativity*, vol. 15, no. 1, Sep. 2012, ISSN: 1433-8351. DOI: 10.12942/lrr-2012-10. [Online]. Available: http://dx.doi.org/10.12942/lrr-2012-10.

[29] J. R. Brownstein and J. W. Moffat, "The bullet cluster 1e0657-558 evidence shows modified gravity in the absence of dark matter," *Monthly Notices of the Royal Astronomical Society*, vol. 382, no. 1, pp. 29–47, Oct. 2007, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2007.12275.x. eprint: https://academic.oup.com/mnras/article-pdf/382/1/29/3057374/mnras0382-0029.pdf. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2007.12275.x.

[30] Z. Xie, B. Liu, J. Liu, Y.-F. Cai, and R. Yang, "Limits on the primordial black holes dark matter with future mev detectors," *Phys. Rev. D*, vol. 109, p. 043 020, 4 Feb. 2024. DOI: 10.1103/PhysRevD.109.043020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.109.043020.

[31] B. J. Kavanagh, *Pbhbounds: Release version*, version 1.0, Nov. 2019. DOI: 10.5281/zenodo.3538999. [Online]. Available: https://doi.org/10.5281/zenodo.3538999.

[32] K. Freese, "Death of baryonic dark matter," *Physics Reports*, vol. 333-334, pp. 183–201, 2000, ISSN: 0370-1573. DOI: https://doi.org/10.1016/S0370-1573(00)00022-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370157300000223.

[33] G. Facchinetti, M. Lucca, and S. Clesse, "Relaxing cmb bounds on primordial black holes: The role of ionization fronts," *Physical Review D*, vol. 107, no. 4, Feb. 2023, ISSN: 2470-0029. DOI: 10.1103/physrevd.107.043537. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.107.043537.

[34] T. Boybeyi, S. Clesse, S. Kuroyanagi, and M. Sakellariadou, "Search for a gravitational wave background from primordial black hole binaries using data from the first three ligo-virgo-kagra observing runs," *Phys. Rev. D*, vol. 112, p. 023 551, 2 Jul. 2025. DOI: 10.1103/zphk-3ld9. [Online]. Available: https://link.aps.org/doi/10.1103/zphk-3ld9.

[35] S. D. McDermott, H.-B. Yu, and K. M. Zurek, "Turning off the lights: How dark is dark matter?" *Phys. Rev. D*, vol. 83, p. 063 509, 6 Mar. 2011. DOI: 10.1103/PhysRevD.83.063509. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.83.063509.

[36] P. Huet and E. Sather, "Electroweak baryogenesis and standard model cp violation," *Phys. Rev. D*, vol. 51, pp. 379–394, 2 Jan. 1995. DOI: 10.1103/PhysRevD.51.379. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.51.379.

[37] N. Craig, *Naturalness: A snowmass white paper*, 2022. arXiv: 2205.05708 [hep-ph]. [Online]. Available: https://arxiv.org/abs/2205.05708.

[38] A. S. Chou et al., *Snowmass cosmic frontier report*, 2022. arXiv: 2211.09978 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2211.09978.

[39] K. A. et al., "Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations," *Nature*, vol. 580, no. 7803, pp. 339–344, 2020. DOI: 10.1038/s41586-020-2177-0. [Online]. Available: https://doi.org/10.1038/s41586-020-2177-0.

[40] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints," *Physics Reports*, vol. 405, no. 5–6, pp. 279–390, Feb. 2005, ISSN: 0370-1573. DOI: 10.1016/j.physrep.2004.08.031. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2004.08.031.

[41] S. Davidson, E. Nardi, and Y. Nir, "Leptogenesis," *Physics Reports*, vol. 466, no. 4–5, pp. 105–177, Sep. 2008, ISSN: 0370-1573. DOI: 10.1016/j.physrep.2008.06.002. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2008.06.002.

[42] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of pseudoparticles," *Phys. Rev. Lett.*, vol. 38, pp. 1440–1443, 25 Jun. 1977. DOI: 10.1103/PhysRevLett.38.1440. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.38.1440.

[43] P. Sikivie, "Axion dark matter," *Nuclear Physics B*, vol. 1003, p. 116 500, 2024, Special Issue of Nobel Symposium 182 on Dark Matter, ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2024.116500. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S055032132400066X.

[44] C. B. Adams et al., *Axion dark matter*, 2023. arXiv: 2203.14923 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2203.14923.

[45] K. A. Beyer, G. Marocco, R. Bingham, and G. Gregori, "Light-shining-through-wall axion detection experiments with a stimulating laser," *Phys. Rev. D*, vol. 105, p. 035 031, 3 Feb. 2022. DOI: 10.1103/PhysRevD.105.035031. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.105.035031.

[46] K. Arisaka et al., "Expected sensitivity to galactic/solar axions and bosonic super-wimps based on the axio-electric effect in liquid xenon dark matter detectors," *Astroparticle Physics*, vol. 44, pp. 59–67, 2013, ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.12.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927650512002344.

[47] "Global fits of gut-scale susy models with gambit," *The European Physical Journal C*, vol. 77, no. 12, Dec. 2017, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5167-0. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-017-5167-0.

[48] K. Bechtol et al., *Snowmass2021 cosmic frontier white paper: Dark matter physics from halo measurements*, 2023. arXiv: 2203.07354 [hep-ph]. [Online]. Available: https://arxiv.org/abs/2203.07354.

[49] J. S. Bullock and M. Boylan-Kolchin, "Small-scale challenges to the Λcdm paradigm," *Annual Review of Astronomy and Astrophysics*, vol. 55, no. Volume 55, 2017, pp. 343–387, 2017, ISSN: 1545-4282. DOI: https://doi.org/10.1146/annurev-astro-091916-055313. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-astro-091916-055313.

[50] H. Deng, M. P. Hertzberg, M. H. Namjoo, and A. Masoumi, "Can light dark matter solve the core-cusp problem?" *Phys. Rev. D*, vol. 98, p. 023 513, 2 Jul. 2018. DOI: 10.1103/PhysRevD.98.023513. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.98.023513.

[51] A. Tumasyan et al., "Search for invisible decays of the higgs boson produced via vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev. D*, vol. 105, p. 092 007, 9 May 2022. DOI: 10.1103/PhysRevD.105.092007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.105.092007.

[52] A. De Roeck, "Dark matter searches at accelerators," *Nuclear Physics B*, vol. 1003, p. 116 480, 2024, Special Issue of Nobel Symposium 182 on Dark Matter, ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2024.116480. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0550321324000464.

[53] L. Sahoo, "Probing sterile neutrinos using decay width measurements at colliders," *International Journal of Theoretical Physics*, vol. 61, no. 7, p. 200, Jul. 2022, ISSN: 1572-9575. DOI: 10.1007/s10773-022-05194-8. [Online]. Available: https://doi.org/10.1007/s10773-022-05194-8.

[54] D. P. Aguillard et al., "Measurement of the positive muon anomalous magnetic moment to 127 ppb," *Phys. Rev. Lett.*, vol. 135, p. 101 802, 10 Sep. 2025. DOI: 10.1103/7clf-sm2v. [Online]. Available: https://link.aps.org/doi/10.1103/7clf-sm2v.

[55] A. Crivellin and B. Mellado, "Anomalies in particle physics and their implications for physics beyond the standard model," *Nature Reviews Physics*, vol. 6, no. 5, pp. 294–309, May 2024, ISSN: 2522-5820. DOI: 10.1038/s42254-024-00703-6. [Online]. Available: https://doi.org/10.1038/s42254-024-00703-6.

[56] T. Daylan et al., "The characterization of the gamma-ray signal from the central milky way: A case for annihilating dark matter," *Physics of the Dark Universe*, vol. 12, pp. 1–23, 2016, ISSN: 2212-6864. DOI: https://doi.org/10.1016/j.dark.2015.12.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212686416000030.

[57] M. Ackermann et al., "The fermi galactic center gev excess and implications for dark matter," *The Astrophysical Journal*, vol. 840, no. 1, p. 43, May 2017. DOI: 10.3847/1538-4357/aa6cab. [Online]. Available: https://dx.doi.org/10.3847/1538-4357/aa6cab.

[58] M. M. Muru et al., "Fermi-lat galactic center excess morphology of dark matter in simulations of the milky way galaxy," *Phys. Rev. Lett.*, vol. 135, p. 161 005, 16 Oct. 2025. DOI: 10.1103/g9qz-h8wd. [Online]. Available: https://link.aps.org/doi/10.1103/g9qz-h8wd.

[59] J. Conrad and O. Reimer, "Indirect dark matter searches in gamma and cosmic rays," *Nature Physics*, vol. 13, no. 3, pp. 224–231, Mar. 2017, ISSN: 1745-2481. DOI: 10.1038/nphys4049. [Online]. Available: https://doi.org/10.1038/nphys4049.

[60] S. Abdollahi et al., "Combined dark matter search towards dwarf spheroidal galaxies with fermi-lat, hawc, h.e.s.s., magic, and veritas," *Journal of Cosmology and Astroparticle Physics*, vol. 2026, no. 03, p. 035, Mar. 2026. DOI: 10.1088/1475-7516/2026/03/035. [Online]. Available: https://doi.org/10.1088/1475-7516/2026/03/035.

[61] M. G. Aartsen et al., "All-flavour search for neutrinos from dark matter annihilations in the milky way with icecube/deepcore," *The European Physical Journal C*, vol. 76, no. 10, p. 531, Sep. 2016, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4375-3. [Online]. Available: https://doi.org/10.1140/epjc/s10052-016-4375-3.

[62] M. Filzinger et al., "Ultralight dark matter search with space-time separated atomic clocks and cavities," *Phys. Rev. Lett.*, vol. 134, p. 031 001, 3 Jan. 2025. DOI: 10.1103/PhysRevLett.134.031001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.134.031001.

[63] J. Aalbers et al., "Search for new physics in low-energy electron recoils from the first lz exposure," *Physical Review D*, vol. 108, no. 7, Oct. 2023, ISSN: 2470-0029. DOI: 10.1103/physrevd.108.072006. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.108.072006.

[64] S. E. Vahsen et al., *Cygnus: Feasibility of a nuclear recoil observatory with directional sensitivity to dark matter and neutrinos*, 2020. arXiv: 2008.12587 [physics.ins-det]. [Online]. Available: https://arxiv.org/abs/2008.12587.

[65] I. Giomataris et al., "A Novel large-volume spherical detector with proportional amplification read-out," *Journal of Instrumentation*, vol. 3, no. 09, P09007, Sep. 2008. DOI: 10.1088/1748-0221/3/09/P09007. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/3/09/P09007.

[66] A. Aguilar-Arevalo et al., "Results on low-mass weakly interacting massive particles from an 11 kg d target exposure of damic at snolab," *Phys. Rev. Lett.*, vol. 125, p. 241 803, 24 Dec. 2020. DOI: 10.1103/PhysRevLett.125.241803. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.125.241803.

[67] P. Adari et al., "First direct-detection results on sub-gev dark matter using the sensei detector at snolab," *Physical Review Letters*, vol. 134, no. 1, Feb. 2025, ISSN: 1079-7114. DOI: 10.1103/physrevlett.134.011804. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.134.011804.

[68] N. Carlin et al., "Combined annual modulation dark matter search with cosine-100 and anais-112," *Phys. Rev. Lett.*, vol. 135, p. 121 002, 12 Sep. 2025. DOI: 10.1103/9j7w-qp1c. [Online]. Available: https://link.aps.org/doi/10.1103/9j7w-qp1c.

[69] P.-A. Amaudruz et al., "Design and construction of the deap-3600 dark matter detector," *Astroparticle Physics*, vol. 108, pp. 1–23, 2019, ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2018.09.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927650518300914.

[70] P. Adhikari et al., "Pulse-shape discrimination against low-energy ar-39 beta decays in liquid argon with 4.5 tonne-years of deap-3600 data," *The European Physical Journal C*, vol. 81, no. 9, p. 823, Sep. 2021, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09514-w. [Online]. Available: https://doi.org/10.1140/epjc/s10052-021-09514-w.

[71] D. Akerib et al., "The lux-zeplin (lz) experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 953, p. 163 047, 2020, ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2019.163047. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900219314032.

[72] E. Aprile et al., "The xenonnt dark matter experiment," *The European Physical Journal C*, vol. 84, no. 8, p. 784, Aug. 2024, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-12982-5. [Online]. Available: https://doi.org/10.1140/epjc/s10052-024-12982-5.

[73] Z. Bo et al., "Dark matter search results from 1.54 Tonne·Year exposure of pandax-4t," *Phys. Rev. Lett.*, vol. 134, p. 011 805, 1 Jan. 2025. DOI: 10.1103/PhysRevLett.134.011805. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.134.011805.

[74] F. Acerbi et al., "Darkside-20k sensitivity to light dark matter particles," *Communications Physics*, vol. 7, no. 1, p. 422, Dec. 2024, ISSN: 2399-3650. DOI: 10.1038/s42005-024-01896-z. [Online]. Available: https://doi.org/10.1038/s42005-024-01896-z.

[75] C. Amole et al., "Dark matter search results from the pico-60 CF_3I bubble chamber," *Phys. Rev. D*, vol. 93, p. 052 014, 5 Mar. 2016. DOI: 10.1103/PhysRevD.93.052014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.93.052014.

[76] R. Agnese et al., "Projected sensitivity of the supercdms snolab experiment," *Physical Review D*, vol. 95, no. 8, Apr. 2017, ISSN: 2470-0029. DOI: 10.1103/physrevd.95.082002. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.95.082002.

[77] G. Angloher et al., "First observation of single photons in a cresst detector and new dark matter exclusion limits," *Phys. Rev. D*, vol. 110, p. 083 038, 8 Oct. 2024. DOI: 10.1103/PhysRevD.110.083038. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.110.083038.

[78] R. Anthony-Petersen et al., "Demonstration of the herald superfluid helium detector concept," *Physical Review D*, vol. 110, no. 7, Oct. 2024, ISSN: 2470-0029. DOI: 10.1103/physrevd.110.072006. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.110.072006.

[79] D. Baxter et al., "First demonstration of a scintillating xenon bubble chamber for detecting dark matter and coherent elastic neutrino-nucleus scattering," *Phys. Rev. Lett.*, vol. 118, p. 231 301, 23 Jun. 2017. DOI: 10.1103/PhysRevLett.118.231301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.118.231301.

[80] C. Levy et al., "Xenon bubble chambers for direct dark matter detection," *Journal of Instrumentation*, vol. 11, no. 03, p. C03003, Mar. 2016. DOI: 10.1088/1748-0221/11/03/C03003. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/11/03/C03003.

[81] L. Balogh et al., *Exploring light dark matter with the darksphere spherical proportional counter electroformed underground at the boulby underground laboratory*, 2023. arXiv: 2301.05183 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2301.05183.

[82] C. Hare, *Directdetectionplots: Plots for dm-nucleus and dm-electron scattering experiments*, https://github.com/cajohare/DirectDetectionPlots, GitHub repository, accessed 2025-10-08, 2025.

[83] D. S. Akerib et al., *Snowmass2021 cosmic frontier dark matter direct detection to the neutrino fog*, 2022. arXiv: 2203.08084 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2203.08084.

[84] S. E. Vahsen et al., *Cygnus: Feasibility of a nuclear recoil observatory with directional sensitivity to dark matter and neutrinos*, 2020. arXiv: 2008.12587 [physics.ins-det]. [Online]. Available: https://arxiv.org/abs/2008.12587.

[85] D. Baxter et al., "Recommended conventions for reporting results from direct dark matter searches," *The European Physical Journal C*, vol. 81, no. 10, p. 907, Oct. 2021, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09655-y. [Online]. Available: https://doi.org/10.1140/epjc/s10052-021-09655-y.

[86] J. Engel, "Nuclear form factors for the scattering of weakly interacting massive particles," *Physics Letters B*, vol. 264, no. 1, pp. 114–119, 1991, ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(91)90712-Y. [Online]. Available: https://www.sciencedirect.com/science/article/pii/037026939190712Y.

[87] G. Dūda, A. Kemper, and P. Gondolo, "Model-independent form factors for spin-independent neutralino–nucleon scattering from elastic electron scattering data," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 04, p. 012, Apr. 2007. DOI: 10.1088/1475-7516/2007/04/012. [Online]. Available: https://doi.org/10.1088/1475-7516/2007/04/012.

[88] J. Vergados, H. Ejiri, and K. Savvidy, "Theoretical direct wimp detection rates for inelastic scattering to excited states," *Nuclear Physics B*, vol. 877, no. 1, pp. 36–50, 2013, ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2013.09.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0550321313004859.

[89] D. Smith and N. Weiner, "Inelastic dark matter," *Phys. Rev. D*, vol. 64, p. 043 502, 4 Jul. 2001. DOI: 10.1103/PhysRevD.64.043502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.64.043502.

[90] J. Aalbers et al., "Dark matter search results from 4.2 Tonne−Years of exposure of the lux-zeplin (lz) experiment," *Phys. Rev. Lett.*, vol. 135, p. 011 802, 1 Jul. 2025. DOI: 10.1103/4dyc-z8zf. [Online]. Available: https://link.aps.org/doi/10.1103/4dyc-z8zf.

[91] C. A. J. O'Hare, "New definition of the neutrino floor for direct dark matter searches," *Phys. Rev. Lett.*, vol. 127, p. 251 802, 25 Dec. 2021. DOI: 10.1103/PhysRevLett.127.251802. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.251802.

[92] J. Billard et al., "Direct detection of dark matter—appec committee report," *Reports on Progress in Physics*, vol. 85, no. 5, p. 056 201, Apr. 2022, ISSN: 1361-6633. DOI: 10.1088/1361-6633/ac5754. [Online]. Available: http://dx.doi.org/10.1088/1361-6633/ac5754.

[93] R. Bernabei et al., "Results from the dama/libra experiment," *Journal of Physics: Conference Series*, vol. 203, no. 1, p. 012 003, Jan. 2010. DOI: 10.1088/1742-6596/203/1/012003. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/203/1/012003.

[94] R. Itay et al., "Direxeno—an apparatus for measuring correlated scintillation signatures in liquid xenon," *Journal of Instrumentation*, vol. 15, no. 06, T06001, Jun. 2020. DOI: 10.1088/1748-0221/15/06/T06001. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/15/06/T06001.

[95] J. Aalbers et al., "First dark matter search results from the lux-zeplin (lz) experiment," *Phys. Rev. Lett.*, vol. 131, p. 041 002, 4 Jul. 2023. DOI: 10.1103/PhysRevLett.131.041002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.131.041002.

[96] B. J. Mount et al., *Lux-zeplin (lz) technical design report*, 2017. arXiv: 1703.09144 [physics.ins-det]. [Online]. Available: https://arxiv.org/abs/1703.09144.

[97] F. Neves et al., "Measurement of the absolute reflectance of polytetrafluoroethylene (ptfe) immersed in liquid xenon," *Journal of Instrumentation*, vol. 12, no. 01, P01017, Mar. 2017. DOI: 10.1088/1748-0221/12/01/P01017. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/12/01/P01017.

[98] P. Sorensen and K. Kamdin, "Two distinct components of the delayed single electron noise in liquid xenon emission detectors," *Journal of Instrumentation*, vol. 13, no. 02, P02032–P02032, Mar. 2018, ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/02/p02032. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/13/02/P02032.

[99] G. R. Araujo, T. Pollmann, and A. Ulrich, "Photoluminescence response of acrylic (pmma) and polytetrafluoroethylene (ptfe) to ultraviolet light: Limits on low-intensity photoluminescence in support materials of rare-event search experiments," *The European Physical Journal C*, vol. 79, no. 8, Aug. 2019, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7152-2. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-019-7152-2.

[100] K. Abe et al., "Radon removal from gaseous xenon with activated charcoal," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 661, no. 1, pp. 50–57, 2012, ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2011.09.051. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900211018468.

[101] A. Dobi et al., "Study of a zirconium getter for purification of xenon gas," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 620, no. 2, pp. 594–598, 2010, ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2010.03.151. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900210007771.

[102] J. Aalbers et al., *Flow-dependent tagging of ^{214}pb decays in the lz dark matter detector*, 2025. arXiv: 2508.19117 [physics.ins-det]. [Online]. Available: https://arxiv.org/abs/2508.19117.

[103] D. S. Akerib et al., "Projected sensitivity of the lux-zeplin experiment to the two-neutrino and neutrinoless double β decays of ^{134}Xe," *Phys. Rev. C*, vol. 104, p. 065 501, 6 Dec. 2021. DOI: 10.1103/PhysRevC.104.065501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevC.104.065501.

[104] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, and Y. Xu, "The Effective Field Theory of Dark Matter Direct Detection," *JCAP*, vol. 02, p. 004, 2013. DOI: 10.1088/1475-7516/2013/02/004. arXiv: 1203.3542 [hep-ph].

[105] J. Aalbers et al., "New constraints on ultraheavy dark matter from the lz experiment," *Phys. Rev. D*, vol. 109, p. 112 010, 11 Jun. 2024. DOI: 10.1103/PhysRevD.109.112010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.109.112010.

[106] V. Chepel and H. Araújo, "Liquid noble gas detectors for low energy particle physics," *Journal of Instrumentation*, vol. 8, no. 04, R04001, Apr. 2013. DOI: 10.1088/1748-0221/8/04/R04001. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/8/04/R04001.

[107] E. Aprile et al., "Excess electronic recoil events in xenon1t," *Physical Review D*, vol. 102, no. 7, Oct. 2020, ISSN: 2470-0029. DOI: 10.1103/physrevd.102.072004. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.102.072004.

[108] J. Aalbers et al., "The design, implementation, and performance of the lz calibration systems," *Journal of Instrumentation*, vol. 19, no. 08, P08027, Aug. 2024, ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/08/p08027. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/19/08/P08027.

[109] D. Akerib et al., "Simulations of events for the lux-zeplin (lz) dark matter experiment," *Astroparticle Physics*, vol. 125, p. 102 480, 2021, ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2020.102480. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927650520300529.

[110] M. Szydagis et al., *Noble element simulation technique v2.0*, version v2.0.0, Jul. 2018. DOI: 10.5281/zenodo.1314669. [Online]. Available: https://doi.org/10.5281/zenodo.1314669.

[111] R. James, "Signals, backgrounds and statistical inference for dark matter direct detection experiments," Embargoed, will be available at https://lz.lbl.gov/publications/, PhD thesis, University College London, 2023.

[112] D. Kodroff, "Background modeling and first searches for low energy signals in the lux-zeplin dark matter experiment," Ph.D. dissertation, The Pennsylvania State University, University Park, PA, Dec. 2023.

[113] M. S. et al, *A review of nest models, and their application to improvement of particle identification in liquid xenon experiments*, 2023. arXiv: 2211.10726 [hep-ex].

[114] A. Hitachi, T. Doke, and A. Mozumder, "Luminescence quenching in liquid argon under charged-particle impact: Relative scintillation yield at different linear energy transfers," *Phys. Rev. B*, vol. 46, pp. 11 463–11 470, 18 Nov. 1992. DOI: 10.1103/PhysRevB.46.11463. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.46.11463.

[115] C. E. Dahl, "The physics of background discrimination in liquid xenon, and first results from Xenon10 in the hunt for WIMP dark matter," Ph.D. dissertation, Princeton U., 2009.

[116] J. Thomas and D. A. Imel, "Recombination of electron-ion pairs in liquid argon and liquid xenon," *Phys. Rev. A*, vol. 36, pp. 614–616, 2 Jul. 1987. DOI: 10.1103/PhysRevA.36.614. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.36.614.

[117] M. Szydagis et al., "Nest: A comprehensive model for scintillation yield in liquid xenon," *Journal of Instrumentation*, vol. 6, no. 10, P10002, Oct. 2011. DOI: 10.1088/1748-0221/6/10/P10002. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/6/10/P10002.

[118] J. B. Birks, "Scintillations from organic crystals: Specific fluorescence and relative response to different radiations," *Proceedings of the Physical Society. Section A*, vol. 64, no. 10, p. 874, Oct. 1951. DOI: 10.1088/0370-1298/64/10/303. [Online]. Available: https://dx.doi.org/10.1088/0370-1298/64/10/303.

[119] T. Doke et al., "Let dependence of scintillation yields in liquid argon," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 269, no. 1, pp. 291–296, 1988, ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(88)90892-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0168900288908923.

[120] M. Szydagis et al., *A review of nest models for liquid xenon and exhaustive comparison to other approaches v3*, 2024. DOI: https://doi.org/10.3389/fdest.2024.1480975. arXiv: 2211.10726 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2211.10726.

[121] D. S. Akerib et al., "Discrimination of electronic recoils from nuclear recoils in two-phase xenon time projection chambers," *Phys. Rev. D*, vol. 102, p. 112 002, 11 Dec. 2020. DOI: 10.1103/PhysRevD.102.112002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.102.112002.

[122] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "Emcee: The mcmc hammer," *Publications of the Astronomical Society of the Pacific*, vol. 125, no. 925, p. 306, Mar. 2013. DOI: 10.1086/670067. [Online]. Available: https://dx.doi.org/10.1086/670067.

[123] J. Aalbers et al., "Background determination for the lux-zeplin dark matter experiment," *Physical Review D*, vol. 108, no. 1, Jul. 2023, ISSN: 2470-0029. DOI: 10.1103/physrevd.108.012010. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.108.012010.

[124] J. Aalbers et al., *Measurements and models of enhanced recombination following inner-shell vacancies in liquid xenon*, 2025. arXiv: 2503.05679 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2503.05679.

[125] D. J. Temples et al., "Measurement of charge and light yields for $^{1}27$Xe l-shell electron captures in liquid xenon," *Physical Review D*, vol. 104, no. 11, Dec. 2021, ISSN: 2470-0029. DOI: 10.1103/physrevd.104.112001. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.104.112001.

[126] J. Brodsky, C. Tunnell, mszydagis, jbalajth, V. Velan, and J. Huang, *Nestcollaboration/nest: Geant4 integration fixes and updates*, version v2.0.1beta, Mar. 2019. DOI: 10.5281/zenodo.2535713. [Online]. Available: https://doi.org/10.5281/zenodo.2535713.

[127] R. James, J. Palmer, A. Kaboth, C. Ghag, and J. Aalbers, "Flamenest: Explicit profile likelihoods with the noble element simulation technique," *Journal of Instrumentation*, vol. 17, no. 08, P08012, Aug. 2022. DOI: 10.1088/1748-0221/17/08/P08012. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/17/08/P08012.

[128] J. Aalbers and others (The LZ Collaboration), "Two-neutrino double electron capture of 124xe in the first lux-zeplin exposure," *Journal of Physics G: Nuclear and Particle Physics*, vol. 52, no. 1, p. 015 103, Dec. 2024. DOI: 10.1088/1361-6471/ad9039. [Online]. Available: https://doi.org/10.1088/1361-6471/ad9039.

[129] E. Aprile et al., "Double-weak decays of ^{124}Xe and ^{136}Xe in the xenon1t and xenonnt experiments," *Phys. Rev. C*, vol. 106, p. 024 328, 2 Aug. 2022. DOI: 10.1103/PhysRevC.106.024328. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevC.106.024328.

[130] B. Dutta and L. E. Strigari, "Neutrino physics with dark matter detectors," *Annual Review of Nuclear and Particle Science*, vol. 69, no. Volume 69, 2019, pp. 137–161, 2019, ISSN: 1545-4134. DOI: https://doi.org/10.1146/annurev-nucl-101918-023450. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-nucl-101918-023450.

[131] J. Aalbers et al., "A next-generation liquid xenon observatory for dark matter and neutrino physics," *Journal of Physics G: Nuclear and Particle Physics*, vol. 50, no. 1, p. 013 001, Dec. 2022. DOI: 10.1088/1361-6471/ac841a. [Online]. Available: https://dx.doi.org/10.1088/1361-6471/ac841a.

[132] A. W. Nelson et al., "Understanding the radioactive ingrowth and decay of naturally occurring radioactive materials in the environment: An analysis of produced fluids from the marcellus shale," *Environmental Health Perspectives*, vol. 123, no. 7, pp. 689–696, 2015. DOI: 10.1289/ehp.1408855. eprint: https://ehp.niehs.nih.gov/doi/pdf/10.1289/ehp.1408855. [Online]. Available: https://ehp.niehs.nih.gov/doi/abs/10.1289/ehp.1408855.

[133] B. Singh et al., "Nuclear data sheets for a=218," *Nuclear Data Sheets*, vol. 160, pp. 405–471, 2019, ISSN: 0090-3752. DOI: https://doi.org/10.1016/j.nds.2019.100524. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009037521930050X.

[134] S. J. Haselschwardt, J. Kostensalo, X. Mougeot, and J. Suhonen, "Improved calculations of β decay backgrounds to new physics in liquid xenon detectors," *Physical Review C*, vol. 102, no. 6, Dec. 2020, ISSN: 2469-9993. DOI: 10.1103/physrevc.102.065501. [Online]. Available: http://dx.doi.org/10.1103/PhysRevC.102.065501.

[135] John Armstrong and Ethan Ritchey, *SR3 Kr and Ar sampling measurements*, Internal LZ Document, Access on LZTWiki, Jul. 2024.

[136] A. K. A. Musalhi, "Searches for rare nuclear decays in the lux-zeplin experiment," A thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2023, Ph.D. dissertation, University of Oxford, St Catherine's College, Oxford, UK, Sep. 2023.

[137] E. Aprile et al., *Wimp dark matter search using a 3.1 tonne × year exposure of the xenonnt experiment*, 2025. arXiv: 2502.18005 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2502.18005.

[138] D. e. a. Baxter, "Recommended conventions for reporting results from direct dark matter searches," *The European Physical Journal C*, vol. 81, no. 10, Oct. 2021, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09655-y. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-021-09655-y.

[139] L. Lyons, *Discovering the significance of 5 sigma*, 2013. arXiv: 1310.1284 [physics.data-an]. [Online]. Available: https://arxiv.org/abs/1310.1284.

[140] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, no. 2, Mar. 2011, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[141] R. D. Cousins and L. Wasserman, *Phystat informal review: Marginalizing versus profiling of nuisance parameters*, 2024. arXiv: 2404.17180 [physics.data-an]. [Online]. Available: https://arxiv.org/abs/2404.17180.

[142] K. Cranmer, *Practical statistics for the lhc*, 2015. arXiv: 1503.07622 [physics.data-an]. [Online]. Available: https://arxiv.org/abs/1503.07622.

[143] S. Algeri, J. Aalbers, K. D. Morå, and J. Conrad, *Searching for new phenomena with profile likelihood ratio tests*, May 2020. DOI: 10.1038/s42254-020-0169-5.

[144] P. Virtanen et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2.

[145] J. Aalbers, B. Pelssers, V. C. Antochi, P. L. Tan, and J. Conrad, "Finding dark matter faster with explicit profile likelihoods," *Phys. Rev. D*, vol. 102, p. 072 010, 7 Oct. 2020. DOI: 10.1103/PhysRevD.102.072010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.102.072010.

[146] A. B. Owen, *Monte Carlo theory, methods and examples.* https://artowen.su.domains/mc/, 2013.

[147] Y. Tang, *A note on monte carlo integration in high dimensions*, 2023. arXiv: 2206.09036 [stat.ME]. [Online]. Available: https://arxiv.org/abs/2206.09036.

[148] M. A. et al, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/.

[149] National Energy Research Scientific Computing (NERSC), *Architecture- NERSC Documentation*. Accessed: Mar. 30, 2024. [Online]. Available: https://docs.nersc.gov/systems/perlmutter/architecture/.

[150] G. Cowan, *Statistical data analysis*. 1998, ISBN: 978-0-19-850156-5.

[151] D. S. Akerib et al., *Searches for light dark matter and evidence of coherent elastic neutrino-nucleus scattering of solar neutrinos with the lux-zeplin (lz) experiment*, 2025. arXiv: 2512.08065 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2512.08065.

[152] D. S. Akerib et al., *Search for new physics via low-energy electron recoils with a 4.2 tonne× year exposure from the lz experiment*, 2025. arXiv: 2511.17350 [hep-ex]. [Online]. Available: https://arxiv.org/abs/2511.17350.

[153] R. Hammann, T. Wolf, K. D. Morå, and J. Howlett, *Xenonnt/gofevaluation: V0.1.0-zenodo-init*, version v0.1.0-zenodo-init, Oct. 2021. DOI: 10.5281/zenodo.5626909. [Online]. Available: https://doi.org/10.5281/zenodo.5626909.

[154] S. Baker and R. D. Cousins, "Clarification of the use of chi-square and likelihood functions in fits to histograms," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 2, pp. 437–442, 1984, ISSN: 0167-5087. DOI: https://doi.org/10.1016/0167-5087(84)90016-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0167508784900164.

[155] P. Vischia, "Reporting results in high energy physics publications: A manifesto," *Reviews in Physics*, vol. 5, p. 100 046, 2020, ISSN: 2405-4283. DOI: https://doi.org/10.1016/j.revip.2020.100046. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405428320300095.

[156] Daniel Kodroff on behalf of the Sims and Backgrounds WGs, *SR3 Simulations and Backgrounds modeling Rev. B*, Internal LZ Document, Access on LZTWiki, Aug. 2024.

[157] Greg Rischbieter, Chami Amarasinghe, Erin Barillier, Yongheng Xu, Amardeep Chawla, Ellie Bishop, Mark Murdy, Asher Kaboth, *SR3 Calibration Analysis Review: Corrections and Modeling*, Internal LZ Document, Access on LZTWiki, Jun. 2024.

[158] E. Aprile et al., "Double-weak decays of ^{124}Xe and ^{136}Xe in the xenon1t and xenonnt experiments," *Phys. Rev. C*, vol. 106, p. 024 328, 2 Aug. 2022. DOI: 10.1103/PhysRevC.106.024328. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevC.106.024328.

[159] C. Oztireli et al., "Tensorflow graphics: Computer graphics meets deep learning," 2019.

[160] LZ Statistics Team (Robert James, Joe McLaughlin, Joshua Green, Leah Wolf, Makayla Trask, Jacopo Sinciar, Asher Kaboth), *Statistical Analysis of SR3 results (including SR1)*, Internal LZ Document, Access on LZTWiki, Sep. 2024.

[161] R. Ajaj et al., "Search for dark matter with a 231-day exposure of liquid argon using deap-3600 at snolab," *Physical Review D*, vol. 100, no. 2, Jul. 2019, ISSN: 2470-0029. DOI: 10.1103/physrevd.100.022004. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.100.022004.

[162] D. S. Akerib et al., "Results from a search for dark matter in the complete lux exposure," *Phys. Rev. Lett.*, vol. 118, p. 021 303, 2 Jan. 2017. DOI: 10.1103/PhysRevLett.118.021303. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.118.021303.

[163] E. Aprile et al., "Dark matter search results from a one ton-year exposure of xenon1t," *Phys. Rev. Lett.*, vol. 121, p. 111 302, 11 Sep. 2018. DOI: 10.1103/PhysRevLett.121.111302. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.111302.

[164] E. Aprile et al., "First dark matter search with nuclear recoils from the xenonnt experiment," *Phys. Rev. Lett.*, vol. 131, p. 041 003, 4 Jul. 2023. DOI: 10.1103/PhysRevLett.131.041003. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.131.041003.

[165] J. Aalbers et al., "First dark matter search results from the lux-zeplin (lz) experiment," *Phys. Rev. Lett.*, vol. 131, p. 041 002, 4 Jul. 2023. DOI: 10.1103/PhysRevLett.131.041002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.131.041002.

[166] Z. Bo et al., "Dark matter search results from 1.54 Tonne·Year exposure of pandax-4t," *Phys. Rev. Lett.*, vol. 134, p. 011 805, 1 Mar. 2025. DOI: 10.1103/PhysRevLett.134.011805. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.134.011805.

[167] Abantika Gosh, Aiham Al Musalhi, Amy Cottle, Bramwell Shafer, Daniel Hunt, Donghee Yeum, Drew Ames, Emily Perry, Gavin Cox, Harkirat Singh Riyat, Harvey Birch, Issy Darlington, Jack Bargeman, Jose Soria, Kaan Oyulmaz, Kate Lawes, Linda de Felice, Marcus Converse, Meghna K. Kannichankandy, Paulo Br´as, Ruben Coronel, Ryan Gibbons, Simran Dacve, Sparshita Dey, Yongheng Xu, and Yue Wang, *SR3 Cuts and Efficiencies*, Internal LZ Document, Access on LZTWiki, Jul. 2024.

[168] Paulo Brás, Daniel Hunt, Nicholas Fieldhouse, Marcus Converse, *TPC S2 Trigger and LZap Single Scatter Identification Efficiencies in SR3*, Internal LZ Document, Access on LZTWiki, Jul. 2024.

[169] Harvey Birch, Marcus Converse, Eryk Druszkiewicz, Miguel Hernandez, and Dev Ashish Khaitan, *Trigger Tuning for SR3*, Internal LZ Document, Access on LZTWiki, Nov. 2024.

[170] Emily Perry, Simran Dave, Jacob McLaughlin, Nicolas Angelides, Olivia Valentino, *Radon Backgrounds and Pb-Tagging in SR3*, Internal LZ Document, Access on LZTWiki, Jul. 2024.

[171] J. Allison et al., "Recent developments in geant4," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 835, pp. 186–225, 2016, ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.125. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900216306957.

[172] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991. DOI: 10.1109/18.61115.

[173] Sam Eriksen, Ewan Fraser, Harvey Birch, Alberto Us´on, Nicholas Fieldhouse, Jonathan Orpwood, Makayla Trask, *SR3 Neutron Rates and Efficiencies*, Internal LZ Document, Access on LZTWiki, Mar. 2025.

[174] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979, ISSN: 03036898, 14679469. Accessed: Jan. 13, 2026. [Online]. Available: http://www.jstor.org/stable/4615733.

[175] Conn, Andrew R. and Gould, Nicholas I. M. and Toint, Philippe L., *Trust Region Methods*. Society for Industrial and Applied Mathematics, 2000. DOI: 10.1137/1.9780898719857. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9780898719857. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9780898719857.