Background Mitigation in the LUX-ZEPLIN Experiment Using Statistical Analysis of Waveforms

Anna David

University College London

Department of Physics and Astronomy

Submitted to University College London (UCL) in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy.

Thesis submission date: **13th April 2024**

Declaration

I, Anna David, confirm that the work presented in this thesis is my own. In the following paragraphs, I will highlight my own contributions to each chapter, and clarify where information has been derived from other sources.

Rather than original work, chapter 2 presents a review of literature on the topic of dark matter, referring to many publications.

The LUX-ZEPLIN (LZ) experiment is described in chapter 3. I contributed to the operation of the detector through remote shifts, both to monitor the photomultiplier tubes and data acquisition, and to optimise the voltages of the electrodes following the first science run (SR1). On-site work included deployment of calibration sources and upgrades to monitoring sensors.

The Physics Readiness Monitor discussed in this chapter (including the tools described as existing features) was developed by E. Fraser. As a core member of the PREM team, I developed the two-dimensional comparison tool, Algorithm Groups, and LZap version comparisons, as well as other general improvements to the user interface. I developed and maintained the plots, algorithms and tuned algorithm limits from the TPC_Health module, shown in sections 3.7–3.9. Other plots shown were produced using the Skin Health module, maintained by L. Wolf. I also played a vital role in the SR2 data quality validation conducted by the PREM team.

The focus of chapter 4 is the analysis carried out by the LZ collaboration which produced the SR1 WIMP search result. My contributions to this chapter were the production of acceptance curves for the SR1 data quality cuts using an AmLi calibration source, as described in subsection 4.5.1. I contributed significantly to this investigation, which was carried out in collaboration with members of the UCL dark matter group as an independent verification of the SR1 WIMP signal efficiency deter-

mination. The further analyses described below did not directly impact this published result, but were developed using SR1 data, to produce a framework which could be used for future LZ science runs.

The tuning of simulations shown in chapter 5 is my own work, where code from T. Fruth was initially used as a basis for creating light collection efficiency histograms. The S1 area correction map used for comparisons between data and simulations was generated by G. Rischbieter and M. Murdy. I altered this code to produce separate maps for the top and bottom PMT arrays. The derivation of the g_1 value, using the Doke plot method described in section 5.2, was carried out by G. Blockinger.

The development of the likelihood-based data quality cuts presented in chapters 6 and 7 is my own work, with existing LZ Flamedisx tools used to generate projections of sensitivity and discovery potential.

Anna David
London, United Kingdom
13th April 2024

Abstract

The LUX-ZEPLIN (LZ) dark matter direct detection experiment searches for recoils of target xenon nuclei in a time projection chamber following an interaction with a Weakly Interacting Massive Particle (WIMP). LZ produced world-leading limits for spin-independent (SI) WIMP-nucleon interactions using data from its initial 60-liveday science run. During the expected 1000-liveday total runtime, LZ will explore new regions of electroweak parameter space, leading to a possible first discovery of WIMPs. As LZ acquires more data, accidental-coincidence backgrounds, resulting from random pairings of lone primary (S1) and secondary (S2) scintillation signals, become increasingly dominant.

This thesis presents a statistical approach to target these events, utilising the full time dependence provided by S1 pulse waveforms. The detection times of scintillation photons can be used to identify events with either a non-physical time separation between the S1 and S2, or a spurious pulse shape resulting from the pile-up of single photoelectrons misclassified as an S1. Data-driven templates of S1 waveforms are used to perform likelihood ratio tests targeting these background pathologies, as well as a background-agnostic S1 goodness of fit test. These form the basis of three test statistics used as input features in a boosted decision tree, which classifies events as signal or background by taking all three criteria into consideration. This machine learning approach improves the projected sensitivity of LZ to 40 GeV-mass WIMPs by 8%. The discovery potential for a 40 GeV WIMP with an SI WIMP-nucleon cross-section of $3.99 \times 10^{-48}\,\mathrm{cm}^2$ is improved by 11%.

The development of waveform-based cuts benefits from the use of accurate optical simulations, so this thesis additionally presents the optimisation of several material parameters, in order to better reproduce the light collection efficiency observed in data. Also described in this work is the development of data quality monitoring tools, which are essential to achieve a robust science result for an extremely rare signal.

Impact Statement

The STFC Particle Physics Advisory Panel highlights the importance of dark matter research in the mission of the UK particle physics community "to push the limits of human understanding of the fundamental description of nature at the smallest scales and highest energies" [1]. The work presented in this thesis contributes to the analysis of data from the LZ dark matter detection experiment. After achieving world-leading limits on WIMP dark matter with its first science result, LZ continues to explore new parameter space, with potential to make the first discovery of WIMPs. The data quality monitoring tools developed in this work are vital for ensuring both a robust result, from the removal of anomalous data, and efficient data taking, enabled by prompt feedback on detector conditions. It has also been shown that the data quality cuts developed in this work increase the potential for a future WIMP discovery by LZ, which would provide a breakthrough in our understanding of the nature of dark matter.

The XLZD consortium, formed by the LZ, XENON and DARWIN collaborations, will provide a next-generation xenon observatory to probe the remaining WIMP parameter space, as well as many other physics channels. Work developed in this thesis, such as the data-driven likelihood approach to waveform analysis and the BDT models, can provide a framework for future analyses in both LZ and the next-generation detector, saving both human effort and computation time. Skills with machine learning tools also have a wide range of applications within particle physics and beyond, for example in healthcare [2] and urban planning [3].

Participation in outreach events, such as Neutrino Day at the Sanford Underground Research Facility and Science Museum Lates in London, was facilitated by knowledge of the LZ experiment and the field of dark matter. This offered an opportunity to members of the public, who may not have a background in science, to engage with particle physics in a stimulating manner. Work included in this thesis was also presented to physics undergraduate students, to provide guidance regarding a future in academia. Outreach is vital for creating confidence in research and

inspiring future scientists who will continue the search for new physics.

Dedication

For my granny, Sheena, who graduated from UCL back in 1951; I am inspired by you and proud to follow in your footsteps.

Acknowledgements

This thesis would not exist without the kindness and generosity of many people, and I would like to express my gratitude towards them. First, thank you to Jim for bringing so much care and patience to your role as my supervisor. I was repeatedly blown away by your level of knowledge about every aspect of the LZ experiment and the incredibly detailed feedback you provided at every stage. Thank you also to Cham for always offering me your support, and Matteo for your encouragement as my second supervisor.

I could not have asked for a better research group than the UCLDM gang. Thank you for creating such a warm and supportive environment, where I have never felt afraid to share my thoughts. Special thanks to Rob for all your help, both as a stats pro and as a Number 1 karaoke partner. Fi, thanks for being there from our first days as PhD newbies. Thanks also to everyone in the UCL HEP group for the friendship, guidance and Friday pub trips.

Thank you to everybody in the LZ collaboration who has supported me with my work, in particular Ewan and Theresa for replying to my many, many Slack messages about PREM and simulations, respectively. Thank you to everyone who made sure that my time in and around South Dakota was anything but boring. Days spent paddleboarding at Lake Pactola, climbing in the Black Hills, and exploring Yellowstone National Park will always be very happy memories. Thank you also to Kevin, Aaron and everyone who welcomed me at Berkeley lab for six weeks.

Thanks to all my friends for sticking around and making sure I still had something resembling a social life. Sophie and Iris, thanks for making our gals' flat a lovely and somewhat calm place to be, even when I was in the depths of thesis writing. Thanks to Sam and Riana for bringing some routine to my life with our weekly bouldering sessions. Em, thank you for being there through every single high and low of the PhD (and for answering every question I was too embarrassed to ask anyone else). I am so fortunate to have shared this entire experience with you. Katie, Saraya, Will, Rachel, Izzy, Jaya and many others, thank you all for your

friendship over the years.

Thank you to all my family for the love and support. Thanks to the original Dr David, Uncle Tom, for the many years of proofreading and encouragement with physics. All my love to Mum, Dad and Christopher. Thank you for always being there; I could not have done it without you.

Contents

Table of Acronyms

Notation	Description
$2\nu\beta\beta$	Two-neutrino double beta
2νDEC	Two-neutrino Double Electron Capture
ACS	AccidentalChopStitch
AFT	Area Fraction Time
ALPACA	LZ analysis framework
BACCARAT	Geant4-based simulations package
BBN	Big Bang Nucleosynthesis
CMB	Cosmic Microwave Background
BDT	Boosted Decision Tree
DAQ	Data Acquisition
DB	Database
DD	Deuterium-deuterium
DER	Detector Electronics Response
DPE	Double Photoelectron
ER	Electron Recoil
e-train	Electron Train
Gd-LS	Gadolinium-loaded Liquid Scintillator
GOF	Goodness of Fit
GXe	Gaseous Xenon
ICV	Inner Cryostat Vessel
ISR	Initial State Radiation
ΛCDM	Lambda Cold Dark Matter
LCE	Light Collection Efficiency
LEE	Look Elsewhere Effect
LHC	Large Hadron Collider
LXe	Liquid Xenon
LZ	LUX-ZEPLIN
LZap	LZ Analysis Package
LZLAMA	LZ Light Analysis Monte Carlo Application
MACHO	Massive Compact Halo Object

Notation	Description
MCTruth	Monte Carlo Truth
MOND	Modified Newtonian Dynamics
MPE	Multiple Photoelectron
MS	Multiple Scatter
NEST	Noble Element Simulation Technique
NR	Nuclear Recoil
OCV	Outer Cryostat Vessel
OD	Outer Detector
PBH	Primordial Black Hole
phd	Photons Detected
ph-train	Photon Train
PMT	Photomultiplier Tube
PREM	Physics Readiness Monitor
PTFE	Polytetrafluoroethylene
QE	Quantum Efficiency
RFR	Reverse Field Region
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RQ	Reduced Quantity
S1c	Corrected primary scintillation signal
S2c	Corrected secondary scintillation signal
SD	Spin-dependent
SE	Single Electron
SHM	Standard Halo Model
SI	Spin-independent
SPE	Single Photoelectron
SR1/2/3	Science Run 1/2/3
SS	Single Scatter
SURF	Sanford Underground Research Facility
TBA	Top-bottom Asymmetry
TPC	Time Projection Chamber
UDT	Unphysical Drift Time
VUV	Vacuum Ultraviolet
WIMP	Weakly Interacting Massive Particle
XGBoost	Extreme Gradient Boosting

List of Figures

List of Tables

Chapter 1

Introduction

The presence of a dominant non-luminous and non-baryonic matter component in the Universe is well-motivated by a range of astrophysical and cosmological evidence, including galactic rotation curves and temperature fluctuations in the Cosmic Microwave Background. Several dark matter candidates have been postulated, on mass scales ranging from $\sim 10^{-21}$ eV to several hundred solar masses [4]. Weakly Interactive Massive Particles (WIMPs) on the GeV scale are a particularly compelling hypothesis, as they are able to predict the calculated dark matter relic density and arise in underlying theories such as supersymmetry. Chapter 2 reviews the evidence and candidates for dark matter, as well as the various WIMP direct detection techniques, with a focus on dual-phase time projection chambers (TPCs), which are used to search for primary (S1) and secondary (S2) scintillation signals following interactions with a target material. Xenon is an appealing choice, due to its large WIMP cross-section, lack of long-lived radioisotopes and excellent self-shielding properties.

This is the technology employed by the LUX-ZEPLIN (LZ) experiment, which consists of a TPC containing a 7-tonne liquid xenon active volume, surrounded by two veto systems and a water tank to provide shielding from backgrounds. Chapter 3 describes these detector components, as well as the dominant background signals, which must be well-understood and mitigated. Identification of data quality issues is vital for determining which acquisitions are suitable for use in analysis and for providing prompt feedback to detector operators. Chapter 3 shows how this was achieved through the use of the Physics Readiness Monitor, for which new tools and analysis modules were developed to assist with monitoring of the detector health. This framework was used in the validation of data from the first and second LZ science runs, and is in place for the third.

LZ has produced world-leading limits on spin-independent WIMP-nucleon interactions using data from the first science run, which had a total exposure of 60 livedays with a 5.5-tonne fiducial mass. This analysis is described in chapter 4, with a focus on the background resulting from the accidental coincidences of isolated S1 and S2 pulses. This includes the origins of these lone pulses, the data quality cuts that target them, and the modelling of their rate and distribution. The use of calibration sources to characterise the detector is also discussed, as well as the calculation of the signal acceptance of data quality cuts using AmLi neutron events from one such calibration campaign. Finally, chapter 4 describes the process of using a profile likelihood ratio to determine that the observed data were in agreement with the background-only hypothesis, and thus set exclusion limits.

Chapter 5 presents the optimisation of Geant4-based optical simulations. The optical parameters of LXe-PTFE reflectivity, LXe absorption length and LXe Rayleigh scattering length were tuned to reproduce the position-dependent photon collection efficiency observed in data, for both the top and bottom photomultiplier tube (PMT) arrays. Accurate optical simulations produce digitised waveforms and hence facilitate the understanding of complex pathologies and the development of pulse-based data quality cuts.

Chapter 6 exhibits work on a new waveform-based analysis developed for rejection of the problematic accidental-coincidence background. This was achieved by exploiting the position-dependent photon timing offset observed between the top and bottom PMT arrays. Data-driven templates of the S1 pulse were used to calculate the likelihood that a pulse is a real S1 within a true Single Scatter, resulting in three test statistics. The signal acceptance and background rejection efficiency of linear cuts placed on each of these test statistics is presented. Correlation between the three cuts indicates that a combined approach would be beneficial.

Chapter 7 presents a boosted decision tree (BDT) analysis, developed to optimise the rejection of accidental-coincidence backgrounds. The BDT model was trained with the three test statistics as input features. Chapter 7 describes the way in which the approach evolved to take account of interesting behaviour resulting from the definition of the likelihood ratios, which introduced an implicit dependence on the number of photons in the S1 pulse. The final model consists of 14 BDT models trained separately, for datasets binned by number of S1 photons. The improvement to the projected WIMP sensitivity and discovery potential of LZ, resulting from both this analysis and the linear cut approach, is

presented. The BDT models produce an 8% decrease in the most stringent limit from a 1000-liveday search, providing sensitivity to additional WIMP parameter space. An increase in discovery potential of 11% is also achieved for a 40 GeV WIMP with a spin-independent WIMP-nucleon cross-section of $3.99 \times 10^{-48}\,\mathrm{cm}^2$. A more efficient removal of accidental-coincidence events would allow for further improvements, so chapter 7 also describes future developments to the current BDT framework which would enable this.

Chapter 2

Dark Matter

Approximately 27% of the energy density of the Universe is made up of non-luminous matter [5], yet it remains undetected. As a result, the nature of dark matter is one of the most compelling questions in both cosmology and particle physics. This chapter will summarise the historical evidence for dark matter, as well as its properties and some of the most well-motivated candidates. Experimental approaches for dark matter detection will also be discussed, with a particular focus on the direct detection of Weakly Interacting Massive Particles.

2.1 Evidence

Evidence for dark matter exists on a wide range of scales, with astrophysical evidence consisting of gravitational effects on both galaxies and galaxy clusters. In the 1930s, Fritz Zwicky made observations of the redshift of galaxies in the Coma cluster to measure velocity dispersions. He applied the virial theorem, which relates the total kinetic energy to the gravitational potential energy of the cluster to estimate the total cluster mass. This revealed a high mass-to-light ratio, with a calculated mass 400 times larger than the "luminous mass" estimated from the light emitted by the cluster [6]. Zwicky attributed this to a non-visible component, which had a significantly higher mass than the observable matter, increasing the velocity dispersion.

Further evidence arises from gravitational lensing, the process by which, following special relativity, massive objects bend spacetime, resulting in distortion of background sources, such as galaxies. Effects can range from a temporary brightening (microlensing), to a mild distortion (weak lensing), to multiple separate images of the source (strong lensing). The extent of the distortion allows the mass of a foreground object to be deduced. Mass-to-light estimates can be produced without the re-

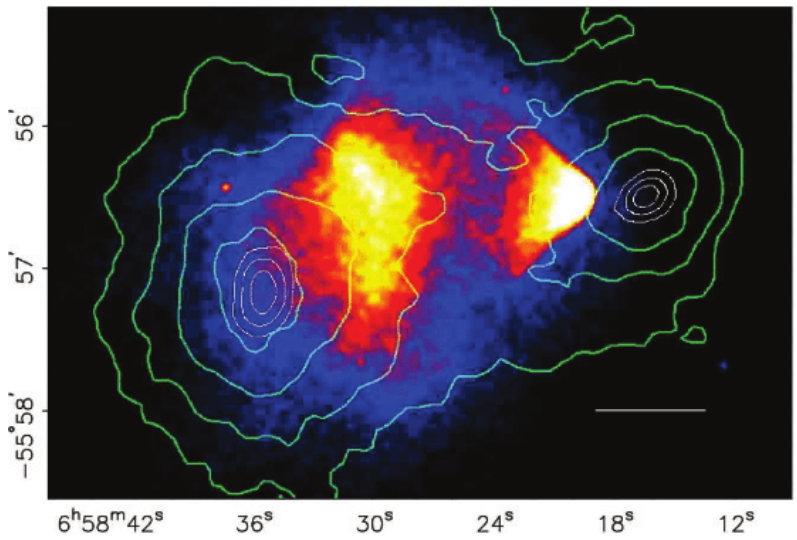


Figure 2.1: X-ray image from the Chandra Observatory of the merging cluster 1E0657-558, known as the Bullet Cluster, with green contours showing the inferred mass distribution from weak gravitational lensing. The white bar shows a 200 kpc distance for scale. Figure from Ref. [10].

quirement for estimating velocities [7], and dark matter structures can be mapped [8] [9].

One prominent example of the use of weak gravitational lensing is in measurements of the mass distribution of two colliding galaxy clusters, known collectively as the "Bullet Cluster", shown in Figure 2.1. The dominant baryonic component of the galaxies consists of X-ray-emitting plasma, which experiences ram pressure during the collision. The galaxies behave similarly to collisionless particles, moving past each other and away from the point of impact, leading to spatial separation from the plasma. The mass peaks of the system, provided by the gravitational lensing signal, are aligned near the galaxies, rather than the plasma, indicating that the dominant contribution to the mass of the system is non-baryonic [10].

Rotation curves of galaxies describe the orbital velocity, v, of stars as a function of radial distance, r. For a galaxy obeying Keplerian motion, this takes the form:

$$v(r) = \sqrt{\frac{GM(r)}{r}}, \tag{2.1.1}$$

where G is the gravitational constant and $M(r)$ is the distribution of mass as a function of radius. Assuming that mass is concentrated in a sphere of radius R with constant density, the velocity is expected to increase as $v(r) \propto r$ for $(r \leq R)$ and decay as $v(r) \propto r^{-1/2}$ for $(r > R)$, where the mass distribution is expected to remain roughly constant. In the 1970s, using an extremely sensitive spectrometer to measure the red-

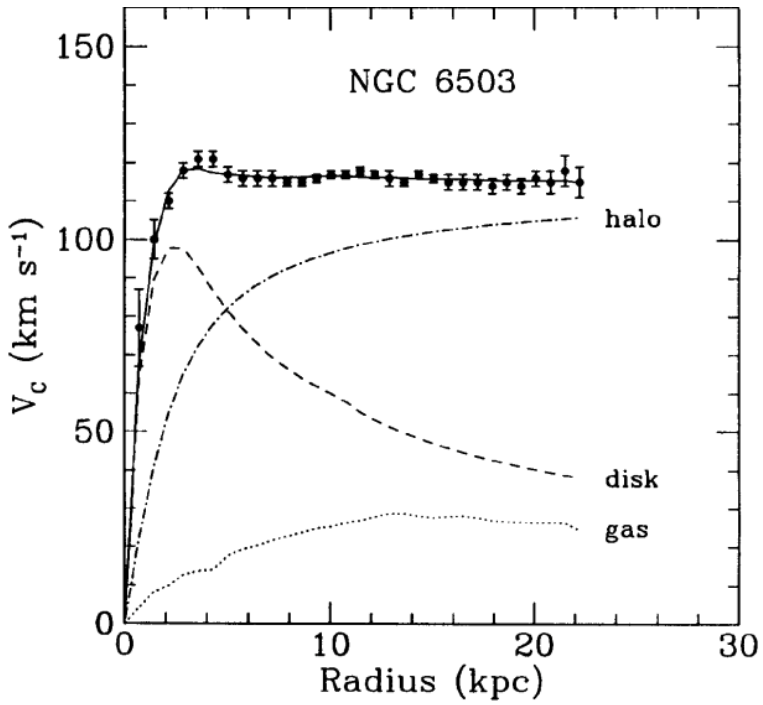


Figure 2.2: Galactic rotation curve for NGC 6503. Baryonic contributions from the disk and gas are shown, together with the dark matter halo contribution required to replicate the data. Figure from Ref. [15].

shift of stars in different parts of the galaxy, Vera Rubin was able to precisely determine the rotation curve of Andromeda, and observed that the orbital velocity approached a constant value with increasing radius [11]. It was shown that this could be accounted for by introducing a new mass component satisfying $M(r) \propto r$ [12] [13]. Similar behaviour was observed in later measurements of many galaxies [14], including the example shown in Figure 2.2, providing strong evidence for a non-luminous halo component.

Cosmological evidence for the existence of dark matter is provided by measurements of the Cosmic Microwave Background (CMB). In the early Universe, free nuclei and electrons existed in a plasma. Thomson scattering with free electrons meant that the mean free path of photons was short, and ensured that the photons were in equilibrium, so acted as a black body spectrum. Recombination occurred approximately 380,000 years after the Big Bang, and was the epoch at which the Universe had cooled sufficiently for electrons and nuclei to first form atoms. Photons then decoupled from the newly-formed atoms, and the Universe became transparent, with photons free streaming since then. As the Universe expanded, the photons continued to act as a black body, which today has a temperature of 2.73 K and has been redshifted into the microwave portion of the electromagnetic spectrum [16].

The first 25 years of surveying the CMB indicated that it was en-

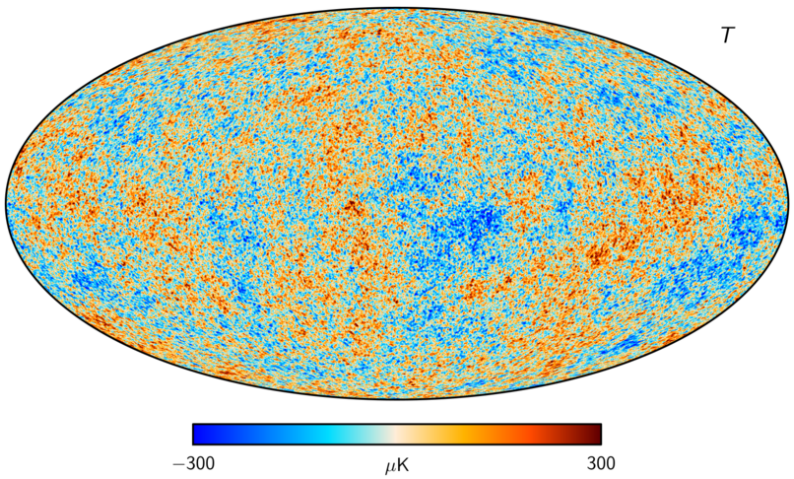


Figure 2.3: Map of fluctuations in the temperature of the CMB as measured by the Planck space observatory [19].

tirely smooth, however, in 1992, the COBE satellite mission discovered anisotropies. These took the form of temperature fluctuations on the order of 10^{-5}, following subtraction of the dipole anisotropy resulting from the Doppler shift of the radiation due to the motion of the Earth with respect to the CMB rest frame [17]. Prior to recombination, both dark matter and baryons were pulled towards the centre of density fluctuations, but, as the density of the photon-baryon fluid increased, pressure was created that opposed this gravitational force. Baryons therefore experienced acoustic oscillations resulting from competing gravitational and radiation pressures, whereas dark matter continued to fall inward. As the CMB photons were released, the pressure support was removed, resulting in the gravitational collapse of matter to create stars and galaxies. The resulting variations in temperature left an imprint in the CMB, providing evidence of these motions [18]. The anisotropies are mapped at various angular scales, which are quantified by the multipole moments of a spherical harmonic. A small multipole moment, l, corresponds to a large angular scale, and vice versa. Figs. 2.3 and 2.4 show the latest measurement of the temperature fluctuations and power spectrum from the Planck collaboration [5], which was able to observe the anisotropies on much smaller angular scales than COBE.

The location of the first peak corresponds to the curvature of the Universe, whereas the ratio of odd and even peaks informs the baryonic and dark matter densities. A higher baryonic density would allow the baryons to fall further into the gravitational potential wells, so that the odd peaks representing compressions would be amplified compared to the even peaks indicating rarefactions. A high dark matter density increases the size of the potential wells that the baryons can fall back into following

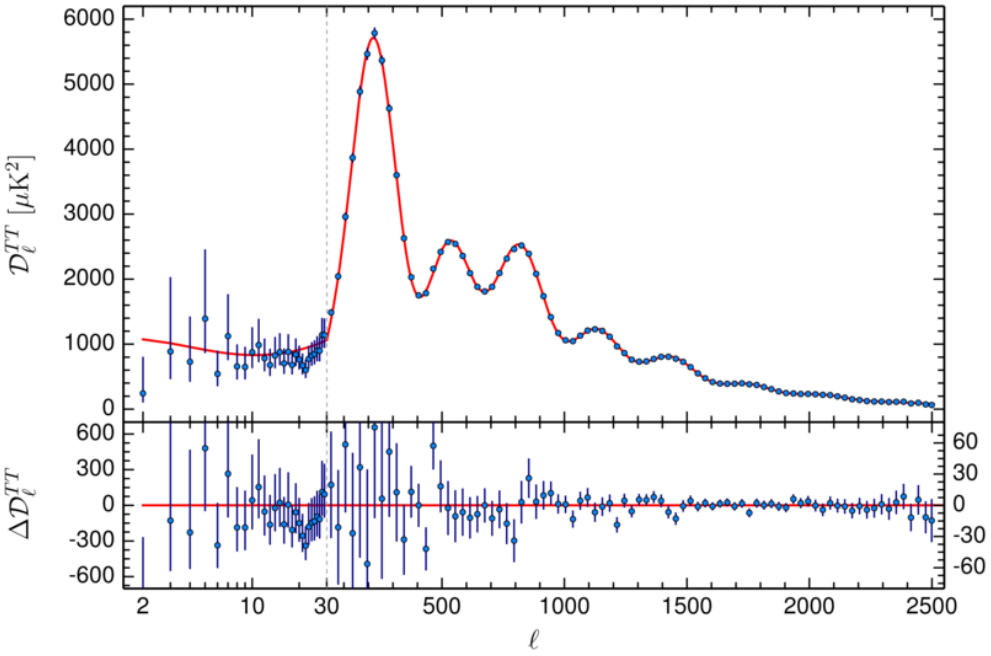


Figure 2.4: CMB power spectrum as measured by the Planck collaboration [5], where l is the multipole moment corresponding to angular scale. The fit assumes a ΛCDM model and the lower panel shows the residuals between this fit and the data.

a rarefaction. A higher proportion of dark matter thus enhances the third peak relative to the second [20]. The Planck collaboration measured the densities of baryonic matter and dark matter to be $\Omega_b h^2 = 0.0224 \pm 0.0001$ and $\Omega_c h^2 = 0.120 \pm 0.001$, respectively, corresponding to a non-luminous component comprising 84.3% of matter in the Universe. Here, the density parameter Ω_x for the relevant species is the ratio of the physical density, ρ_x, to the critical density, ρ_c, which provides a Universe of flat geometry:

$$\Omega_x = \frac{\rho_x}{\rho_c} = \frac{8\pi G \rho_x}{3H^2}, \tag{2.1.2}$$

where H is the Hubble parameter. Due to the large historical uncertainty on H, abundances are often quoted as $\Omega_x h^2$, where $h = H/(100$ kms^{-1}Mpc$^{-1})$.

2.2 Properties

The CMB power spectrum is well accounted for by the Lambda Cold Dark Matter (ΛCDM) model, which includes the existence of dark matter particles with several assumed properties [21]. For example, they must be electrically neutral, as they do not interact with the fundamental electromagnetic force. Evidence that DM particles were present before recombination indicates that they must also be long-lived.

Furthermore, dark matter particles appear to have been non-relativistic (cold) at the time of Universal "freeze-out", when their annihilation rate became smaller than the Hubble expansion rate. Dark matter was able to cluster in gravitational wells to form structures in the early Universe, when baryons were still coupled to photons. However, relativistic dark matter with a large free-streaming length would have been able to escape this gravitational pull, leading to inconsistencies with observations of large super clusters [22]. N-body simulations using cold dark matter are capable of reproducing these large-scale structures, but show some discrepancies with observations on sub-galactic scales, leading to the emergence of modifications to ΛCDM, such as warm dark matter [23].

Measurements of the CMB, combined with predictions from Big Bang nucleosynthesis (BBN), signify that baryonic matter comprises only a small portion of the total matter content of the Universe. BBN predicts the abundance of light elements produced within minutes of the Big Bang, when protons and neutrons were first able to form nuclei [16]. The rate of these reactions is dependent on the density of baryons in the early Universe, which is usually expressed normalised by the relic photon density. Measurements of the abundances of light elements placed constraints on this free parameter of baryon-to-photon ratio in the BBN model. A larger baryon-to-photon ratio leads to more reactions and thus a more efficient transformation of deuterium into ^{4}He. The abundance of deuterium is thus highly sensitive to the baryon-to-photon ratio. Additionally, deuterium is only created during BBN and is destroyed at a predictable rate throughout stellar evolution, so observations of abundance close to the primordial value are possible [24]. The first high-precision measurements of deuterium were used to calculate a baryon density of $\Omega_b h^2 = 0.020 \pm 0.002$ [25]. This baryon density alone would be unable to produce the gravitational effects discussed previously, implying that dark matter particles must be non-baryonic.

Observations of cluster collisions also place constraints on the self-interaction of dark matter particles, with simulations of the Bullet Cluster constraining the self-interaction cross-section per unit mass to $\sigma/m < 1.25\,\mathrm{cm}^2\mathrm{g}^{-1}$ [26].

In order to be considered a good candidate for dark matter, a new particle must also be compatible with current astrophysical and experimental constraints. Since no known particle is able to meet these requirements, detection of dark matter would be compelling evidence for new physics beyond the Standard Model.

2.3 Candidates

A variety of candidates have been proposed for dark matter, with an extremely broad mass range, spanning from $\sim 10^{-21}$ eV to several hundred solar masses [4]. Some of the most popular candidates will now be summarised, many of which are motivated by existing discrepancies in the Standard Model of particle physics.

2.3.1 Neutrinos

Standard Model neutrinos were an early candidate for dark matter, but constraints on their mass, from cosmological observations and measurements of the tritium β-decay spectrum in the KATRIN experiment, indicate that they can only make up a small portion of the total dark matter density in the Universe [27] [28]. Their relativistic nature also leads to inconsistencies with observations of structure, as discussed. A new species, the right-handed sterile neutrino, would interact only via gravity and through a weak coupling to left-handed Standard Model neutrinos. It is postulated that sterile neutrinos would be produced through oscillations of active neutrinos. Although significant constraints exist on mixing angle, for example from the non-observation of X-rays produced in the decay of the sterile neutrino to the active neutrino, sterile neutrinos can be considered primarily as a candidate for warm dark matter with a mass on the keV scale [29] [30].

2.3.2 Axions

The axion is a quasi-Nambu-Goldstone boson introduced to solve the "strong CP problem"; the absence of charge-parity violation in the strong interaction [31]. Most mass ranges have been ruled out by accelerator searches, astrophysical constraints from supernovae and red giants, and cosmological constraints from the axion relic density [32]. However, axions within the mass range of $1\,\mu\text{eV} \lesssim m \lesssim 10\,\text{meV}$ are candidates for dark matter. Thermal axions would move too fast to reproduce observed structures in galaxies and clusters, so axion dark matter must be produced non-thermally [33]. The ADMX collaboration uses haloscope technology to search for axions by converting them to detectable microwave photons using a strong magnetic field via the Inverse Primakoff Effect [34]. Experiments also search for axions via their couplings with electrons, in a process analogous to the photoelectric effect [35] [36].

2.3.3 Massive Compact Halo Objects

Massive compact halo objects (MACHOs) are non-luminous astrophysical bodies consisting of baryonic matter, that could make up some fraction of the dark matter in the Universe. Example MACHOs include brown dwarfs, large planets, and neutron stars [37]. The MACHO project monitored the brightness of stars in the Large Magellanic Cloud to observe 13-17 gravitational microlensing events over 5.7 years, but ruled out a 100% MACHO halo at the 95% confidence level. The MACHO halo fraction was instead placed at 20%, with typical MACHO masses between 0.15 $M_\odot$ and 0.9 $M_\odot$ [38].

Primordial black holes (PBHs) are a subcategory of MACHOs that would have formed from the gravitational collapse of density fluctuations in the very early Universe [39]. The LIGO interferometer detected gravitational waves from a merger of two black holes of mass $\sim$30 $M_\odot$ [40], and there was discussion of whether they could be primordial in nature and constitute dark matter [41]. Observational constraints have been placed on the abundance of PBHs at various mass ranges, for example using velocity dispersions of star clusters or microlensing [39]. Gravitational waves observed by LIGO and Virgo place the percentage of dark matter consisting of PBHs at $\lesssim 1\%$ for masses between 1 $M_\odot$ and 300 $M_\odot$ [42] [43].

2.3.4 Modified Gravity

Shortcomings of the dark matter model include an inability to reproduce behaviour defined by Renzo's rule (which states that every feature in the luminosity profile of a galaxy has a corresponding feature in its rotation curve) and the Tully-Fisher relation between the baryonic mass and rotation speed of a galaxy [44]. An alternative hypothesis to dark matter exists in the form of modifications to Newtonian gravity, a form of which was first presented in 1983 [45]. Known as modified Newtonian dynamics (MOND), the proposal originated as an explanation of the motion of galaxies, in which the gravitational force is altered at very low accelerations. Despite success with gravitational effects, the MOND theory struggles to reproduce all observations that motivate dark matter, including the CMB power spectrum [44]. Analysis of observations of wide binary stars by the Gaia satellite were used to claim a deviation from Newtonian expectations, and report a gravitational anomaly in accordance with a MOND model at weak accelerations of $g_N \lesssim 10^{-9}\,\mathrm{ms}^{-2}$ [46], with similar conclusions drawn in Ref. [47]. However, a subsequent

Bayesian analysis, using much stricter data quality cuts, found that the same data were consistent with Newtonian gravity, and excluded MOND at 16σ confidence. Although these results appear to falsify MOND, it is made clear that they cannot be used to argue that the ΛCDM model is necessarily correct [48].

2.3.5 Weakly Interacting Massive Particles

It may be assumed that dark matter particles were produced thermally in the early Universe, and were in thermal equilibrium with baryonic matter. As the Universe expanded, the density of dark matter particles per comoving volume reduced with decreasing temperature, until the moment of freeze-out, when the density of dark matter particles became constant. This provides a dark matter relic density of

$$\Omega h^2 \approx \frac{3 \times 10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle \sigma_{\text{ann}} v \rangle}, \tag{2.3.1}$$

where $\langle \sigma_{\text{ann}} v \rangle$ is the thermally-averaged dark matter annihilation cross-section [49]. The calculated dark matter relic density can be predicted by a particle with mass on the GeV scale that couples to the Standard Model via the weak interaction, with an annihilation cross-section of $\langle \sigma_{\text{ann}} v \rangle \sim G_F^2 m_\chi^2$ [50].

Dark matter particles with these properties are known as Weakly Interacting Massive Particles (WIMPs), and their ability to reproduce the expected relic density is known as the "WIMP miracle". One widely-discussed candidate for the identity of the WIMP is the lightest of the four neutralinos in the Minimal Supersymmetric Standard Model (MSSM): electrically neutral fermionic states which are composites of the superpartners of the neutral gauge bosons and the neutral Higgs bosons. This particle would be stable and could only be destroyed as a result of pair annihilation, making it a good candidate for dark matter [51]. WIMPs have been a primary focus of dark matter detection experiments in recent years.

2.4 Detection

As shown by Figure 2.5, three complementary methods are employed in dark matter searches, exploiting interactions in the form of dark matter annihilation, production and scattering off Standard Model particles. Each mode will be discussed, with a particular focus on direct detection of WIMP-nucleus scattering.

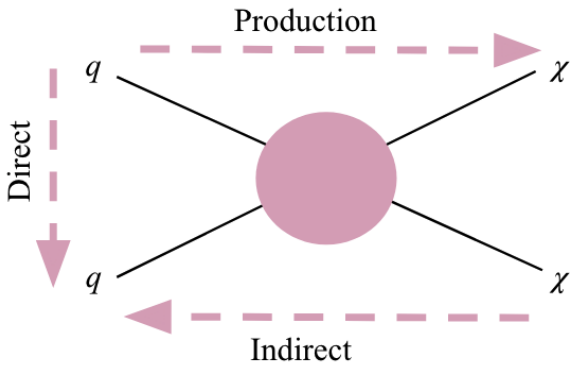


Figure 2.5: Cartoon Feynman diagram illustrating the three modes of dark matter detection, where χ and q represent dark matter and Standard Model particles, respectively.

2.4.1 Collider Searches

Dark matter produced through collisions of Standard Model particles cannot be observed directly, but its presence could be inferred by missing transverse momentum [52]. The Large Hadron Collider (LHC) is the highest-energy particle collider currently operating, and a goal of the multipurpose ATLAS and CMS experiments is to search for physics beyond the Standard Model (BSM), including dark matter. The minimal experimental signature of dark matter production is a single visible Standard Model "X" particle combined with missing transverse momentum, known as a mono-X process [53].

Many models exist for the interactions between Standard Model and dark matter particles, so collider searches often make minimal assumptions about the detectable products, so as to be model agnostic. This can be achieved by requiring that the recoiling visible particles are governed by Standard Model processes [54]. This includes initial state radiation (ISR) in the form of jets [55], photons and vector bosons [56] [57]. If the mass of the mediator is large compared to the energy of the collision, an effective field theory (EFT) can be used, and the unspecified coupling between the Standard Model and dark matter particles can be modelled as a contact interaction described by a set of operators, as shown in the Feynman diagram in Figure 2.6a [54].

As couplings of the Higgs boson to light quarks and gluons are suppressed, the rate of Higgs ISR is not significant. Mono-Higgs searches instead focus on interactions in which the Higgs couples directly to a BSM particle which participates in the production of dark matter. An example process is shown in Figure 2.6b, where a heavy neutral Z' mediator decays to a Higgs boson, h, and a pseudoscalar, A^0, of a two-Higgs-doublet model, where A^0 then decays to a pair of dark matter particles [58]. This

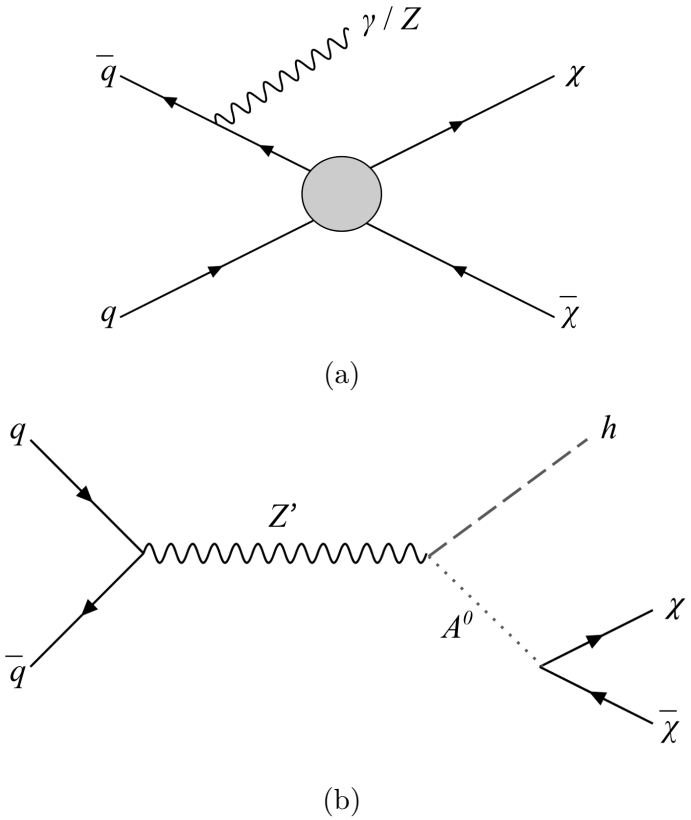


Figure 2.6: Feynman diagrams of mono-X signatures in collider searches for dark matter. (a) shows a process with photon or Z boson initial state radiation, where the interaction between Standard Model quarks and antiquarks (q and $\bar{q}$), and dark matter particles (χ and $\bar{\chi}$) is modelled using an effective field theory. (b) shows an example of a simplified model, in which a mediator Z' decays to a Higgs boson, h, and a pseudoscalar, A^0, with the latter then decaying to a pair of dark matter particles.

is an example of a "simplified model", in which the collision energy is comparable to or higher than the mediator mass, and so interactions are resolved into single-particle exchanges [53]. These models include additional degrees of freedom associated with the mediator particle, and bridge the gap to "complete models" with many parameters, such as MSSM. As no evidence of dark matter has been found, LHC searches have been able to set constraints for various models.

2.4.2 Indirect Detection

Indirect dark matter searches aim to detect the products of annihilations or decays of dark matter particles. Dark matter could annihilate directly to γ rays, and models with hadronic final states would also result in a γ signal. γ rays have the benefit of travelling directly from the source with very little absorption [59]. The Fermi Gamma-Ray Telescope provides the leading sensitivity to γ rays in the GeV region, and has reported

an anomalous flux in this range, which is well-explained by WIMP dark matter, but also by pulsars [60]. Meanwhile, the HESS air Cherenkov telescope sets the tightest constraints on TeV dark matter annihilation in the galactic centre [61].

Dark matter annihilation can also result in antimatter, which can be studied with cosmic-ray observations. Cosmic rays allow probing of very high energies, but their propagation is not well understood. Experiments including PAMELA have reported an excess in 10 - 1000 GeV positrons [62], and AMS-02 observed an excess of 5 - 10 GeV antiprotons [63], which hint at dark matter candidates with masses on the $\sim$ TeV and $\sim$ 100 GeV scales, respectively [64] [65].

Searches for secondary particles also include high-energy neutrinos resulting from the annihilation of WIMPs in the Sun, as well as from the galactic centre. Due to the relative difficulty of detecting neutrinos, constraints from these searches are generally weaker than those from γ signals and cosmic rays [66]. The IceCube underground neutrino observatory detected a TeV - PeV neutrino flux [67], which could include a component from dark matter decays [68], however the non-observation by the Fermi telescope of the γ rays that would accompany this decay provides constraints [69]. Unlike collider searches, indirect detection is able to set limits on the lifetime of dark matter particles [70].

2.4.3 Direct Detection

The dark matter halo of the Milky Way would produce a WIMP flux on Earth of the order $10^5(100\,\mathrm{GeV}/m_\chi)\mathrm{cm}^{-2}\mathrm{s}^{-1}$, where m_χ is the WIMP mass. This flux is sufficiently large that a measurable fraction of WIMPs are expected to scatter off target nuclei in direct detection experiments. The event rate of WIMPs is given by

$$\frac{dN}{dE} = \epsilon \frac{\rho}{m_\chi m_N} \int_{v_{\min}} v f(\mathbf{v}) \frac{d\sigma}{dE} dv. \tag{2.4.1}$$

$\epsilon = tM_T$ is the experimental exposure, where t is the observation time, and M_T is the detector mass [71]. This dependence on mass and time motivates the preference for large-scale detectors and long livetimes. ρ is the local dark matter density and m_N is the mass of the target nucleus. $v_{\min}$ is the minimum speed required to produce a WIMP recoil of energy E.

The velocity distribution of WIMPs in the Milky Way, $f(\mathbf{v})$, is the main astrophysical input to the rate equation. Direct detection experiments typically assume a Standard Halo Model (SHM), where WIMPs

follow an isotropic Maxwell-Boltzmann velocity distribution of the form

$$f(\mathbf{v}) = \frac{1}{(2\pi)^{3/2}\sigma^3} \exp\left(-\left(\mathbf{v} - \mathbf{v}_{\text{lag}}\right)^2 / 2\sigma^2\right) \tag{2.4.2}$$

where $\mathbf{v}_{\text{lag}}$ is the velocity of the Earth with respect to the Galactocentric rest frame and has an average value of $|\mathbf{v}_{\text{lag}}| \approx 230$ km s^{-1}, and σ is the velocity dispersion, with a typical value of 156 km s^{-1}.

The total WIMP-nucleus cross-section, $\frac{d\sigma}{dE}$, depends on the interaction model used. The total cross-section is often considered as a combination of spin-independent (SI) and spin-dependent (SD) components, corresponding to scalar or vector couplings to quarks, and axial-vector couplings to quarks, respectively [71]:

$$\frac{d\sigma}{dE} = \frac{m_N}{2\mu_{\chi N}^2 v^2}\left[\sigma_0^{SI} F_{SI}^2(E) + \sigma_0^{SD} F_{SD}^2(E)\right]. \tag{2.4.3}$$

Here, $\mu_{\chi N}$ is the WIMP-nucleus reduced mass, $\sigma_0^{SI,SD}$ are the SI and SD WIMP-nucleus cross-sections at zero momentum transfer and $F_{SI,SD}$ are form factors describing the nuclear charge density. The SI model assumes identical interactions with neutrons and protons, and considers the mass and charge distribution of the entire nucleus. For SI scattering, the Helm form factor is commonly used, and is a convolution of the charge density of a spherical nucleus, which has a uniform distribution inside a certain radius and is zero outside, with an exponential factor to allow for the soft edge of the nucleus. $\sigma_0^{SI,N}$ is defined as:

$$\sigma_0^{SI,N} = \frac{4\mu_{\chi N}^2}{\pi}\left[Z f_p + (A - Z) f_n\right]^2, \tag{2.4.4}$$

where A is the atomic mass, Z is the proton number and f_p and f_n are the effective WIMP couplings to protons and neutrons, respectively [72]. Assuming that $f_p = f_n$,

$$\sigma_0^{SI,N} = \left(\frac{\mu_{\chi N}}{\mu_{\chi n}}\right)^2 A^2 \sigma^{SI}, \tag{2.4.5}$$

where $\mu_{\chi n}$ is the WIMP-nucleon reduced mass and σ^{SI} is the WIMP-nucleon SI cross-section. The dependence on A^2 means that target materials with a high mass number have an increased sensitivity to SI interactions. Figure 2.7 shows the event rate as a function of nuclear recoil energy for several target materials, for a WIMP mass of 40 GeV, as well as the dependence of rate on WIMP mass.

The event rate for SD interactions is given by

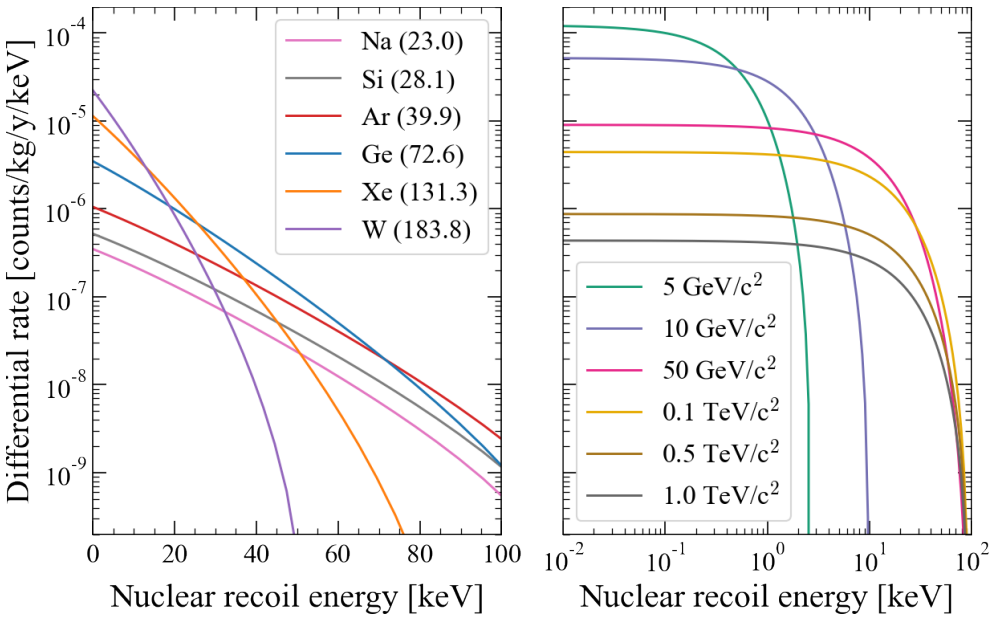


Figure 2.7: Reproduction of Figure 1.5 from Ref. [73], produced using the wimprates package [74]. Shown is the differential WIMP rate, assuming SI interactions, as a function of nuclear recoil energy. Left is a 40 GeV WIMP interacting with various target materials, and right is a selection of WIMP masses scattering in xenon. A value of $\sigma^{SI} = 1 \times 10^{-46}\,\mathrm{cm}^2$ has been chosen.

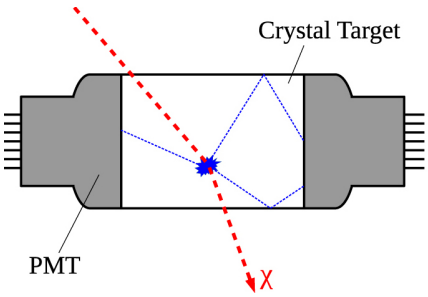
$$\sigma_0^{\mathrm{SD}} = \frac{32 G_F^2 \mu_{\chi N}^2}{\pi} \frac{(J+1)}{J} \left[a_p \langle S_p \rangle + a_n \langle S_n \rangle\right]^2 . \tag{2.4.6}$$

where J is the total nuclear spin of the target and $\langle S_p \rangle$ and $\langle S_n \rangle$ are expectation values of the proton and neutron spin operators from calculations using nuclear models. a_p and a_n are the couplings to protons and neutrons, one of which is usually set to zero when quoting SD results, assuming WIMPs couple to protons or neutrons only. Neutron-odd targets are most sensitive to WIMP-neutron interactions, while proton-odd targets are most sensitive to WIMP-proton interactions. For heavy target nuclei such as xenon, the SI component is dominant.

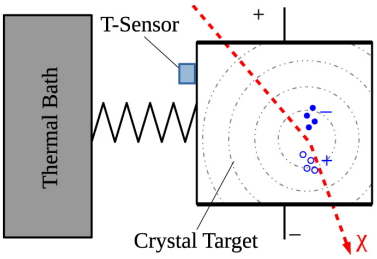
Use of an effective field theory can provide a more general approach, by parameterising WIMP-nucleon interactions in terms of fourteen four-field operators. The inclusion of momentum dependence in these operators allows for models beyond the typical SI and SD interactions [75].

2.5 Direct Detection Techniques

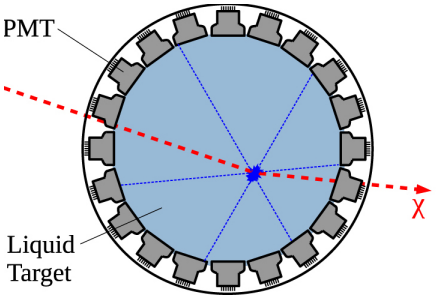
Various direct detection techniques aim to measure the energy of the WIMP-nucleus interaction by observing light, charge, heat, or some combination of the three. Six of the most common detector designs, informed by Ref. [76] and shown in Figure 2.8, are summarised, along with exam-



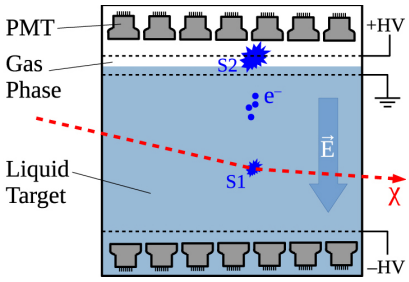
(a) Anorganic crystal detector



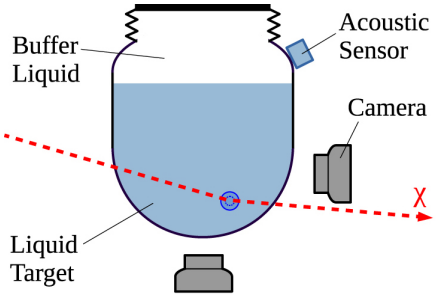
(b) Cryogenic solid-state detector



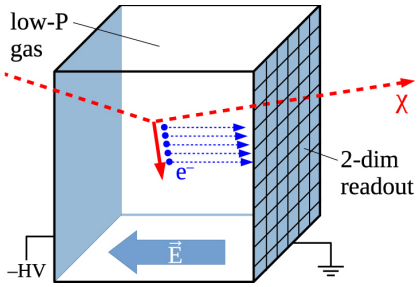
(c) Single-phase noble liquid detector



(d) Dual-phase time projection chamber



(e) Bubble chamber



(f) Directional detector

Figure 2.8: Diagrams of the six direct detection techniques described in the text. The scatter of a WIMP, χ, is shown in red and the propagation of recoil energy is shown in dark blue. All images from Ref. [76].

ples of each type of experiment.

- **Anorganic crystal detectors:** The first WIMP direct detection experiment used a high-purity germanium crystal diode to measure charge in the form of electron-hole pairs [77]. Current experiments use Ge (CDEX-10 [78]) and Si (DAMIC [79], SENSEI [80]) semiconductors to achieve very low thresholds, due to the small amount of energy required to create electron-hole pairs, but high levels of electronic noise limit the mass of these detectors to the kg-scale. Alternatively, light signals from arrays of high-purity kg-scale crystals can be observed by photomultiplier tubes (PMTs), as shown in Figure 2.8a. Intrinsic backgrounds are high, and background discrimination techniques such as fiducialisation, where a clean inner part of the active volume is defined, are not possible. The WIMP

wind, and hence the rate of recoils, varies throughout the year due to the orbital motion of the Earth. Background signals are not expected to experience this periodic behaviour [81], so experiments such as DAMA/LIBRA [82], COSINE-100 [83] and ANAIS-112 [84], using thallium-doped NaI, and KIMS [85], using thallium-doped CsI, search for an annual modulation signal. Section 2.7 presents a discussion of the modulation effect reported by DAMA/LIBRA, which is in tension with other experimental results.

- **Cryogenic solid-state detectors:** Semiconductor detectors are maintained at millikelvin temperatures by a cryogenics system to reduce thermal noise. The temperature increase resulting from heat signals in the form of phonons is then measured with superconducting transition edge sensors. Simultaneously measuring the ionisation signal using an applied voltage allows for background discrimination. All modern cryogenic detectors therefore exploit this two-channel detection approach, with the SuperCDMS [86] and EDELWEISS [87] collaborations also measuring the ionisation signal using Ge, and CRESST [88] measuring the scintillation signal using $CaWO_4$. Cryogenic detectors provide excellent energy resolution, but single detectors are limited to the kg-scale, as a small heat capacity is required. This can be overcome by using arrays of crystals, but the larger surface-to-volume ratio increases the impact of surface contaminants [76]. Searches that operate with low energy thresholds offer improved sensitivity to low-mass dark matter particles. For example, in the CDMSlite mode, large amplification of the phonon signal by a high bias voltage across the detector allows for a much lower energy threshold than in the normal SuperCDMS operating mode [89]. However, without the ionisation signals recorded in normal operation, there is less background discrimination.

- **Single-phase noble liquid detectors:** The noble elements xenon and argon are excellent scintillators, and do not suffer from the high intrinsic backgrounds associated with long-lived isotopes in krypton. They can both be liquified to form a dense target for WIMPs. Following an interaction, single-phase detectors measure the scintillation signal only, so noble-liquid targets are often spherical and surrounded by PMTs to ensure high light collection. The position of an interaction is determined by the photon timing and the PMT hit pattern. Argon detectors, such as DEAP-3600 [90], make use of excellent pulse-shape discrimination allowing for powerful suppression of backgrounds, necessary due to high levels of ^{39}Ar.

Conversely, xenon detectors, including XMASS [91], must rely on fiducialisation.

- **Dual-phase time projection chambers:** Time projection chambers (TPCs) use electric fields to drift electrons through a sensitive volume in order to perform three-dimensional position reconstruction. Dual-phase dark matter detectors also observe the primary scintillation signal in liquid, utilising PMTs in arrays at the top and bottom of the (typically cylindrical) detector. The ionisation signal is collected by drifting electrons up to the liquid surface using a vertical electric field, where they are emitted into an additional gas layer by a stronger extraction field. Here, a secondary scintillation signal is produced from collisions with gas atoms. The measurement of both signals allows for mm-precision position reconstruction. The precisely-measured time separation between the two signals, together with knowledge of the electron drift velocity, provides the height of the interaction. The XY position is determined from the location of secondary scintillation light in the top PMT array. The relative excitation and ionisation yields depend on the type of interaction, so the ratio between primary and secondary signals allows for background discrimination. The Darkside-50 [92] collaboration uses argon as the target material, whereas LUX-ZEPLIN [93], XENONnT [94] and PandaX-4T [95] use xenon TPCs, which will be discussed in further detail in section 2.6.

- **Bubble chambers:** Bubble chambers contain superheated liquids held just below their boiling point. Energy depositions will cause a local phase transition, resulting in the creation of a bubble. The chamber is observed by cameras, allowing for mm-precision position determination, although energy reconstruction is not possible, and significant deadtime is introduced by the requirement for compression and decompression following each event. The isotope ^{19}F has the highest sensitivity to SD WIMP-proton couplings, so is typically contained in the target fluid, while iodine provides good sensitivity to SI interactions, due to its high mass number. The PICO collaboration uses a C_3F_8 target material and currently reports the highest sensitivity to WIMP-proton scattering [96].

- **Directional detectors:** Directional detectors search for an excess of nuclear recoils from the direction of the WIMP wind. This often exploits the daily modulation resulting from the rotation of the Earth with respect to the WIMP flux. Low-pressure gas targets

such as C_3F_4 are the most commonly-used materials due to the long ionisation tracks produced by nuclear recoils, allowing for improved direction reconstruction, and sensitivity to SD interactions is often a focus. Tracks are either observed by cameras, or charge is drifted under an electric field and collected on multi-wire readout planes. Good discrimination of backgrounds is achieved using track range and ionisation density. The DRIFT-II TPC uses a CS_2 + C_3F_4 + O_2 target material [97].

2.6 Dual-phase Xenon TPCs

The principle behind dual-phase TPCs was summarised in section 2.5. The use of xenon as a target material in this type of detector will now be discussed, and a more detailed description of the microphysics processes involved in the generation of the primary and secondary scintillation signals will be given.

2.6.1 Interactions in Xenon

There are several advantages to using xenon as the target material. First, liquid xenon (LXe) allows for excellent self-shielding due to its high density, meaning that a low-background inner fiducial volume can be defined [98]. The high density also allows for a high target mass within a relatively small cryostat. The high mass number of xenon ($A = 131.3$) increases sensitivity to SI interactions, which depend on A^2, and the natural abundance of neutron-odd ^{129}Xe and ^{131}Xe provides high sensitivity to SD interactions. LXe experiments are also sensitive to many dark matter models beyond SI and SD scattering using an EFT approach, due to the large atomic mass and neutron-rich composition [99]. As well as this, xenon has an absence of long-lived isotopes, besides ^{136}Xe and ^{124}Xe, which have extremely long half-lives [100].

As discussed in section 2.5, interactions in target materials result in energy deposits in the form of excitation, ionisation and heat, allowing for three channels of detection. The energy, E_0, transferred to a medium can thus be written as

$$E_0 = N_i E_i + N_{ex} E_{ex} + \text{heat}, \tag{2.6.1}$$

where E_{ex} and E_i are the mean energies required to excite or ionise an electron and N_{ex} and N_i are the number of excited or ionised atoms. The de-excitation of atoms creates primary scintillation, known as the

S1 signal, while the ionisation electrons result in the secondary S2 signal.

The signature of a WIMP interaction, as well as that of neutron backgrounds, is a nuclear recoil (NR), whereas γ and β backgrounds result in electron recoils (ERs). The energy transferred to heat is much more significant for NRs than ERs, and is not detected in the TPC, but instead dissipates in the LXe [101]. The ratio N_{ex}/N_i was also found to differ depending on the interaction type, with a value in the range [0.06 - 0.2] for ERs in LXe [102], compared to ~ 1 for NRs [103]. This allows for discrimination between NRs and ERs, which permits background rejection.

2.6.2 Primary Scintillation: S1 Signal

The primary scintillation signal can be produced through two distinct mechanisms [101], the first of which is the excitation of a xenon atom, which, with a neighbouring atom, forms an excimer; a strongly-bound diatomic molecule in the excited state. The de-excitation of the excimer to the ground state leads to the emission of a photon in the vacuum ultraviolet (VUV) range, with a wavelength peaking at 178 nm, denoted by $h\nu$.

$$\begin{array}{ll} e^- + \mathrm{Xe} \rightarrow \mathrm{Xe}^* + e^- & \text{impact excitation} \\ \mathrm{Xe}^* + \mathrm{Xe} \rightarrow \mathrm{Xe}_2^{*,v} & \text{excimer formation} \\ \mathrm{Xe}_2^{*,v} + \mathrm{Xe} \rightarrow \mathrm{Xe}_2^* + \mathrm{Xe} & \text{relaxation} \\ \mathrm{Xe}_2^* \rightarrow \mathrm{Xe} + \mathrm{Xe} + h\nu & \text{VUV emission} \end{array} \tag{2.6.2}$$

The superscript v indicates excited states with vibrational excitation, as opposed to purely electronic excitation.

Scintillation also arises from the recombination of diatomic positive xenon ions with ionisation electrons. The diatomic ions are formed through interactions of xenon ions with neighbouring atoms, before recombination results in an excited atom. An excimer is then produced with a neighbouring atom, which leads to the emission of a VUV photon through a similar mechanism to the process described in Equation 2.6.2.

$$\begin{array}{ll} e^- + \mathrm{Xe} \rightarrow \mathrm{Xe}^+ + 2e^- & \text{ionisation} \\ \mathrm{Xe}^+ + \mathrm{Xe} + \mathrm{Xe} \rightarrow \mathrm{Xe}_2^+ + \mathrm{Xe} & \\ e^- + \mathrm{Xe}_2^+ \rightarrow \mathrm{Xe}^{**} + \mathrm{Xe} & \text{recombination} \\ \mathrm{Xe}^{**} + \mathrm{Xe} \rightarrow \mathrm{Xe}^* + \mathrm{Xe} + \text{ heat} & \\ \mathrm{Xe}^* + \mathrm{Xe} + \mathrm{Xe} \rightarrow \mathrm{Xe}_2^* + \mathrm{Xe} + \text{ heat} & \\ \mathrm{Xe}_2^* \rightarrow \mathrm{Xe} + \mathrm{Xe} + h\nu & \text{VUV emission} \end{array} \tag{2.6.3}$$

The recombination fraction depends on both the applied electric field and the recoil energy of the interaction.

During the scintillation process, $Xe_2^* \rightarrow Xe + Xe + h\nu$, the molecular Xe excimer decays with both fast and slow components, corresponding to the de-excitation of singlet and triplet states, respectively. This process can be described by:

$$\text{Emission Time } = C_1 e^{-t/\tau_1} + C_3 e^{-t/\tau_3} \tag{2.6.4}$$

where τ_1 and τ_3 are the time constants of the decay of the singlet and triplet states, respectively. For LXe, measurements of τ_1 range from 2 to 4 ns, while measurements of τ_3 range from 21 to 28 ns. The ratio of singlet to triplet states, $(C_1\tau_1/C_3\tau_3)$ depends on the interaction type, so the pulse shape will vary between ER and NR events [104]. This effect can be exploited in the method of pulse shape discrimination, particularly in argon TPC experiments, as the LAr decay times are separated by two orders of magnitude, with $\tau_1 = 7$ ns and $\tau_3 = 1.6$ µs [105]. Although this is an appealing property of liquid argon, detectors using this target material require an exposure roughly five times that of LXe experiments due to the enhancement of the interaction cross-section for xenon [100].

2.6.3 Secondary Scintillation: S2 Signal

Ionisation electrons that do not undergo recombination are drifted, under the influence of an electric field, to the liquid surface, where they must overcome a potential barrier to be extracted into the gas phase. This is achieved through the application of a much stronger extraction field, which accelerates the electrons to sufficient energies to excite xenon atoms in the gas, leading to secondary scintillation, known as electroluminescence. The light yield per cm depends on both the electric field strength, E, and the gas pressure, P, through

$$\frac{dN_{ph}}{dx} = \alpha E - \beta P - \gamma. \tag{2.6.5}$$

The gas-specific coefficients have been calculated to be $\alpha = 0.137\,\text{V}^{-1}$, $\beta = 177$ bar^{-1}cm^{-1} and $\gamma = 45.7$ cm^{-1} for xenon [101]. The secondary scintillation signal is larger than the S1 signal by orders of magnitude. For example, for values of $E = 10.2$ kV and $P = 1.8$ bar, $\mathcal{O}(10^3)$ VUV photons would be produced by a single electron travelling over a distance of 0.8 cm between the liquid surface and the anode [106]. It is therefore sensitivity to the S1 signal that limits the minimum detectable recoil energy.

2.6.4 Light and Charge Yields

Following from Equation 2.6.1, the energy of an interaction can be written as

$$E_0 = \frac{W}{L}(N_i + N_{ex}), \tag{2.6.6}$$

where W is the energy required to produce an ionisation electron or excimer, which has a value of 13.5 eV for LXe, using the linear function presented in Ref. [107]. L is a nuclear quenching factor, which accounts for energy lost to dissipated heat. The number of photons, n_γ, and electrons, n_e, escaping an interaction site can be related to N_{ex} and N_i by the probability of recombination, r through

$$n_e = N_i(1 - r) \tag{2.6.7}$$

$$n_\gamma = N_{ex} + N_i r, \tag{2.6.8}$$

allowing Equation 2.6.6 to be rewritten as [108]

$$E_0 = \frac{W}{L}(n_e + n_\gamma). \tag{2.6.9}$$

It should be noted that

$$n_e + n_\gamma = N_{ex} + N_i \tag{2.6.10}$$

is always true, regardless of the recombination fraction.

The S1 and S2 signals can be written in terms of n_γ and n_e as

$$\mathrm{S1} = n_\gamma g_1 \tag{2.6.11}$$

$$\mathrm{S2} = n_e(\epsilon N_{ph} g_{1,\mathrm{gas}}) = n_e g_2, \tag{2.6.12}$$

where g_1 is the light collection efficiency of scintillation photons and $g_{1,\mathrm{gas}}$ is the light collection efficiency in the gas phase. ϵ is the electron extraction efficiency and N_{ph} is the number of electroluminescence photons per electron. Equation 2.6.9 can therefore be rewritten as

$$E_0 = \frac{W}{L}\left(\frac{\mathrm{S1}}{g_1} + \frac{\mathrm{S2}}{g_2}\right). \tag{2.6.13}$$

For ER interactions, the nuclear quenching factor, L is assumed to be 1, as minimal energy is lost to heat, whereas for NRs it varies with recoil energy and can be modeled by the Lindhard theory [103], [109]. S1 and

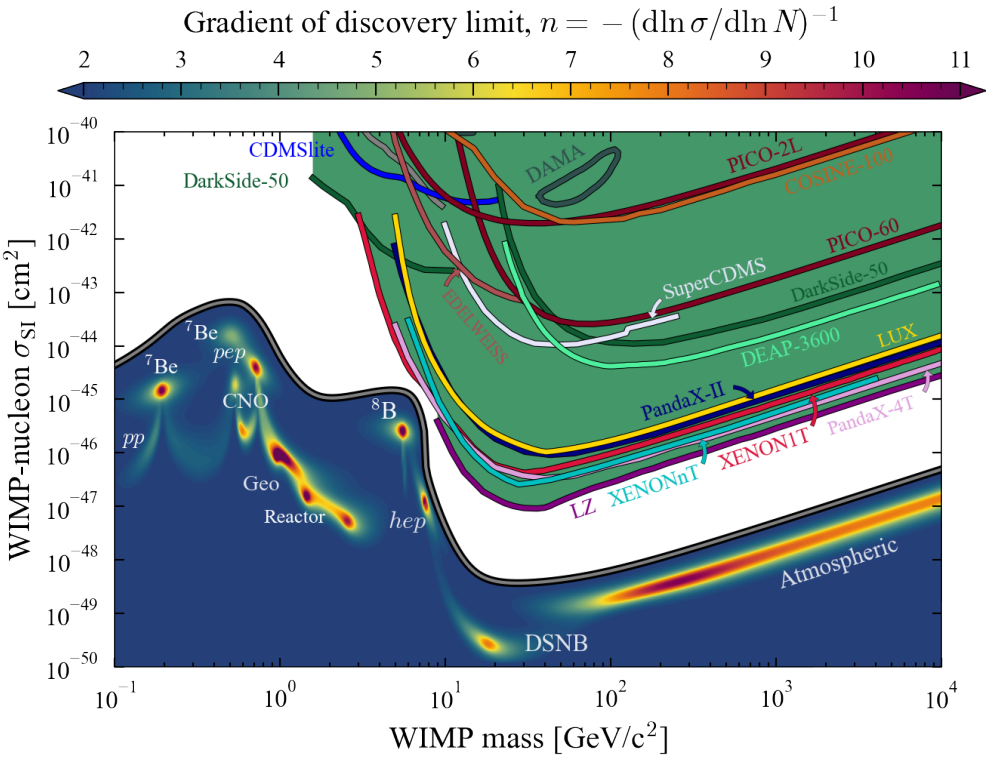


Figure 2.9: Reproduction of Figure 1.7 from [73], which was originally based on Ref. [111], showing the current exclusion limits for spin-independent WIMP-nucleon scattering. The gradient of the neutrino fog, $n = -(\mathrm{d}\ln\sigma/\mathrm{d}\ln N)^{-1}$, represents the difficulty in claiming a WIMP observation, with a larger value of n requiring a higher number of events to report a discovery. The different sources of neutrinos are indicated, including the DSNB (diffuse supernova neutrino background).

S2 signal sizes are measured in units of photons detected (phd) rather than photoelectrons, to account for the $\sim 20\%$ probability of one VUV photon inducing two photoelectrons at the PMT photocathode, known as the double photoelectron (DPE) effect [110].

2.7 Current Status of Direct Detection

Figure 2.9 shows the current limits in the parameter space of WIMP mass and WIMP-nucleon cross-section set by direct detection experiments, assuming SI interactions. Results from the DAMA/LIBRA collaboration have been demonstrating an annual modulation signal consistent with dark matter for over 20 years, reaching a significance of $\sim 12\sigma$ [82]. The inconsistency of the DAMA/LIBRA signal with the limits provided by null results from other direct detection experiments has been a cause for concern. It has been suggested that the modulation could be due to changes in the detector environment or the analysis methods used. The DAMA/LIBRA collaboration subtracted the average rate over each

year to obtain the residual rate, with each cycle of data taking starting at roughly the same time of year. A slowly increasing background rate could lead to a sawtooth wave, which also fits the data well and could provide an explanation for the observed modulation [112]. In 2021, the ANAIS-112 experiment released results after three years of searching for an annual modulation signal, using the same detector material of NaI(Tl), that were incompatible with DAMA/LIBRA at 3.3 and 2.6 σ in [1-6] and [2-6] keV energy ranges [84]. The SABRE collaboration will further test the DAMA/LIBRA claim using similar detectors in the northern and southern hemispheres to disentangle seasonal effects, such as a temperature dependence [113].

The LUX-ZEPLIN experiment has set the most stringent limits for WIMP masses above 9 GeV, with cross-sections above $9.2 \times 10^{-48}\,\mathrm{cm}^2$ rejected at a confidence level of 90% for a WIMP mass of 36 GeV [93]. Above a WIMP mass of 3 GeV, the strongest constraints are all set by experiments which use LXe TPCs, which have been leading the field in these mass ranges since the first results from ZEPLIN-II and XENON10 in 2007. The competition between LXe projects has led to developments in technology that have since been adopted by the community. Examples include improvements to photosensors, purification methods, calibration sources and the Noble Element Simulation Technique (NEST), which models light and charge yields, as well as detector responses [100].

Searches for low-mass WIMPs are limited by detector thresholds, but sensitivity to this parameter space could be achieved using the Migdal effect. It is predicted that during a scattering process, a small displacement of the target nucleus relative to the electron shells can, with a small probability, excite or ionise the atom. Sub-GeV dark matter that would produce undetectable NRs of sub-keV energy could produce keV-scale ERs through the Migdal effect, above the detection threshold of several existing detectors [114]. Although the Migdal effect is yet to be observed, several experiments have reported improved WIMP exclusion limits by incorporating the process into their signal models, with Darkside-50 achieving the most stringent limits in the 40 MeV - 3.6 GeV mass range [115].

Direct detection experiments are limited by the so-called neutrino floor, where WIMP-nucleus scattering becomes indistinguishable from coherent neutrino-nucleus scattering due to solar neutrinos, atmospheric neutrinos and the diffuse supernova neutrino background (DSNB), a hypothesised population of neutrinos produced in the core collapse of supernovae throughout the history of the Universe [116]. It should be

noted that this limit can in fact be overcome with sufficient statistics for most WIMP masses, unless the WIMP and neutrino signals are an exact match. As well as this, the position of the limit is dependent on a choice of astrophysical parameters. Therefore, the term "neutrino fog" has recently been adopted by some [111]. In Figure 2.9, the neutrino fog is defined by the parameter n, which describes how a WIMP discovery limit scales with the number of observed background events. A higher value of n represents an increased number of WIMP events required to claim a discovery. Directional detectors could provide an avenue to search for dark matter beyond the neutrino floor.

Projected sensitivities for the LUX-ZEPLIN experiment leave only a further order of magnitude of available parameter space before the neutrino fog is reached, which is sufficient to explore with a single future generation three (G3) multi-tonne LXe TPC experiment. The LUX-ZEPLIN, XENON and DARWIN collaborations have formed the XLZD consortium, which aims to create the ultimate xenon observatory, searching for neutrinoless double β decay, solar and astrophysical neutrinos, and other dark matter candidates, as well as WIMPs [117].

Chapter 3

The LZ Detector and Data Quality

The LUX-ZEPLIN (LZ) experiment is based 1478.3 m underground in the Davis Cavern at the Sanford Underground Research Facility (SURF) in South Dakota. LZ utilises two-phase xenon technology to operate low-background searches for WIMP signals. This chapter will discuss the LZ detector and software, as well as backgrounds and data quality, which must be well-understood in order to produce a robust science result.

3.1 The LZ Detector

As shown in Figure 3.1, LZ is a multi-detector experiment, consisting of an inner TPC containing the LXe target material, surrounded by two active veto systems, the Outer Detector (OD) and Skin, for background rejection and in-situ characterisation. These primary components will be discussed in the following sections.

3.1.1 The Time Projection Chamber

A cylindrical TPC containing a 7-tonne active volume of LXe, with a thin layer of gaseous xenon (GXe) at the top, is situated inside two titanium cryostats: the Outer Cryostat Vessel (OCV) and Inner Cryostat Vessel (ICV). The TPC is used to detect the primary scintillation (S1) and secondary electroluminescence (S2) signals described in chapter 2. Whereas S2 signals are relatively large due to the high charge yield in GXe, some S1 pulses consist of only a small number of photons, so it was necessary to design the TPC to maximise their collection.

Light is collected by 494 3-inch R11410-22 Hamamatsu photomultiplier tubes (PMTs) arranged in arrays at the top and bottom of the

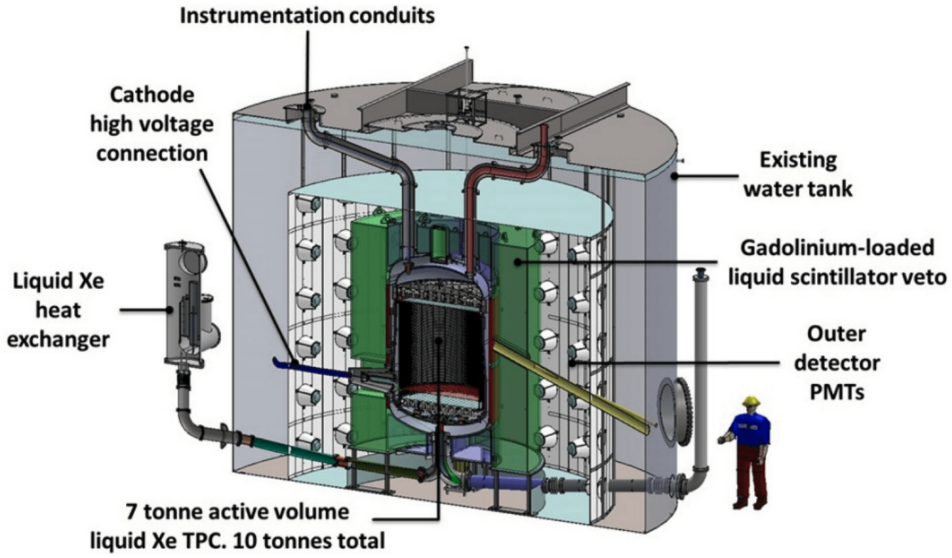


Figure 3.1: Schematic of the LZ experiment from Ref. [106].

TPC. Incident photons propagate through PMT windows, which are made from quartz in order to maximise sensitivity in the VUV range within which S1s and S2s are emitted in xenon. Photons then encounter the photocathode, where electrons are produced via the photoelectric effect. Electron multiplication occurs at a series of dynodes, before collection at the PMT anode. The PMTs were designed specifically for high performance at LXe temperatures and ultra-low radioactivity [118]. The TPC is lined with highly-reflective polytetrafluoroethylene (PTFE). The majority of S1 light is collected by bottom array PMTs, predominantly a result of reflections at the liquid surface, and the immersion of bottom array PMTs in LXe, which has a similar refractive index to the quartz of the PMT windows. The 241 bottom-array PMTs are therefore arranged in a tightly-packed hexagonal array to optimise S1 light collection.

Four high-voltage stainless steel mesh grids produce the vertical electric fields in the TPC, with field-shaping rings in the PTFE walls maintaining field-uniformity. The grids divide the TPC into three regions, as shown in the schematic in Figure 3.2. The first is a reverse field region (RFR) between the bottom and cathode grids. The bottom grid has a small negative bias and exists to protect the bottom PMT array from the strong fields above it. In the second region, between the cathode and gate grids, the drift field causes ionisation electrons to travel upwards. The final region is the extraction region between the gate and anode grids. The electrons are emitted from the liquid into the gas layer by the stronger extraction field between these grids, where they gain sufficient energy to produce the secondary scintillation light known as the S2 signal [106].

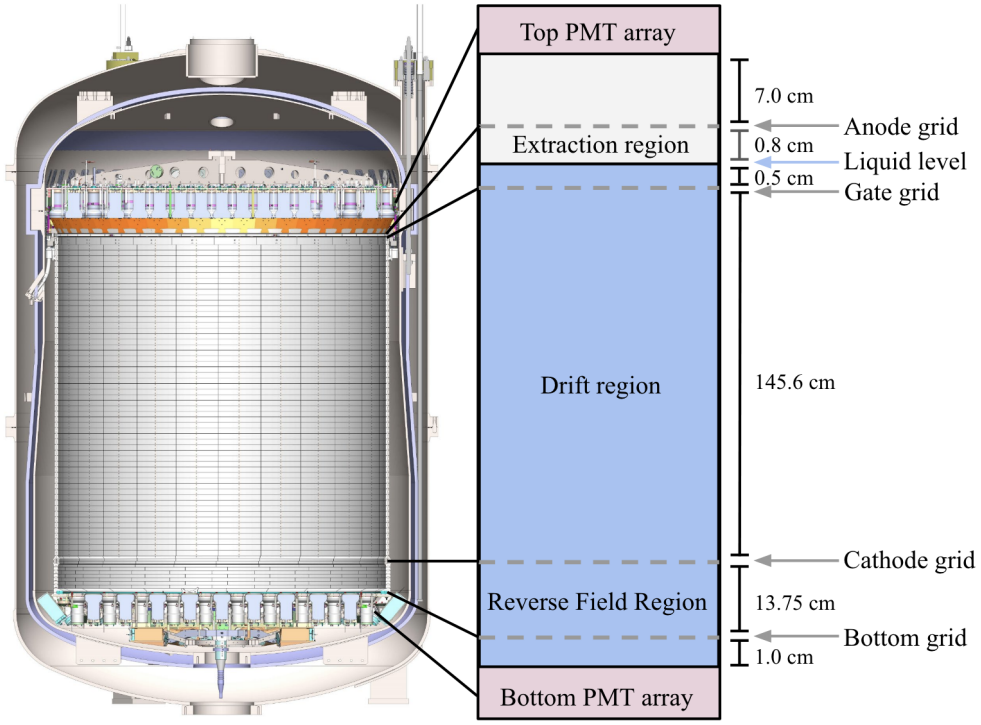


Figure 3.2: Schematic of the TPC inside the ICV and OCV alongside a diagram showing the positioning of the grids and the different field regions that they produce, based on Fig. 2 from Ref. [119]. Blue indicates the LXe, whereas light grey represents the GXe.

Accurate XY position reconstruction, using the location of S2 light in the top PMT array, is vital, particularly to avoid leakage of backgrounds from near the PTFE walls towards the detector centre. Therefore, unlike the bottom array, top-array PMTs overhang the edges of the TPC (where they transition to an evenly-spaced, circular arrangement) to avoid an inward reconstruction bias. As described in chapter 2, the Z-position (or "drift time") is calculated from the time difference between the two signals, using the constant speed of the drifting electrons. The relative ionisation yield is used to distinguish between electron and nuclear recoils.

3.1.2 Veto Detectors

As WIMPs are only expected to scatter once, two veto detectors are employed to tag events with additional energy deposits outside the TPC. The Skin detector is situated inside the ICV, surrounding the TPC, from which it is optically separated. Containing no gaseous xenon, the Skin is a scintillation-only detector. It holds an additional 2 tonnes of LXe that is divided between a cylindrical side-skin region and a dome underneath the TPC. The Skin acts as a scintillator to tag scatters primarily from internal γ rays, as well as neutrons. 93 1-inch Hamamatsu R8520 PMTs

and 20 2-inch Hamamatsu R8778 PMTs view the side skin, while 18 2-inch Hamamatsu R8778 PMTs observe the dome [120].

The OCV is enclosed by an Outer Detector (OD), which acts as a γ and neutron anti-coincidence system, consisting of ten tanks filled with a total of 17.3 tonnes of gadolinium-loaded liquid scintillator (Gd-LS). Gadolinium efficiently captures neutrons, a problematic NR background to WIMPs, which release a γ-ray cascade of up to $\sim$ 8 MeV in total energy. The OD was designed to have a $> 95\%$ neutron tagging efficiency, as the scintillation light produced in the subsequent interactions in the LS generally occur within the same event window as the S1 [121]. 120 8-inch Hamamatsu R5912 PMTs, arranged on ladders around the detector, collect OD scintillation light [106].

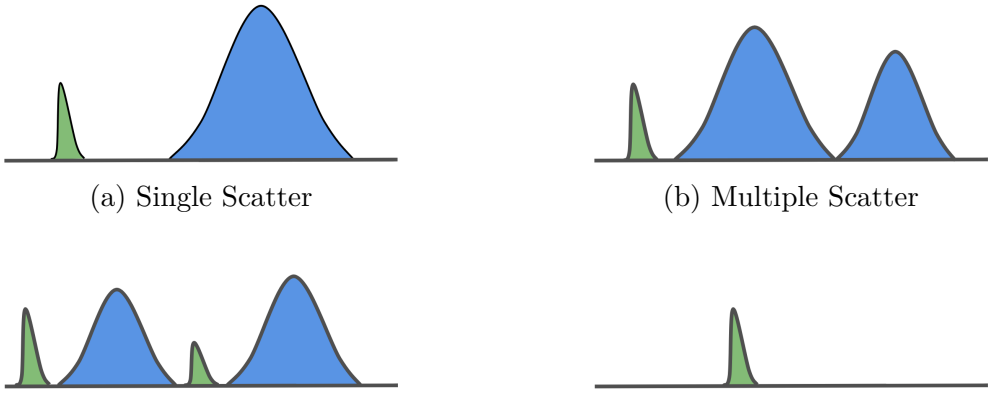
Finally, a water tank holding 228 tonnes of ultrapure water provides further shielding from cavern γs and neutrons.

3.2 LZ Data and Software

Following amplification and shaping, raw signals from PMTs are passed to the LZ data acquisition (DAQ) system for digitisation by a 32-channel Analogue-to-Digital Converter card. In order to reduce the raw waveform volume, data are collected in Pulse Only Digitisation (POD) mode, where only groups of contiguous signals above a threshold in each PMT are stored in circular buffers. Both high- and low-gain signals are acquired, to account for the high dynamic range required to detect small S1 and large S2 signals. A filter is applied to select useful data, which are retrieved and compressed by a Data Extractor and sent to a Data Collector for temporary storage, before being written to disk [106]. WIMP search events are typically triggered by S2s, with the amplitude, channel multiplicity and coincidence window all contributing to the trigger logic. The nominal total rate of S2 triggers is $\sim$ 5 Hz during data taking for the WIMP search. Event windows span from 2 ms before to 2.5 ms after the trigger.

The LZ Analysis Package (LZap) generates Reduced Quantities (RQs) from raw waveforms at the levels of individual PMT channel pulses, detector pulses (generated by summing channel pulses), and the entire event. One such pulse-level RQ, used to describe the signal size, is the integrated pulse area.

LZap contains algorithms to select and classify pulses. Physical pulses with a PMT coincidence of one are, by default, single photoelectron (SPE) pulses. RQs describing the shape of the pulse are employed to



(a) Single Scatter (b) Multiple Scatter

(c) Pile-up (d) Other

Figure 3.3: Examples of the possible interaction types, as categorised by LZap, with S1 pulses shown in green and S2 pulses in blue.

distinguish between S1-like and S2-like pulses, as S1s typically have a more positive skew than S2s, which experience electron diffusion. S1-like pulses must meet an S1 coincidence requirement of three, else they are classified as multiple photoelectron (MPE) pulses. S2-like pulses are categorised as single electrons (SEs), unless they meet a minimum S2 area requirement. An "Other" category exists for all remaining pulses that do not fulfil any of the criteria. Position reconstruction of S2 pulses is achieved using the Mercury LZap algorithm, which determines the coordinates and scintillation intensity that provide the optimal similarity between the observed PMT hit pattern and a light response model [122].

An interaction-finding LZap algorithm analyses the size and shape of pulses within an event to identify prominent S1s and S2s from the primary physics interaction. These are the pulses that govern the type of interaction, leading to classification of events as: Single Scatter (SS), for a single S1 proceeding a single S2; Multiple Scatter (MS), for a single S1 followed by multiple S2s; Pile-up, for multiple S1s and S2s in various orders; and Other, for noise or isolated S1s or S2s. Examples of each topology are shown in Figure 3.3.

Only SS events are considered as candidate WIMPs. The LZ Offline Event Viewer is a web-based tool allowing for inspection of waveforms in all three detectors within an event, with LZap-derived properties, such as pulse classifications and areas, also available to view. A PMT map indicates the hit pattern of light collected. An example SS event on the Offline Event Viewer webpage is shown in Figure 3.4, with magnified images of the S1 and S2 waveforms in Figure 3.5 to demonstrate typical pulse shapes. This event has signals in the Skin and OD that are co-incident with the S1 pulse in the TPC. This corresponds to a neutron scattering in all three detectors. As this neutron scatters a single time in

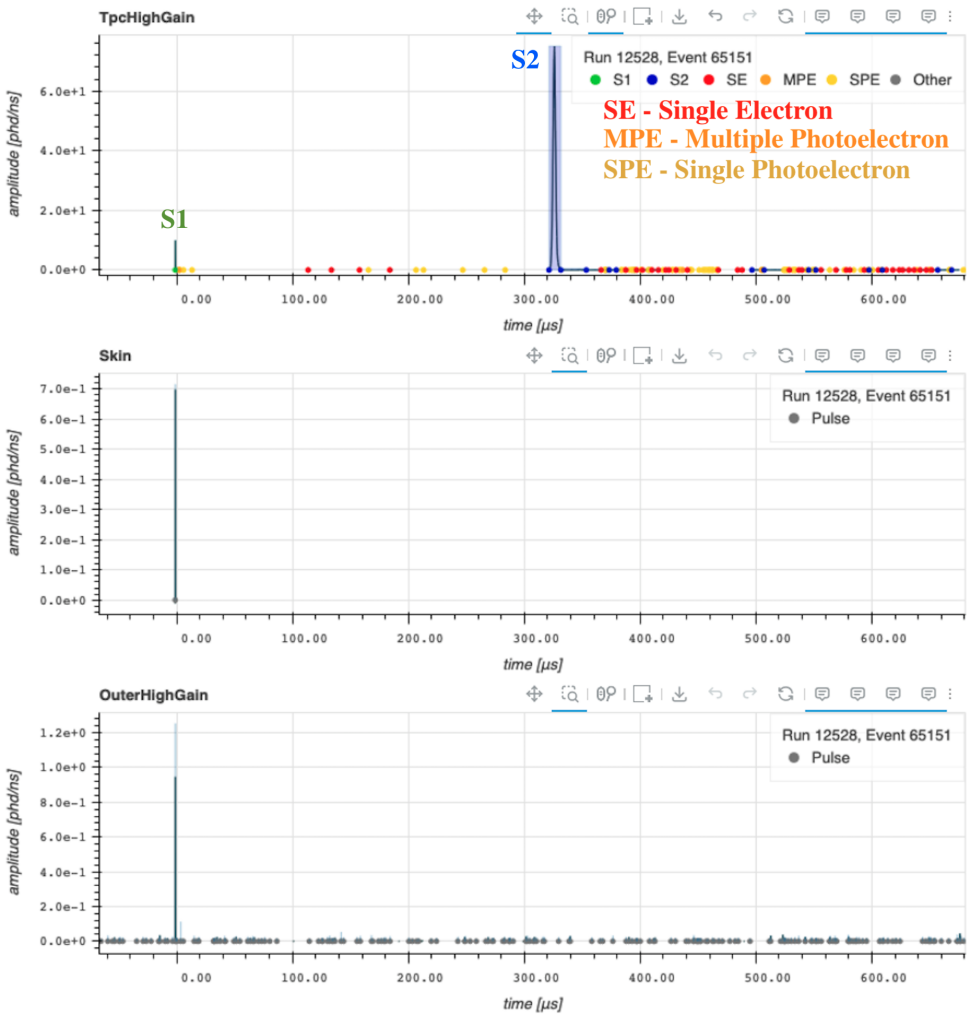
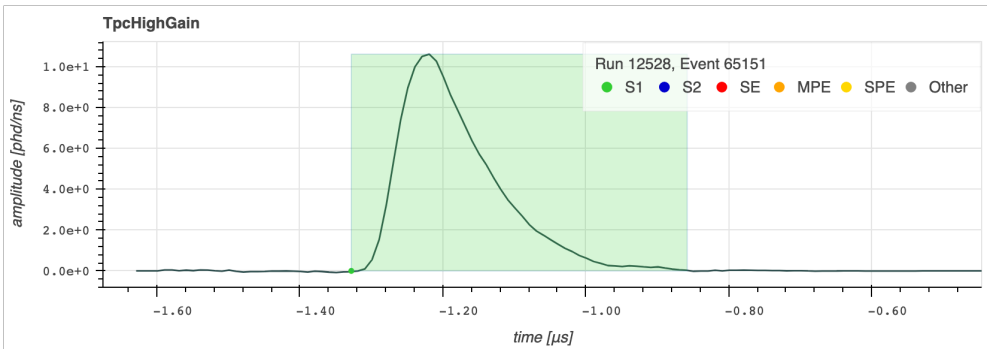


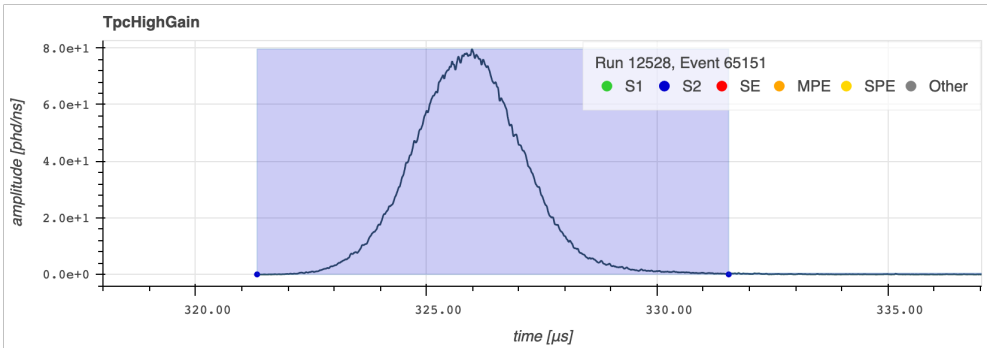
Figure 3.4: From the LZ Event Viewer webpage: a Single Scatter neutron event with a signal in the Skin (middle image) and OD (bottom image) detectors coincident with the S1 pulse in the TPC (top image), corresponding to the neutron scattering in all three detectors. The large S2 pulse and smaller single electrons, multiple photoelectrons and single photoelectrons can also be seen.

the TPC it contributes an important background to the WIMP search, but the coincident signals in the veto detectors would allow this event to be excluded.

LZap RQ files can be used as inputs for modules within the Analysis LZ Package (ALPACA); the C++ analysis framework tailored for LZ. ALPACA standardises analyses by handling underlying features in a shared codebase, and the modular natures allows for easy running of code written by other users. Analysis code is run over data on an event-by-event basis, and generally involves implementation of cuts and generation of histograms, simplified by services within the ALPACA framework.



(a)



(b)

Figure 3.5: From the LZ Event Viewer webpage: magnified waveforms for the S1 (a) and S2 (b) pulses from the event shown in Figure 3.4.

3.3 Backgrounds

Backgrounds must be mitigated as far as possible, and well-understood to allow for accurate modelling. The sources of backgrounds in LZ, both internal and external, will now be discussed. For LZ to achieve sensitivity to WIMP-nucleon cross sections below $3 \times 10^{-48}\,\mathrm{cm}^2$ within three years of data taking, the maximum allowable rate of ER backgrounds from non-astrophysical sources is 37×10^{-6} events/keV/kg/day, with ~ 1 NR background event during the exposure [106]. The collaboration set significantly stricter goals for all backgrounds to reduce the risk of any component violating these requirements. Determination of expected number of events for the most dominant backgrounds in the SR1 WIMP search is discussed in subsection 4.6.2.

3.3.1 Detector Components

Neutron and γ-ray emission in LZ occurs predominantly from naturally-occurring radioactive nuclides within detector materials. This includes the γ emitting ^{40}K and ^{60}C, as well as ^{238}U and ^{232}Th and their progeny, which also generate neutrons via (α, n) reactions and spontaneous fission. Neutrons generally scatter multiple times, but SS neutron NR events are

indistinguishable from WIMPs and rely on OD tagging, which is not 100% efficient. The LZ goal for backgrounds resulting from fixed contamination in detector components is less than 0.4 NR counts and below 1×10^{-6} events/keV/kg/day of ER background, for SS events within the fiducial volume and WIMP search energy range during a three-year exposure [106]. An extensive radioassay campaign was used to screen all materials to sensitivities of $\sim$0.2 mBq/kg. The contribution of detector component backgrounds could therefore be well-defined, as well as minimised through material selection [123].

3.3.2 Surface and Dispersed Contaminants

^{222}Rn and ^{220}Rn can emanate from detector components, as well as from dust that has settled on detector materials during construction, and disperse in the LXe. This is the largest contribution to the expected number of events during WIMP search campaigns, with the dominant radon background resulting from "naked" β emission (not accompanied by γ rays) from the decay of ^{214}Pb to ^{214}Bi. Radon emanation can be attributed to both the recoil of radon atoms and diffusion, which is highly dependent on the material. Detector components were therefore screened for radon emanation [123], with a maximum activity of 10 mBq required throughout the LXe and a goal of 1 mBq [106]. For the avoidance of radioactive dust, transport and assembly of detector components were required to comply with strict cleanliness protocols, with a requirement limiting dust contamination to less than 500 ng/cm^2 on wetted surfaces [124].

To maintain purity, the xenon is continuously circulated through a purification system. As part of this process, the LXe is vapourised and passed through a hot zirconium getter to remove electronegative impurities resulting from outgassing, such as oxygen [120]. As noble elements are not removed by the getter, the xenon was purified using a process of charcoal chromatography before transport to SURF, to reduce contamination from Kr and Ar present in natural xenon, with requirements of 0.02 ppt (g/g) and 4.5×10^{-10} (g/g), respectively [106]. The presence of these noble elements leads to a flat ER background from ^{85}Kr and ^{39}Ar [125].

3.3.3 Cosmogenic and Laboratory Backgrounds

During above-ground storage and transport of the xenon, cosmogenic activation led to the production of ^{127}Xe and ^{37}Ar. Following underground deployment, activation from cosmic rays is no longer significant. ^{127}Xe

and ^{37}Ar have half-lives of 36.4 days and 35.1 days, respectively, so during commissioning and early SR1 acquisitions, these isotopes are still present, but are decaying away. ^{127}Xe decays through electron capture, and the γ produced by the de-excitation of the ^{127}I daughter can usually be tagged by the Skin. In the case that it goes undetected, the L- and M- shell electron captures produce an ER background within the WIMP search region of interest (ROI) [93]. ^{37}Ar also decays through electron capture, with an 2.82 keV Auger cascade from K-shell capture contributing an ER background in the ROI [126]. Cosmogenic activation backgrounds are accounted for in the background model (see subsection 4.6.2).

By locating the detector deep underground, the cosmic-ray muon flux is reduced by a factor of 10^6 compared to at sea level [127]. However, the surrounding rock produces γ-ray backgrounds emitted from the ^{238}U and ^{232}Th decay chains, as well as ^{40}K [128]. Remaining muon-induced cascades can produce neutrons, but, following attenuation in the water tank and tagging by the OD, studies using simulations found that their rate can be considered to be negligible [129].

3.3.4 Physics Backgrounds

Two rare Standard Model processes in xenon contribute a physics background for LZ. The two-neutrino double electron capture (2νDEC) of ^{124}Xe can combine captures from K-, L-, M- and N-shells. The LL decay mode (in which both electrons are captured from the L-shell) and LM decay mode (in which electrons are captured from both the L- and M-shells) produce backgrounds in the WIMP search ROI. The XENON collaboration first observed ^{124}Xe 2νDEC [130] and has since measured its half-life (1.1×10^{22} years) and branching ratio (1.4% and 0.8% for the LL and LM modes, respectively), allowing rates to be modelled [131]. Two-neutrino double-beta ($2\nu\beta\beta$) decay of ^{136}Xe has also been observed and, with a half-life of 2.165×10^{21} years, contributes an ER background up to 2457.8 keV [132].

The dominant NR background to the WIMP search is due to coherent elastic scattering of neutrinos with xenon nuclei, predominantly from solar ^{8}B, with smaller contributions from the *hep* reaction, as well as atmospheric and diffuse supernova neutrinos. *pp* chain neutrinos and, to a lesser extent, ^{7}Be and CNO, produce an additional ER background due to electroweak interactions with the xenon [125].

3.3.5 Accidental-Coincidence Backgrounds

Isolated S1 and S2 pulses may be misidentified as a Single Scatter event if their time separation is less than the maximum possible drift time. Accidental-coincidence events are a significant background for the LZ WIMP search, particularly at low energies, and will be a primary focus of chapter 4.

3.4 Detector Conditions

As data from each of the three science runs, referred to as SR1 (used for the first WIMP-search result), SR2 and SR3 (ongoing), will be discussed, the detector conditions for each are summarised in Table 3.1. A campaign of testing various configurations of electrode voltages was undertaken before the start of SR2, in an effort to reduce accidental-coincidence backgrounds from the grids. This resulted in a reduced extraction field strength, while maintaining the same drift field. Data quality issues experienced during SR2 will be discussed in section 3.9, and ultimately resulted in a reduction of the drift field for SR3, in order to prioritise the health of the detector. As a consequence, the maximum drift time increased from 951 µs to 1046 µs. Throughout SR1, the temperature and pressure of the LXe were stable to within 0.2%, at 174.1 K and 1.791 bar(a), respectively. This was not the case for the other science runs, as during both SR2 and SR3 there was a point at which the temperature gradient in the TPC was inverted in order to produce a state with less mixing of the LXe. This resulted in larger deviations in temperature and pressure, but the mean values are also shown in Table 3.1.

Table 3.1: The values of key parameters during the SR1, SR2 and SR3 WIMP search campaigns.

Detector Parameter	SR1	SR2	SR3
Anode voltage [kV]	4	1.5	3.5
Gate voltage [kV]	-4	-6	-4
Cathode voltage [kV]	-32	-34	-18
Bottom voltage [kV]	-1.25	-1.25	-1.25
Drift field [V cm^{-1}]	193	193	97
Extraction field [kV cm^{-1}]	7.3	3.39	3.38
Maximum drift time [µs]	951	951	1046
Temperature [K]	174.1	174.7	175.0
Pressure [bar(a)]	1.791	1.858	1.860

3.5 Data Quality Requirements

Since the expected WIMP signal is so rare (< 0.0001 events/kg/day), it is essential to monitor data and constantly check for problems. Irregularities can correspond to physical changes in the detector itself (such as the presence of radioactive backgrounds), or the electronics system (such as altered PMT behaviour). As well as this, changes to online configurations (such as an incorrect acquisition mode), or offline data processing (such as updates to LZap) could cause issues. Therefore, it is important to identify data quality defects, in order to influence decisions regarding whether each "run" (an acquisition typically lasting several hours) can be used for the WIMP search or other analyses. Prompt feedback on detector changes should also be provided to experimental operators, so that adjustments can be made accordingly.

3.6 The Physics Readiness Monitor

The Physics Readiness Monitor (PREM) is the offline Data Quality Monitor for the LZ collaboration and aims to rapidly provide a first look at datasets within hours of LZap processing. ALPACA analysis modules can be modified to produce metadata stored in a nested structure inside a JSON file, which is sent to a PREM DB (database). An output ROOT file containing histograms is stored at the National Energy Re-

search Scientific Computing Center (NERSC). The PREM frontend is a website known as PREMweb, where the user submits requests for data to view. PREMweb uses a framework consisting of Cascading Style Sheets (CSS) and HyperText Markup Language (HTML). A screenshot of the top of the PREM webpage is shown in Figure 3.6. Dropdown menus are used to select a Run Type (a campaign such as a specific WIMP search or series of calibrations), run number, PREM module and LZap version. These selections also fill a string of key-value pairs within the URL, allowing for easy distribution of a chosen analysis among collaborators. The backend is a NodeJS application that receives these requests, and queries the database for the relevant information. The histograms are then displayed on PREMweb. Three core PREM modules observe the general health of each of the detectors, while additional dedicated modules monitor specific detector components or physics quantities.

3.6.1 The TPC Health Module

As the TPC is the detector in which a WIMP interaction may be detected, it is essential that any irregularities in its behaviour are identified rapidly. A minimal selection of plots were chosen for the TPC_Health module to allow quick inspection of data. These were categorised according to the navigation tabs shown in Figure 3.6, and include monitoring of the grid health, the rates of pulses, and the position reconstruction performed within LZap. This information has proved to be critical for data quality monitoring, with all observed defects discussed in section 3.9 identifiable using the TPC_Health module. It is expected that offline shifters will conduct PREM detector checks in future, so detailed module documentation was produced, which included example plots of expected behaviour from “golden” runs of known good quality.

3.7 Existing PREM Features

Several PREMweb tools, which had previously been developed to assist with data quality monitoring, will now be described.

3.7.1 One-dimensional Comparison Plots

Presenting data alongside a golden run allows for easy identification of irregularities. After loading the original dataset, a reference run is chosen using the dropdown menu shown in Figure 3.6. For each new run selected, a one-dimensional histogram is overlaid in a different colour,

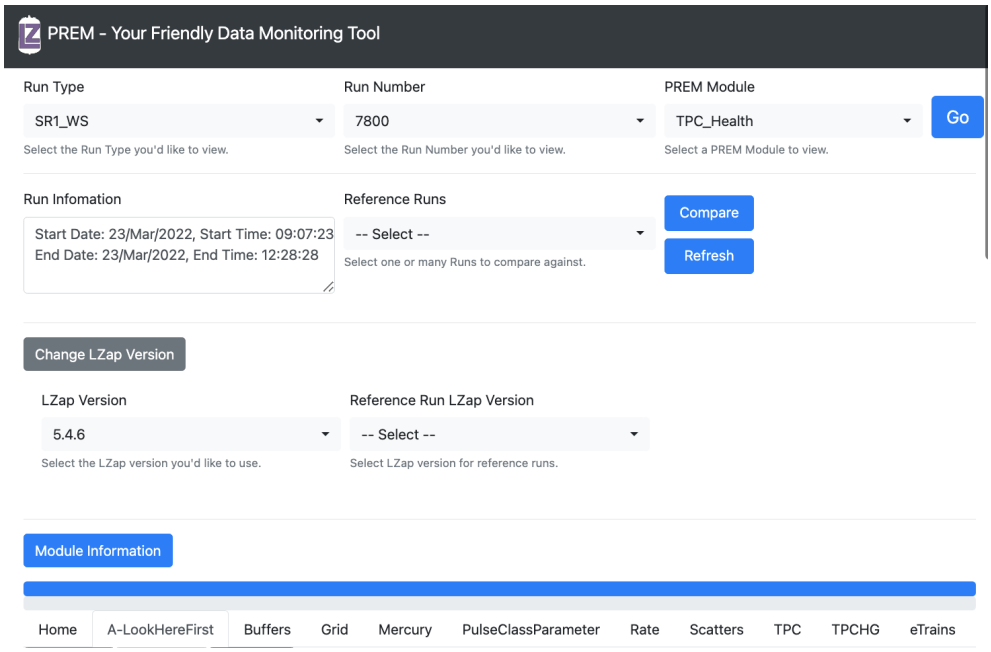


Figure 3.6: Screenshot of the top of the PREM webpage, where a Run Type of SR1_WS, run number of 7800 and PREM module of TPC_Health have been selected. Plots are categorised into the navigation tabs shown.

with a legend displaying the run number, as shown in the upper panel of Figure 3.7 for two runs during SR1. This histogram shows the rate of single electron pulses throughout the run. As these pulses are typically too small to activate the trigger, data from a random trigger were used to achieve an accurate calculation of the rate (indicated by "RND TRG" in the plot title). It can be seen that the chosen run, 7812, has an SE rate that is consistently higher than the reference run, 7800, which signifies an irregularity.

3.7.2 Algorithms

Another useful PREM feature is the algorithm, which is defined as a quantity that can be calculated for each run. The majority of one-dimensional histograms include algorithms monitoring the mean and standard deviation, but more complicated metrics can also be calculated. History plots of algorithm values over many runs are essential, both for detecting transient spikes, and for the monitoring of long-term trends.

The lower panel of Figure 3.7 shows the algorithm history plot for the mean SE rate, which has been calculated within the TPC_Health module for the histogram shown above, and is revealed upon a click of the "meanRate" button. This demonstrates that at run 7812, there was a sharp increase in rate. This corresponds to a known event. As part of the xenon circulation system, the weir reservoir collects the LXe re-

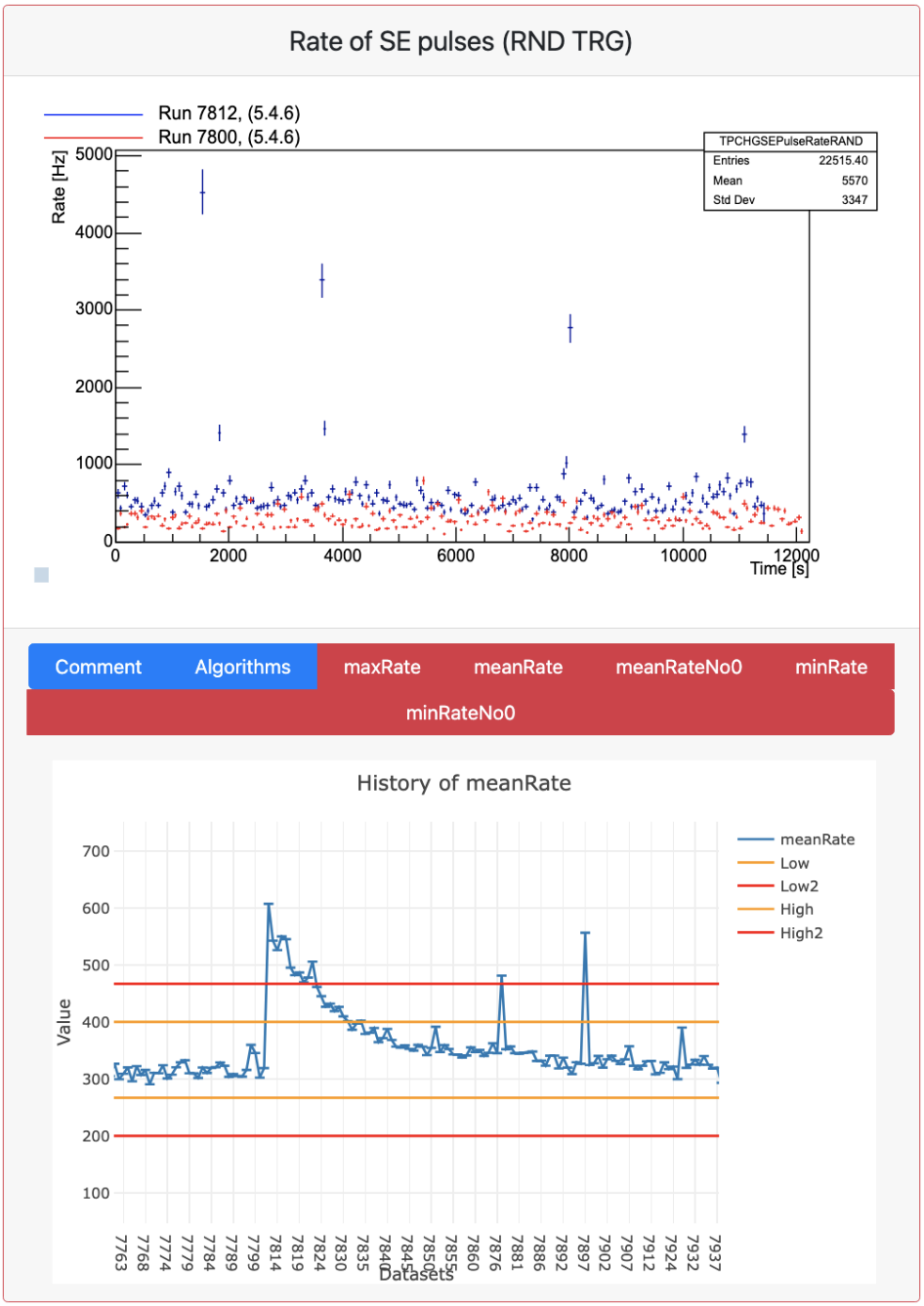


Figure 3.7: Example of the one-dimensional comparison and algorithm features. On the upper panel, the reference run 7800 (before the weir incident) has been overlaid on the histogram showing SE pulse rate for the original run, 7812 (after the weir incident). The "meanRate" algorithm button has been clicked to expand the history plot below. As the algorithm value for the selected run is outside the limits, the button and card border are shown in red. The "meanRateNo0" algorithm performs a similar calculation of the mean rate, but ignores time bins with a rate of zero (for example due to periods of deadtime), which was not relevant for these specific runs.

turning from the detector and prepares it for evaporation. A decrease in TPC pressure caused the contents of the weir reservoir to empty into the TPC, resulting in an increase in impurities. Upper and lower limits are set within the PREM module and, as the value for the selected run, 7812, exceeds the outer red limit, the run is flagged by highlighting the algorithm button and plot border in red. As the circulation system purified the xenon, the SE rate decayed to a nominal value, demonstrating how algorithms can be used to observe trends as well as extreme values.

3.8 Development of PREM Features

As well as general improvements to the user interface and stability of PREMweb, several new features were developed to aid with the evaluation of data quality. This required alterations to the ALPACA core code to change the structure of the JSON metadata; new database queries; and additional elements in the HTML code to create new features on the PREM webpage.

3.8.1 Two-dimensional Comparison Plots

Although it was possible to compare multiple runs for one-dimensional histograms, it was desirable to develop an equivalent feature for two-dimensional comparisons. This was achieved by adding a carousel object to the HTML code, which appears beneath all two-dimensional histograms when a reference run is selected from the dropdown menu. For each new run selected, a new slide is added to the carousel with a caption stating the run number. This allows the user to either compare two runs side by side, or to rapidly cycle through several runs using the carousel arrows, for example to examine how populations move around over the course of several runs. The two runs from before and after the weir incident are shown in Figure 3.8 on a two-dimensional histogram of the XY position of SE pulses. The top panel shows that immediately following the weir event, the pulse rate has increased, and two high-rate regions or "hotspots" are visible, which will be discussed in further detail in subsection 3.9.5.

3.8.2 Algorithm Groups

In some cases it is useful to perform the same calculation many times with different inputs for a single histogram, especially when tracking the same metric for many PMTs. This could be achieved by attaching hundreds of

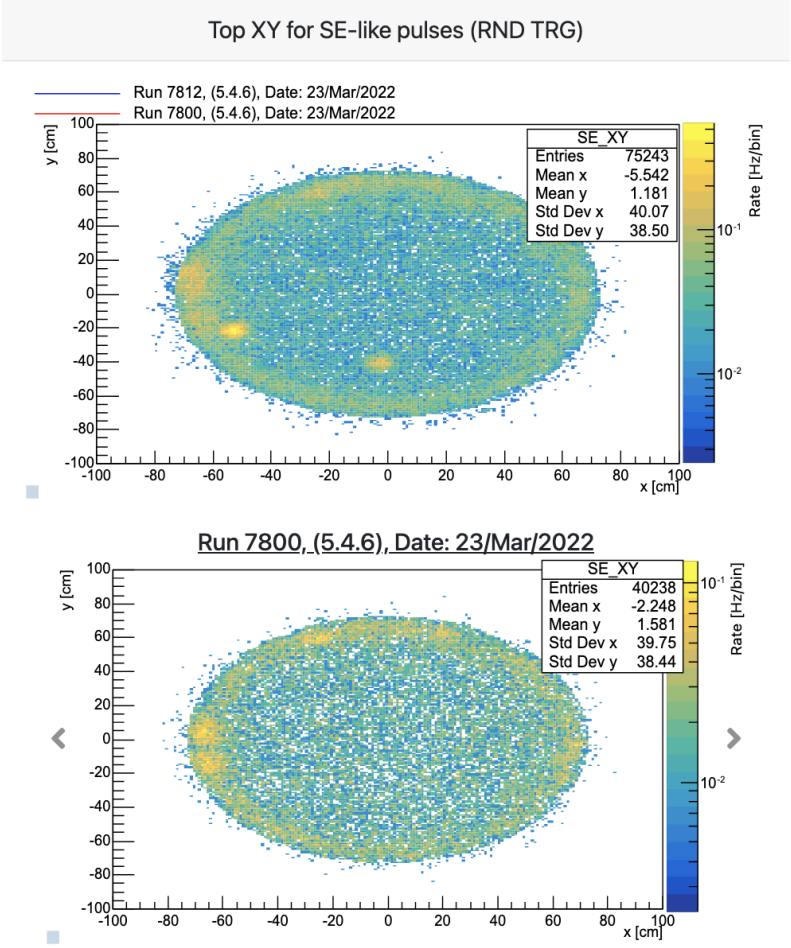


Figure 3.8: Example of the two-dimensional comparison feature. A histogram of the rate of SE pulses in XY space is shown for the original run 7812 (after the weir incident). The reference run 7800 (before the weir event) has been selected, so appears on the carousel below. The carousel arrows can be used to cycle through runs.

similar algorithms to one histogram, but this would require excess code in the PREM module, and many algorithm buttons and separate history plots on PREMweb, making comparisons difficult. This motivated the development of the Algorithm Group, a data structure which contains an array of algorithms.

A searchable dropdown menu and text input box were added to the website interface, allowing the user to select multiple algorithms within the group to view on the same history plot. When the value of an algorithm within a group is outside the specified limits, a typical coloured algorithm button appears to alert the user. Figure 3.9 shows the website display for the "OD_PMT_SPE" module that measures the response of PMTs to a single photoelectron. Here, the mean SPE response in pC is plotted for channels 834, 835 and 836 over several commissioning runs, where an Optical Calibration System was used to inject controlled pulses from LEDs into the OD through optical fibres. The PMT voltages were altered to bring the PMT gains closer together for run 4347 onwards.

3.8.3 LZap Version Comparisons

As well as comparing different runs, it is also useful to make comparisons between the same run processed with different versions of LZap. This allows PREM to monitor the performance of the offline reconstruction software, as well as the quality of the data. Important changes include alterations to the pulse-finding and pulse-classifying algorithms, resulting in different populations of each pulse type, as well as updates to the position reconstruction algorithm and position correction maps. The LZap version was therefore added as a selection in addition to the existing Run Type, run number and PREM module options. The user can make a direct comparison by choosing the same acquisition for the original run and reference run, and selecting a different LZap version for each from a dropdown menu. Figure 3.10 shows SR1 run 6774 processed with both versions 5.4.1 and 5.4.5 of LZap. Updates between these versions included changes to the pulse area thresholds used to select SEs, as well as a new calibration of the SPE size per PMT channel used in LZap. The difference in the distributions of SE pulse area is therefore expected.

3.9 SR2 Data Quality Validation

For SR1 and SR2, the quality of the data was evaluated retroactively, using the three detector health modules: TPC_Health, OD_Health and Skin_Health. All known data quality issues during SR1 were detectable

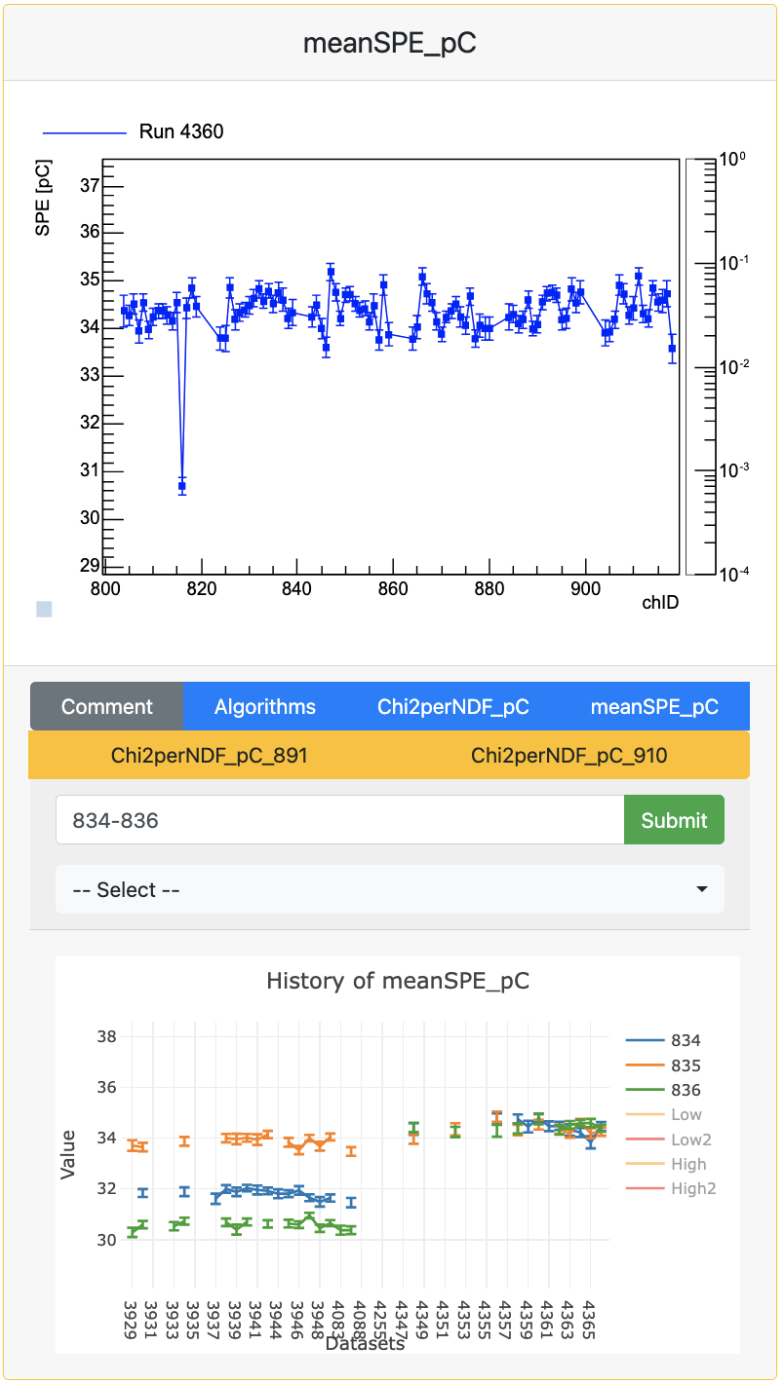


Figure 3.9: This plot was taken from PREMweb as an example of the Algorithm Group feature, and was originally produced using the OD_PMT_SPE module during OD optical calibrations. The "meanSPE_pC" button has been clicked to reveal an algorithm history plot, where the mean SPE response for three selected PMTs is shown. An additional Algorithm Group, "Chi2perNDF_pC" is also attached to the histogram. Two algorithms within this group (corresponding to PMTs 891 and 910) have values outside the inner limits, so have overflowed to create yellow algorithm buttons.

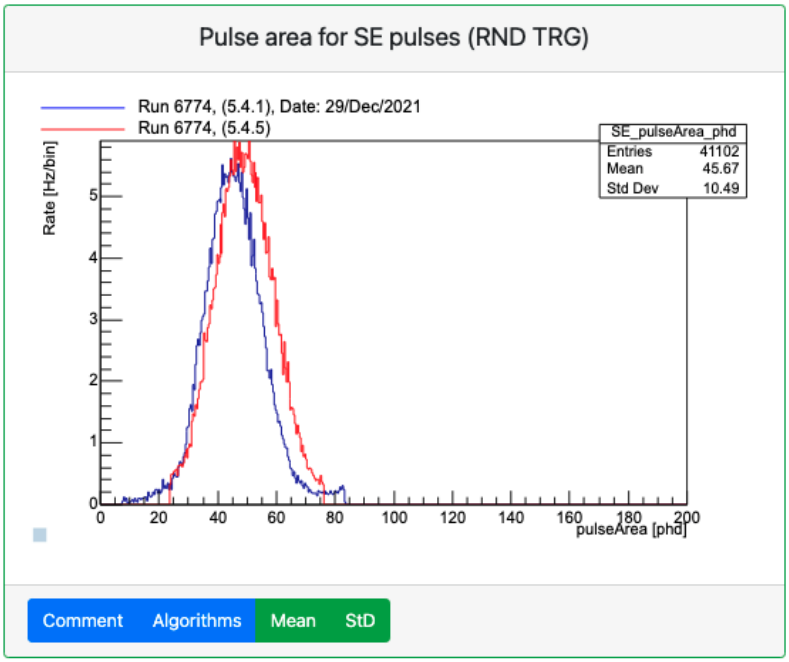


Figure 3.10: Example of the LZap version comparison feature. The SE pulse area for reference run 6774 processed with version 5.4.5 has been overlaid on a histogram for the same run processed using version 5.4.1.

using PREM, and recommendations were made to determine the list of runs suitable for the WIMP search analysis. The evaluation of SR2 data will be the focus of this section. The chosen approach was to inspect algorithm history plots in order to identify both trends and irregularities where the value deviated from the expected range. In the latter case, histogram comparisons were made between the run of concern and a golden run of known good quality. The observations and resulting recommendations from the PREM team will now be described.

3.9.1 Activation of Xenon

For the majority of algorithms, the distributions followed an exponential trend throughout SR1. To account for this when evaluating data quality, exponential limits were defined by performing a fit to the algorithm values as a function of time, and assigning inner and outer thresholds. These were generally situated at 3σ and 4σ, respectively, which were the values deemed to represent a significant deviation from typical behaviour. These limits could be easily displayed using a Jupyter notebook that queries the PREM DB, providing the additional benefits of allowing for multiple Run Types on a single plot, and accounting for gaps in the data taking, rather than jumping to the next value.

An example plot with a clear exponential decay can be seen in Figure 3.11, which shows the total rate of pulses in the TPC as a function of

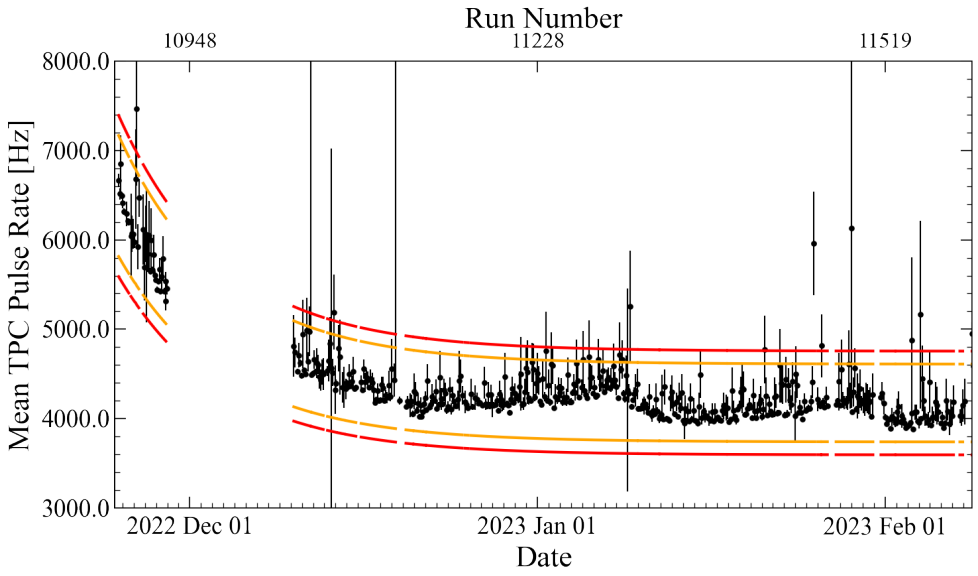


Figure 3.11: Algorithm history plot produced using a Jupyter notebook that queries the PREM DB, showing the mean pulse rate in the TPC as a function of time throughout SR2. The exponential 3σ and 4σ limits are shown in yellow and red, respectively.

time. The increase and subsequent decay in rate is due to the activation of xenon, resulting from neutron calibrations which occurred prior to the start of SR2, leading to additional backgrounds from ^{131m}Xe, ^{129m}Xe, ^{133}Xe, ^{125}Xe and ^{125}I. As well as the pulse rate, many other algorithms also exhibit exponential behaviour, as the population of activation events with specific characteristics becomes less dominant. For example, the mean area of S1 and S2 pulses is pulled down by activation events and, as this population decays away, the mean pulse area gradually increases. These isotopes have short half-lives (the longest of which is 11.8 days for ^{131m}Xe, aside from ^{125}I which is removed by the getter) and high energies outside the WIMP search ROI. These populations were also continuously monitored to ensure that they were decaying at the expected rate. Therefore, this observed trend was not a concern for the WIMP search, and no additional data quality cut was applied to account for it.

3.9.2 Temporal Variation

As well as the overall exponential trend, some additional temporal modulation in the pulse rate can be seen in Figure 3.11. Similar behaviour is observed to various extents for many algorithms, including most rate plots, where it is most significant for S2 pulses, but is also subtly visible for SPEs. The behaviour is also evident in the mean pulse area of SEs and S2s (see Figure 3.12), as well as the mean pulse length. The algorithm limits here are flat, as the temporal modulation dominates the

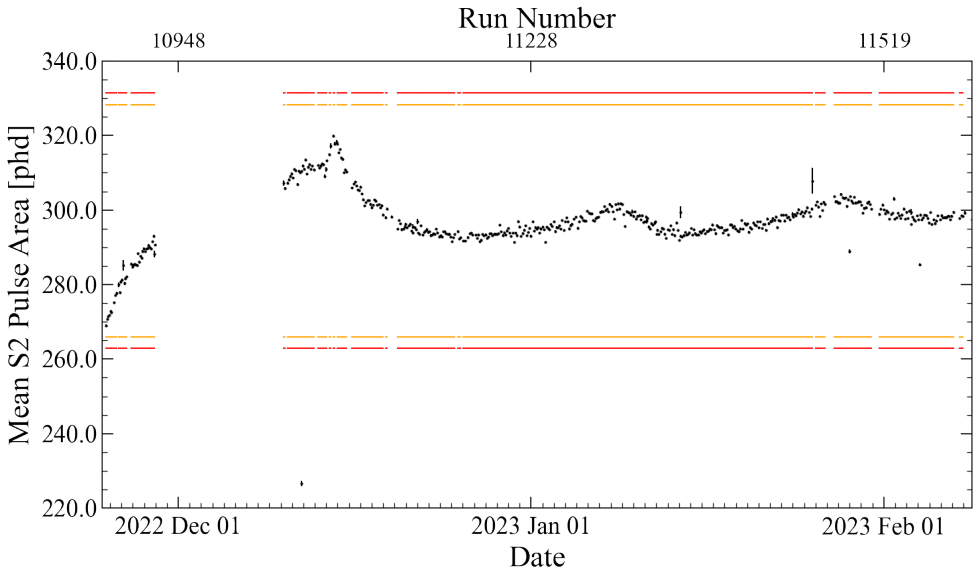


Figure 3.12: Algorithm history plot showing the mean S2 pulse area for pulses < 3000 phd as a function of time throughout SR2. Flat limits were chosen to contain the multiple distributions, and inner and outer limits are shown in yellow and red, respectively.

exponential trend. This was also found to be a position-dependent effect, with variation observed in the mean radial position of S2 pulses. Figure 3.13 shows the pulse length of SE pulses as a function of XY position, for runs at the peak and trough of the modulation of SE pulse length. At the minimum, pulses are narrower in the central region of the TPC, whereas at the maximum, they are more uniform in XY space. Algorithms monitor the mean SE pulse length in several regions in XY space, and the variation was found to be most significant in the central region, which is shown in Figure 3.14.

The cause of this behaviour is unknown, but one hypothesis is that the temperature decrease associated with a circulation change to a state with less mixing of the LXe resulted in a build up of condensation on the anode grid, causing sagging over time. The resulting smaller extraction region would produce a stronger extraction field and hence a higher extraction efficiency. This would lead to larger S2 pulses and a larger S2 pulse rate, deviating from the exponential trend. The increase in SPE rate could be a result of this, as S2 light is expected to produce SPEs through fluorescence of detector materials. A smaller extraction region would lead to narrower SE pulses, and the effects of grid sag would be most significant near the centre of the anode, explaining the position-dependent behaviour. The condensation could then fall under its own weight, before beginning to accumulate again. PREM was a useful tool for clearly demonstrating this trend. It was recommended that, if correc-

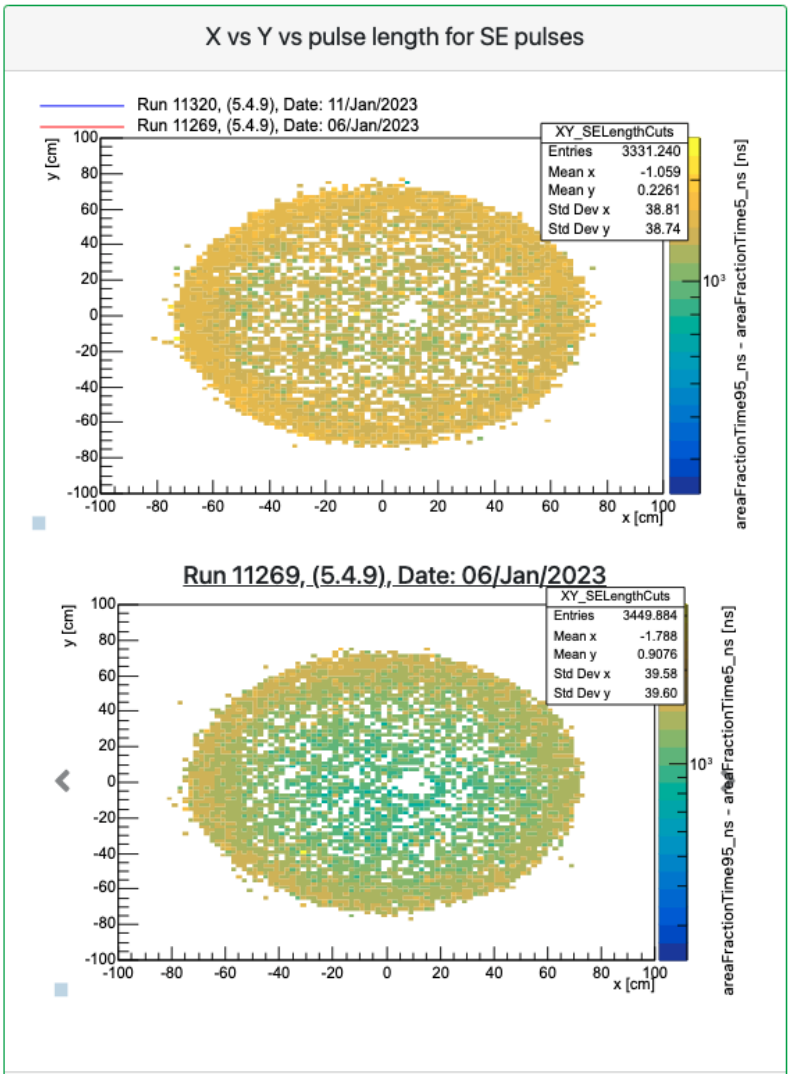


Figure 3.13: Comparison of two-dimensional histograms showing the length of SE pulses in XY space. Only events with total area < 20 electrons were selected, in order to avoid e-train backgrounds (see subsection 4.2.2). The upper panel shows run 11320 at the peak of the pulse length modulation, and the carousel below shows run 11269 at the trough of the modulation.

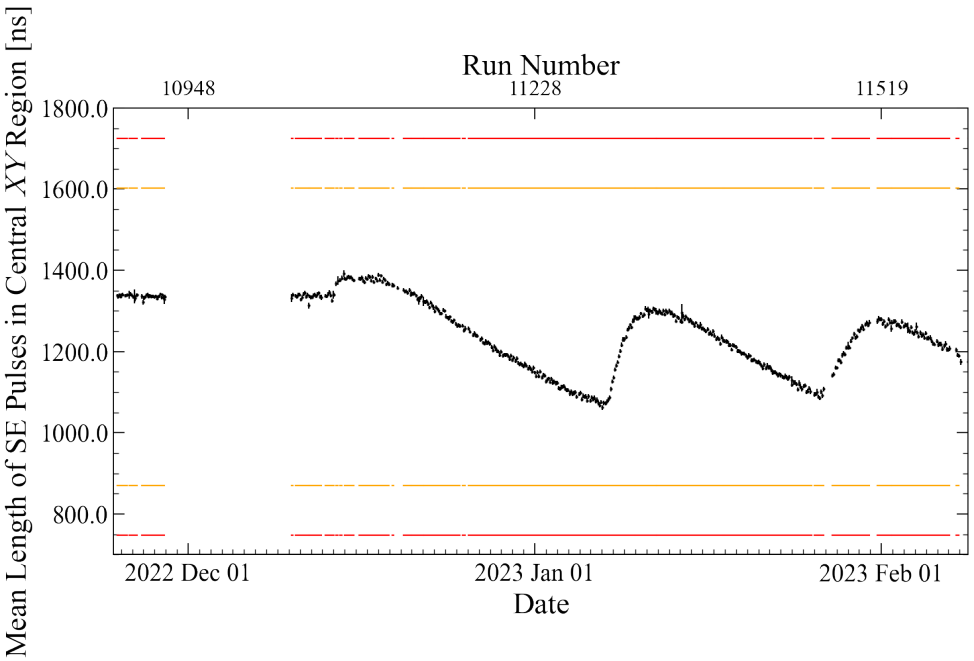


Figure 3.14: Algorithm history plot showing the mean length for SE pulses within a radius of 36.4 cm. Flat limits were chosen to contain the multiple distributions, and inner and outer limits are shown in yellow and red, respectively.

tions were implemented to account for this behaviour, these data would be suitable for the WIMP search. Although the SR2 acquisitions have not yet been used in a WIMP search analysis, S1 and S2 pulse sizes in SR3 data were corrected to account for the temporal variation observed there.

3.9.3 Low Event Counts

Following investigations of long-term trends, spikes in algorithm values from a single run or a small number of acquisitions were investigated. One reason for these deviations that did not result from a physical data quality defect was the low number of events in some runs, where fluctuations in data had a larger impact on algorithm quantities. This could be the result of a short run that was aborted early, but otherwise had no issues and was suitable for the WIMP search. In other cases, a large amount of deadtime was present throughout the run, which could be the result of missing data files, for example due to issues with data collection, transport or processing. The histogram in Figure 3.15 shows the total livetime accumulated over time bins of 60 seconds for a run with very low livetime, compared to a golden run. This plot was particularly useful for flagging runs with possible data-processing issues, providing information that can be shared with data-processing managers. In this case, upon

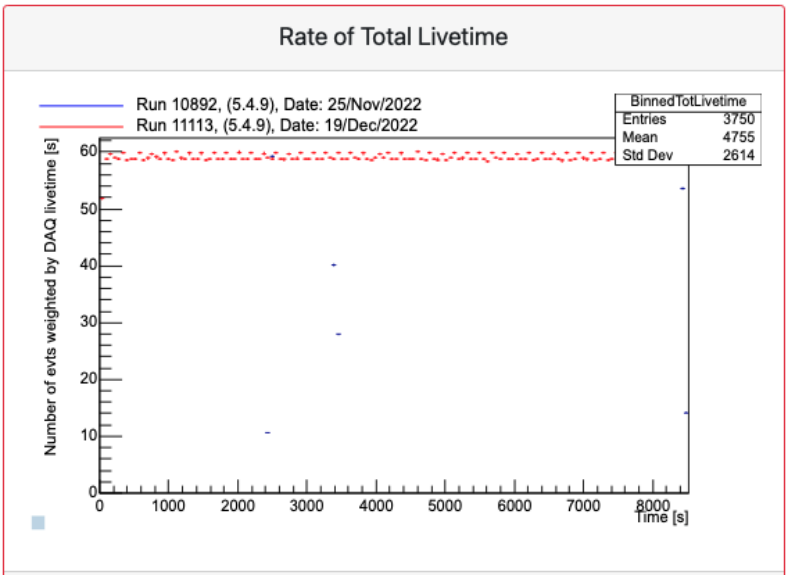


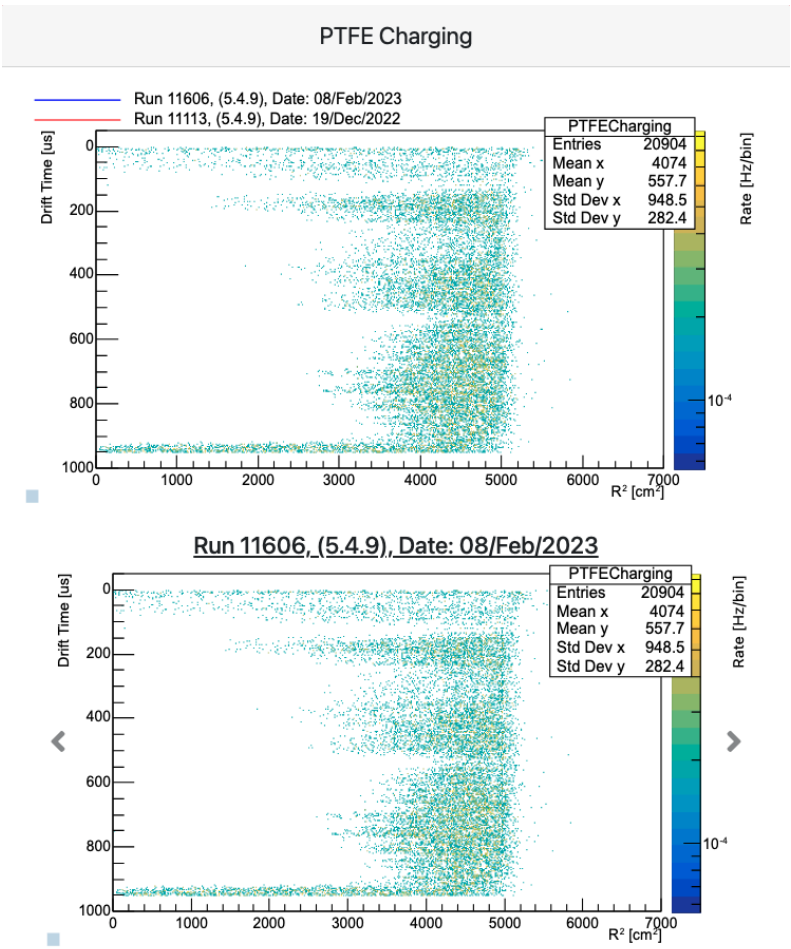
Figure 3.15: Total livetime accumulated in 60-second time bins, for run 10892, shown in blue, which is missing a large amount of livetime, and golden run 11113, shown in red.

further investigation it was found that only a fraction of the total events had been processed using LZap. This was corrected in a later reprocessing of the data.

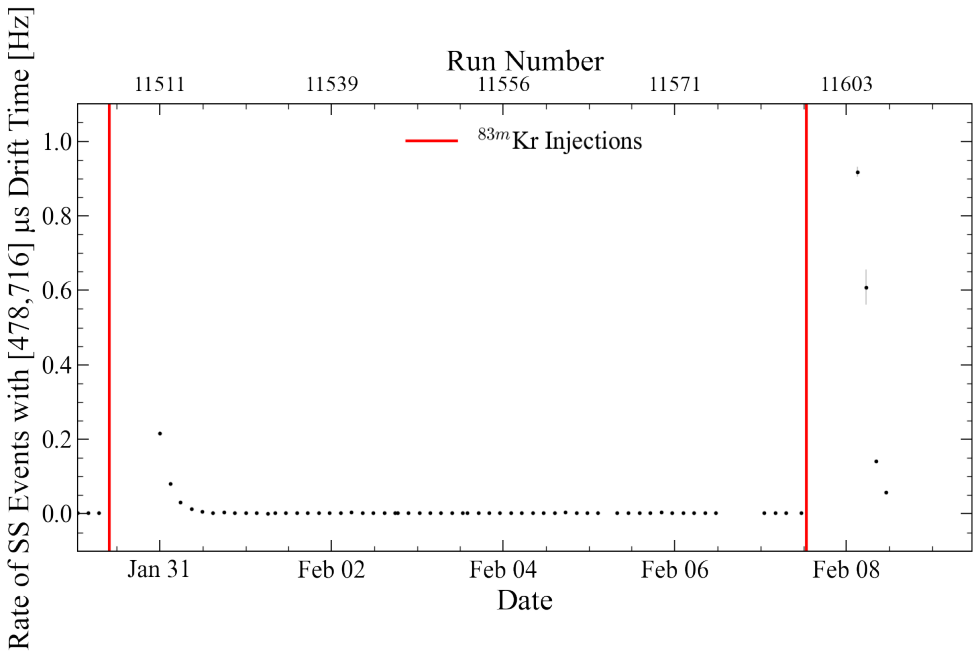
3.9.4 Residual ^{83m}Kr

Spikes with subsequent decays were observed in many algorithm history plots. One of the most clear examples of this is in an algorithm originally designed to monitor build-up of charge on the PTFE walls. The spatial distribution was plotted for Single Scatter events below the ER band, where the size of the S2 is reduced due to charge loss, with an S1 area between 80 and 500 phd. This typically leaves a population close to the walls. Algorithms monitor the rate in four horizontal slices. The "Wall-Rate3" algorithm shown in Figure 3.16 monitors the rate of these events with a drift time in the range [478,716] µs. The algorithm history plot is magnified to show two of the spikes and decays described. These rate increases occurred following two separate injections of a ^{83m}Kr calibration source into the LXe. ^{83m}Kr is a monoenergetic calibration source used for defining the detector response, as described in subsection 4.1.1. The histograms compare run 11606, at the beginning of the second spike, with golden run 11113. An excess of events near the wall can be observed, demonstrating how new xenon entering the TPC remains close to the wall in the more isolated circulation state, rather than fully mixing.

Referring to Equation 2.6.13, the population of events corresponding to the 41.5 keV monoenergetic ^{83m}Kr decay under SR2 detector condi-



(a)



(b)

Figure 3.16: The spatial distribution of Single Scatter events below the ER band is shown in (a) for run 11606 following a ^{83m}Kr injection (upper) and golden run 11113 (lower). The "WallRate3" algorithm history plot in (b) shows two spikes and decays in the rate of these events with a drift time of [478,716] µs, following two separate ^{83m}Kr calibrations.

tions can be clearly seen in S2-S1 space in Figure 3.17. ^{83m}Kr events have a double S1 structure, which is discussed in further detail in section 4.1, and this is observed through the double peak in the histogram of S1 pulse area in Figure 3.18.

PREM has therefore provided strong evidence for excess ^{83m}Kr following calibrations. ^{83m}Kr has a half-life of 1.83 hours, but the effects can be seen for up to 26 hours after larger injections. The energy of ^{83m}Kr is high relative to the expected WIMP signal, but would result in an increase in accidental-coincidence backgrounds inside the ROI. As data quality cuts are able to efficiently remove these backgrounds (see subsection 4.4.3), it was decided by analysis group leads that inclusion of these runs in the WIMP search would be preferable to the livetime loss incurred by their removal.

3.9.5 Hotspots

Localised increases in single-electron rate, known as hotspots, are one of the most significant data quality concerns. Hotspots can result from debris or defects in the wires on the gate grid, which can generate additional electrons. The TPC_Health module was used to identify hotspots by dividing the TPC into 16 segments in XY space and monitoring the SE pulse rate in each. The position of each of the segments is shown in Figure 3.19. Figure 3.20 shows an example algorithm monitoring the SE pulse rate in the region labelled 1a. Values outside the algorithm limits were examined in further detail by inspecting the XY distribution of SE pulses and Single Scatter events, as well as the SE pulse rate in the corresponding region throughout the run. The latter allowed the duration of the hotspot to be established.

Hotspots were classified into two categories. Transient hotspots generally only impact one time bin, so are less than 60 seconds in length. An example of a transient hotspot is shown in Figure 3.21a, where a localised increased rate of SE pulses is seen. The algorithm button for "Region4bRate" is red, indicating that this value is outside the limits. The corresponding rate spike can be seen in a single time bin in the plot of SE rate for region 4b shown in Figure 3.21b. A total of 50 transient hotspots were observed using PREM during SR2, some of which can be attributed to muons traversing the TPC, rather than grid emission.

Sustained hotspots result in increased rates across several time bins, and in some cases multiple runs. Figure 3.22a shows a run with two hotspots in regions 2a and 4a. The former was inspected using the rate plot shown in Figure 3.22b, which demonstrates that the rate continued

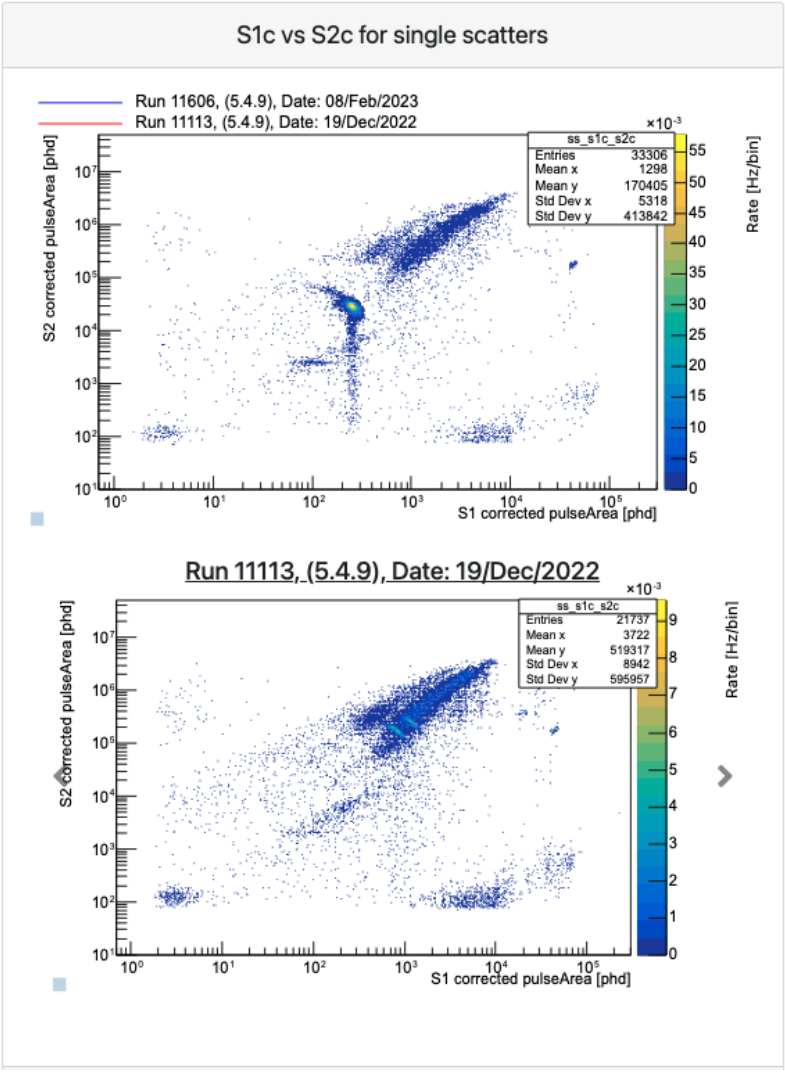


Figure 3.17: Single Scatter events in position-corrected S2-S1 space for a run following a ^{83m}Kr injection (upper) and a golden run (lower). Under SR2 detector conditions, the S1 and S2 signals from the monoenergetic ^{83m}Kr decay with 41.5 keV total energy correspond to the population clearly seen at coordinates of ∼(255, 12100) phd in the upper plot.

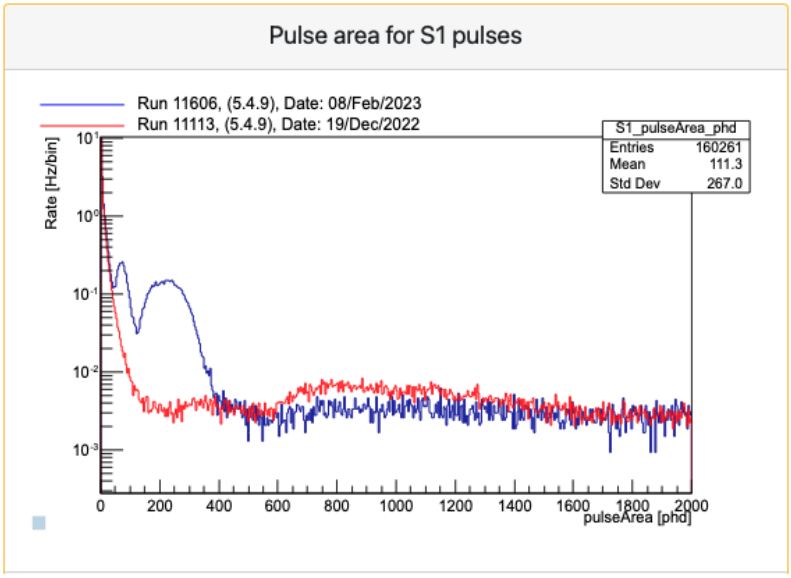


Figure 3.18: S1 pulse area for run 11606 following a ^{83m}Kr injection (blue) and golden run 11113 (red). The two peaks resulting from the ^{83m}Kr double S1 can be seen in blue.

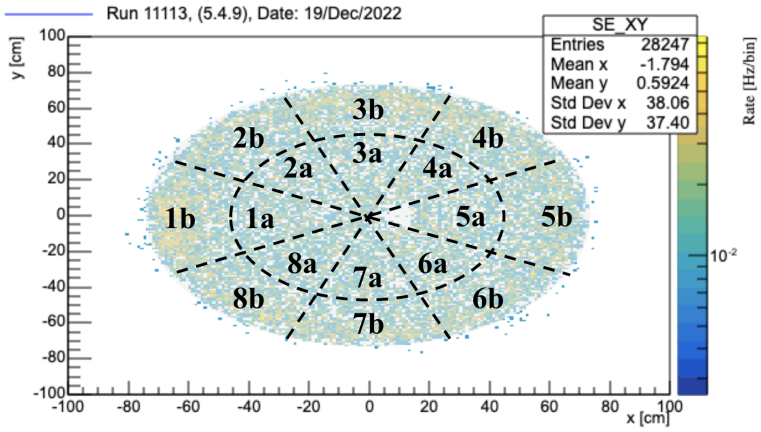


Figure 3.19: Diagram showing the 16 regions in XY space in which the SE pulse rate is monitored for hotspots, overlaid on a histogram of SE rate from golden run 11113.

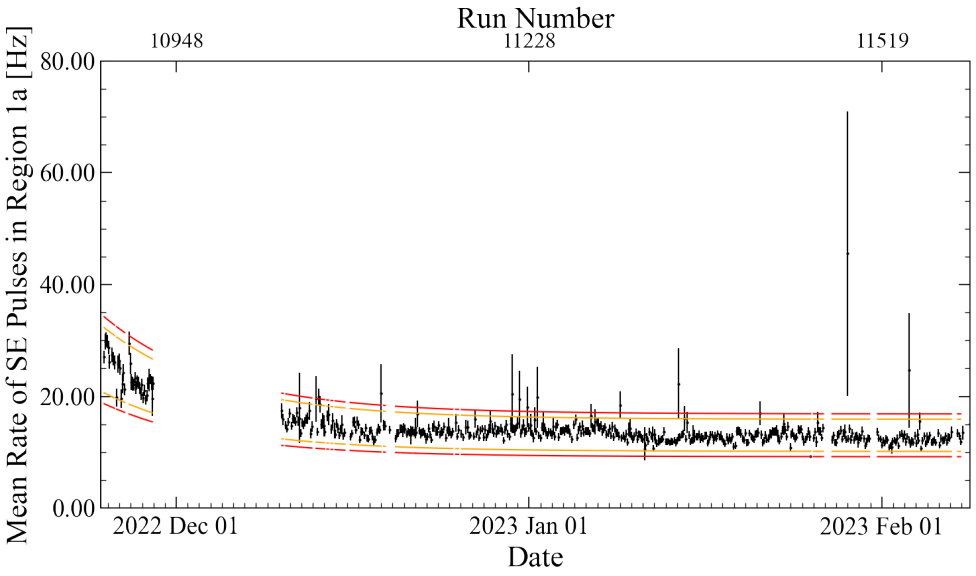


Figure 3.20: Algorithm history plot of the mean rate of SE pulses in region 1a, as defined in Figure 3.19, as a function of time. The exponential 3σ and 4σ limits are shown in yellow and red, respectively.

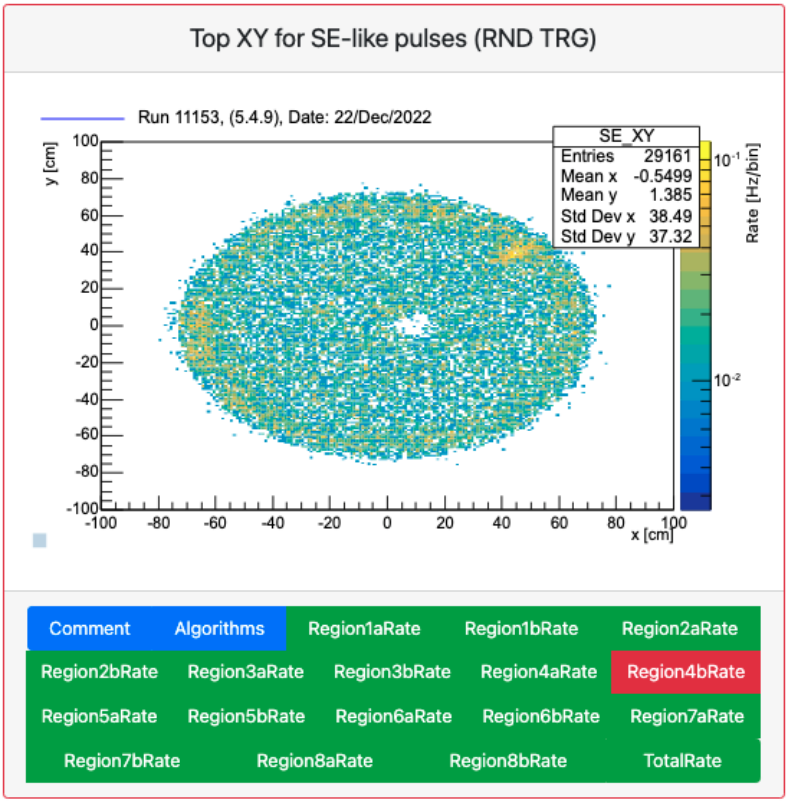
to increase throughout the second half of the run over a time period of more than an hour, suggesting the presence of contamination or a defect in the grid wires. Only 8 such events were observed using PREM during SR2.

Remote shifters observe LZ data from raw unprocessed files immediately as they become available using the Underground Performance Monitor (UPM). Of all SR2 WIMP search acquisitions, only three were marked as containing a hotspot by shifters that were not flagged during the PREM evaluation. Although increased SE activity was present, the mean rate remained within the 3σ limits. In future, the requirements for further inspection of a run could be loosened, or, ideally, the XY histograms would be inspected for every run.

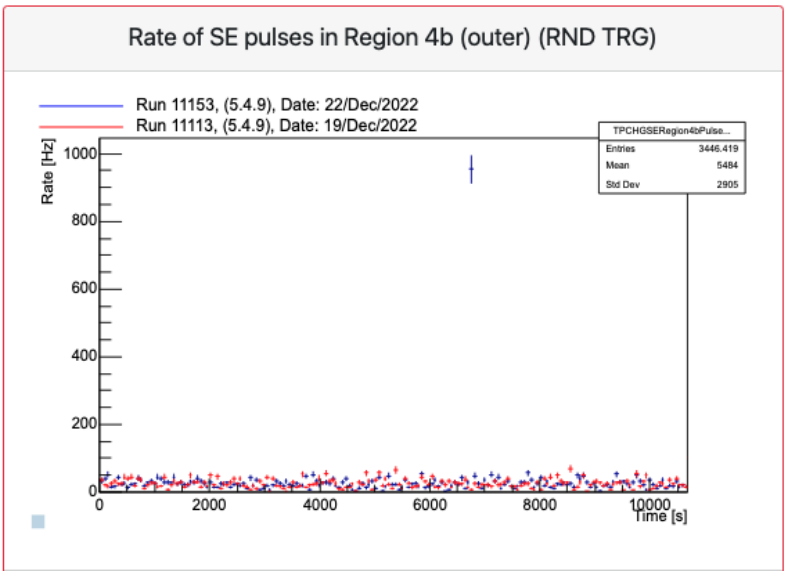
As data quality cuts, primarily the hotspot exclusion cut, are specifically designed to target these irregularities, the recommendation from the PREM team was that these runs would be suitable for the WIMP search following cuts. It was noted that the application of data quality cuts may result in the removal of entire runs in the case of sustained hotspots.

3.9.6 Skin Light Emission

The Skin_Health module was also able to identify data quality issues, the first of which was noise due to a grounding problem with several Skin PMTs. Figure 3.23 shows the number of pulses detected in each Skin PMT during run 11187. An excess is observed in channels 684-

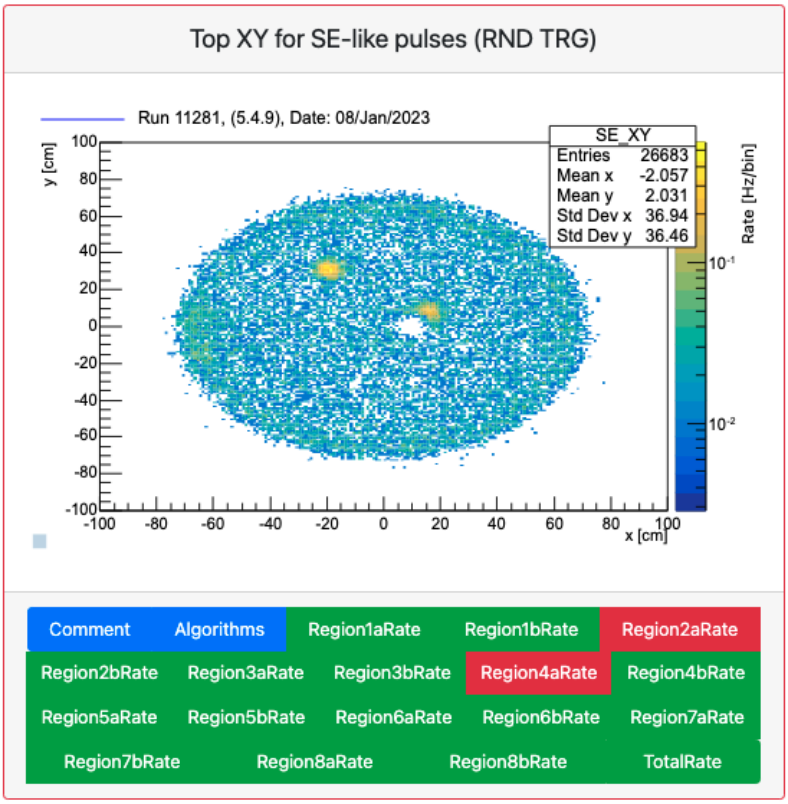


(a)

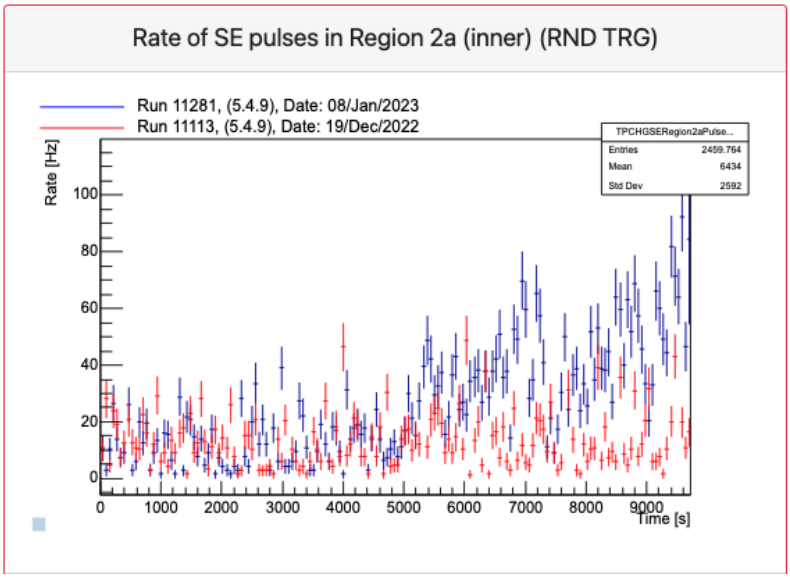


(b)

Figure 3.21: Example of a transient hotspot in run 11153, demonstrated by the SE pulse rate in XY space (a), and the SE pulse rate in region 4b throughout the run, compared to golden run 11113 (b).



(a)



(b)

Figure 3.22: Example of a sustained hotspot in run 11281, demonstrated by the SE pulse rate in XY space (a), and the SE pulse rate in region 2a throughout the run, compared to golden run 11113 (b).

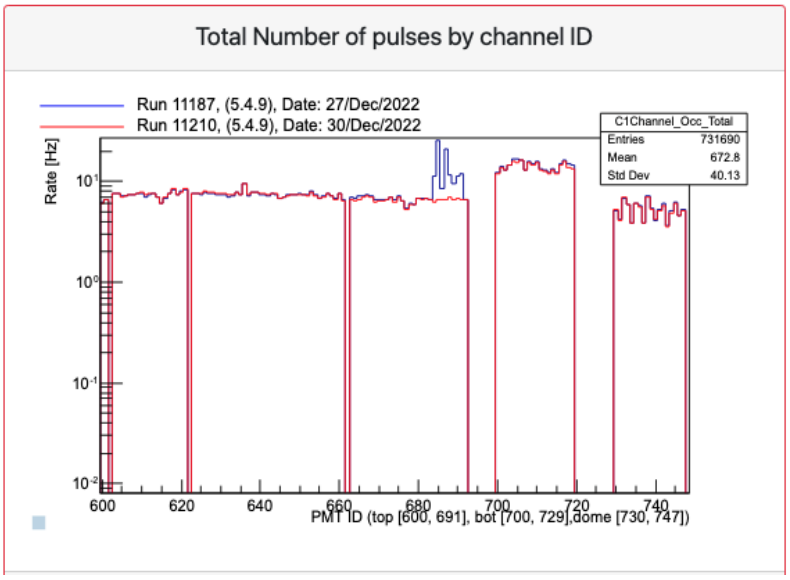


Figure 3.23: Number of pulses in each Skin PMT for run 11187 (blue), compared to reference run 11210 (red). This plot was taken from PREMweb and was originally produced using the Skin_Health module.

694 compared to the reference run. The algorithm shown in Figure 3.24 monitors the ratio of the number of pulses detected by Skin PMTs in the range 684-694, to the number of pulses seen by Skin PMTs in the dome region. This behaviour was seen to impact runs 11005-11206, after which the grounding nut was tightened to resolve the issue. Figure 3.25 shows the rate of pulses that result in activation of the Skin veto. As there is no significant difference between the affected runs and the reference run, it was confirmed that there would be no impact to the WIMP search, and the impacted runs were still suitable for this analysis.

A more significant issue was light emission observed in the Skin detector, primarily in channel 709 in the bottom side-skin. The heatmap in Figure 3.26 shows the reconstructed position of pulses in a run affected by light emission, compared to a golden run. The "theta" position where the pulse rate is highest corresponds to the region closest to the 9 o'clock position on the TPC heatmaps. While the light source itself was localised, the wider spatial distribution observed in the heatmap is a result of the position reconstruction, particularly for pulses with a low number of photons, which are more significantly impacted by statistical fluctuations. An algorithm was added to monitor the ratio of the pulse rate in this region (indicated by the overlaid box) to a typical region in the Skin. Several large spikes can be seen in the algorithm history plot (Figure 3.27), and each of these runs was investigated further.

For some runs, a transient spike was seen in the total Skin pulse rate and the rate of activated Skin vetoes, but no obvious leakage into

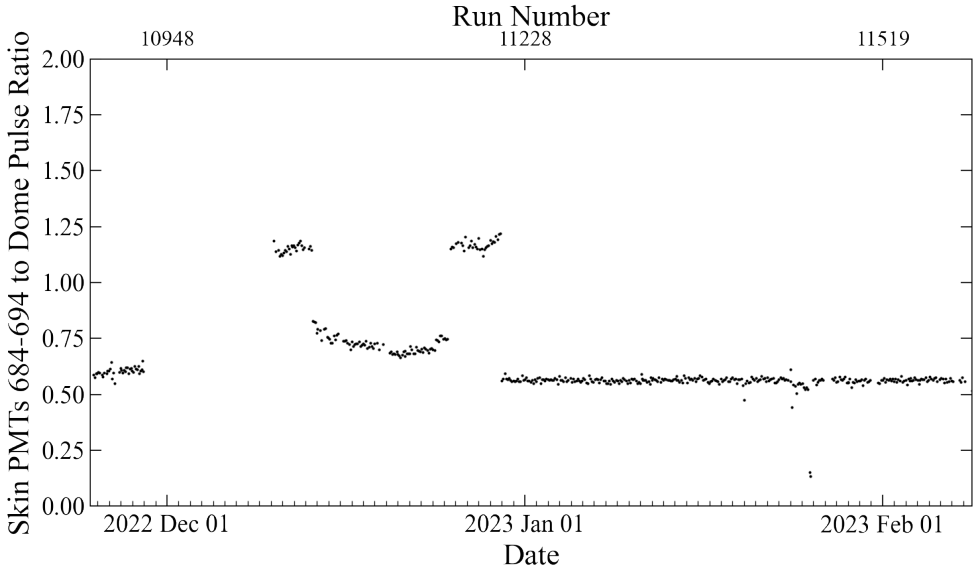


Figure 3.24: This algorithm history plot was taken from the Skin_Health module, and monitors the ratio of number of pulses in Skin PMTs 684-694, to number of pulses in Skin PMTs in the dome region.

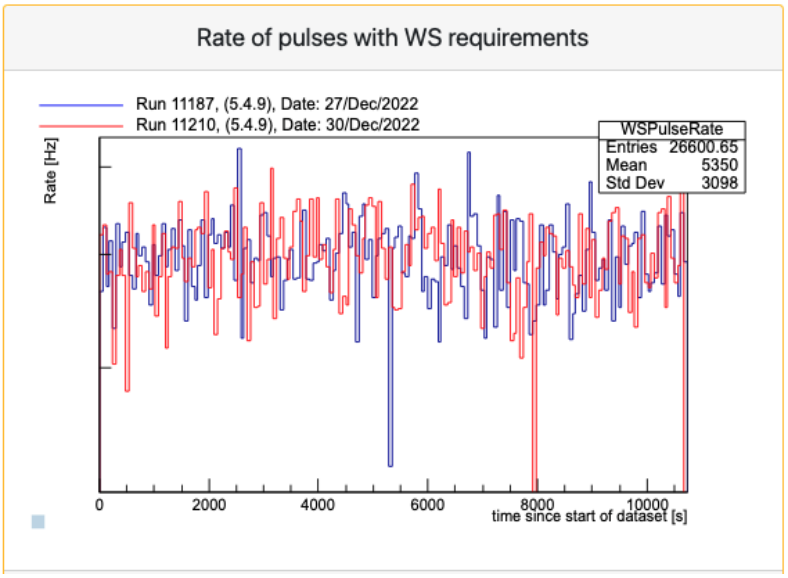


Figure 3.25: Rate of pulses resulting in activation of the Skin veto for run 11187 affected by grounding issues (blue) compared to reference run 11210 (red). This plot was taken from PREMweb and was originally produced using the Skin_Health module.

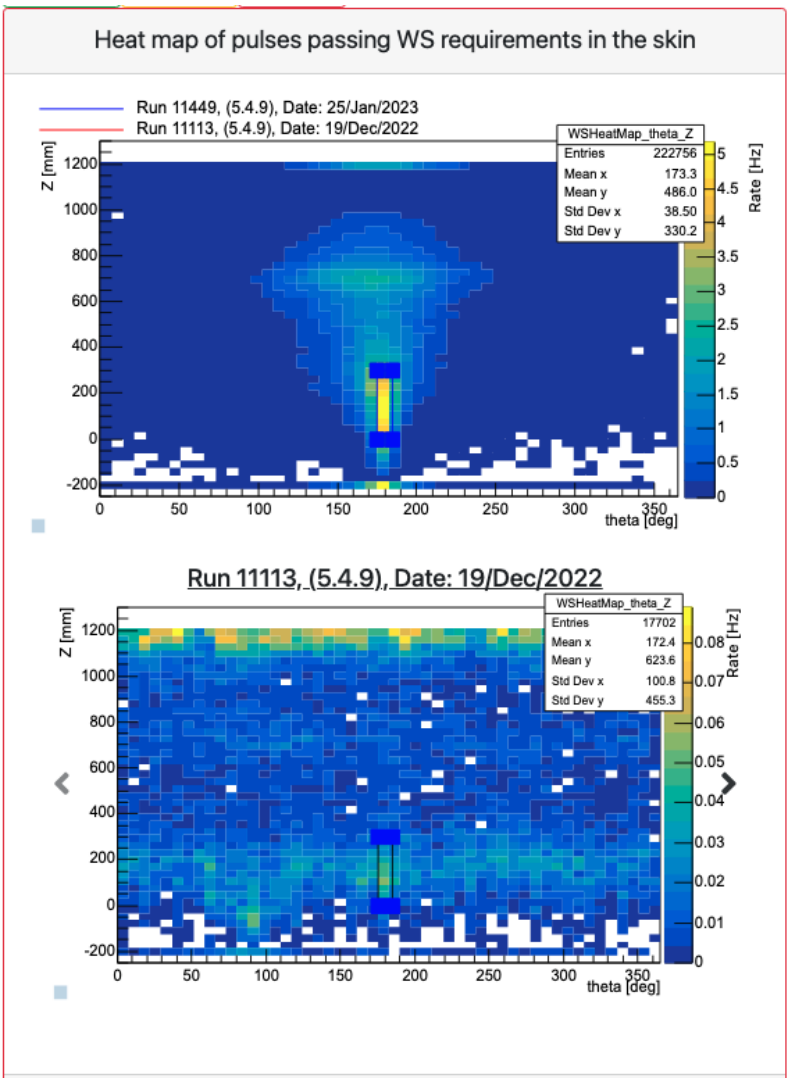


Figure 3.26: Heatmap showing the reconstructed positions of pulses activating the Skin veto, for run 11449 impacted by Skin light emission (upper) and golden run 11113 (lower). The region primarily affected by light emission is indicated by the overlaid box, with blue corner markers and black connecting lines. This plot was taken from PREMweb and was originally produced using the Skin_Health module.

the TPC was observed. It was recommended that the time period surrounding the spike be removed using data quality cuts, and that these runs would then be acceptable for use in the WIMP search. Other runs showed a sustained increase in the number of pulses activating the Skin veto, as shown in Figure 3.28, which would have an impact on the WIMP search. The TPC_Health module was used to observe light leakage into the TPC, as the distribution of SPE light was pulled towards the location of light emission in the Skin, as shown in Figure 3.29. It was therefore decided that these runs were unsuitable for use in the WIMP search.

The reason for the light emission is unknown, although the affected PMT, 709, is close to the RFR resistor chain and, during the Skin irregularities, shifts in the RFR current were observed. One possible expla-

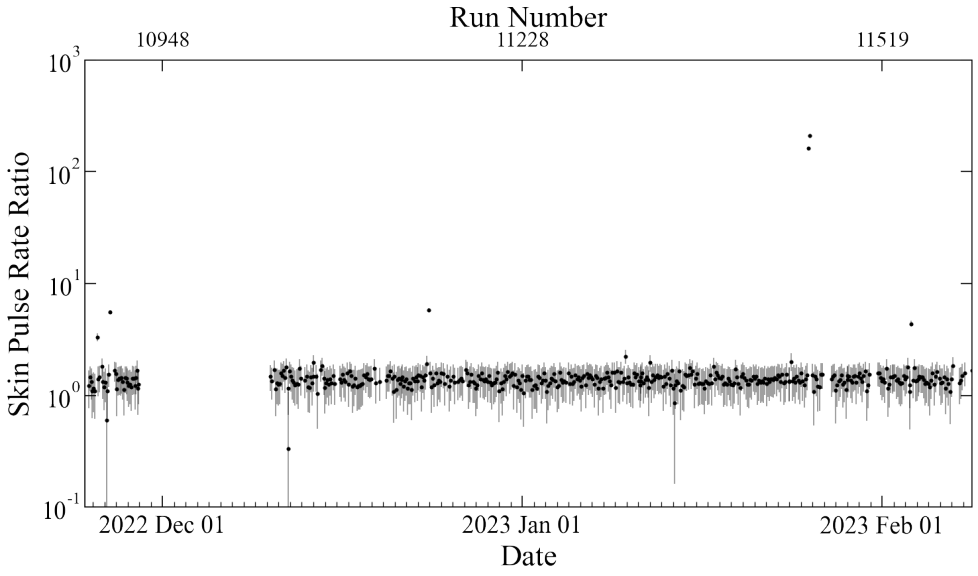


Figure 3.27: This algorithm history plot was taken from the Skin_Health module, and monitors the ratio of pulse rate in the Skin light emission region, as defined in Figure 3.26, to the pulse rate in a typical region in the Skin. Spikes required further inspection to determine if the corresponding runs were suitable for the WIMP search.

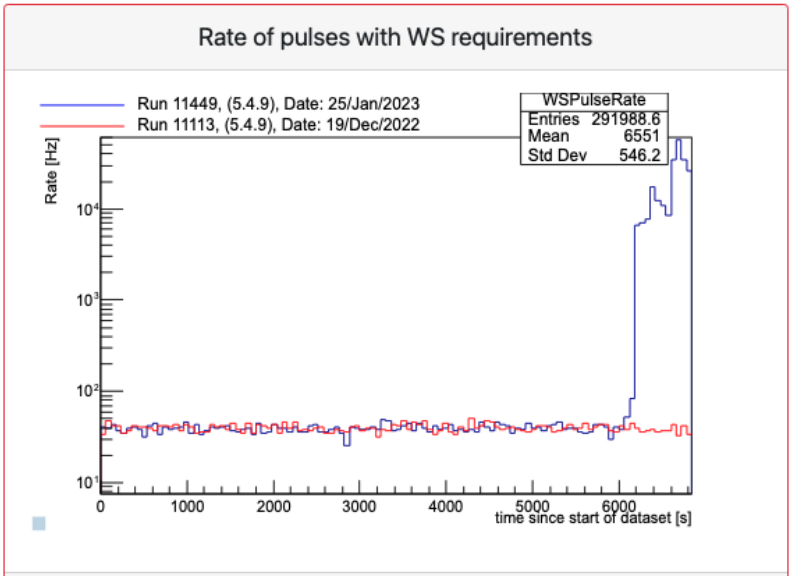


Figure 3.28: Rate of pulses resulting in activation of the Skin veto for run 11449 affected by Skin light emission (blue) compared to golden run 11113 (red). This plot was taken from PREMweb and was originally produced using the Skin_Health module.

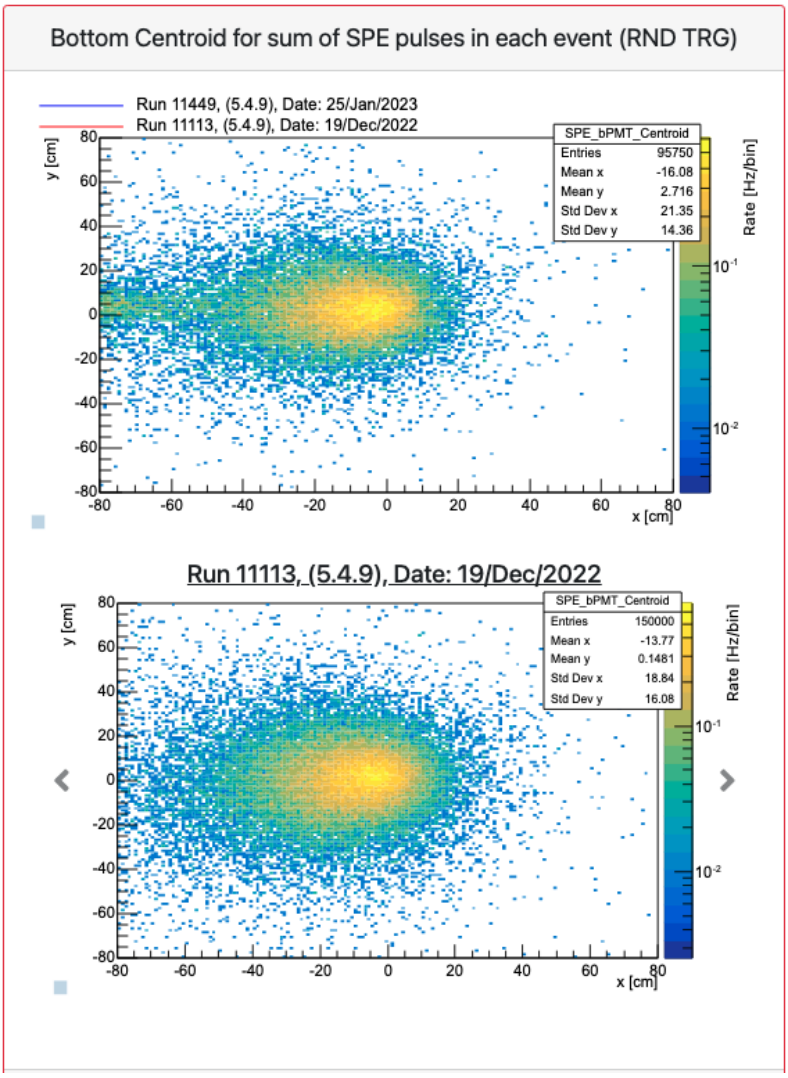


Figure 3.29: Histograms from the TPC_Health module showing the mean position of SPE pulses in the bottom PMT array for each event, for run 11449 affected by light emission (upper) compared to golden run 11113 (lower).

nation is therefore that if a microscopic crack was present in the resistor material, flowing current could cause light emission. The continued presence of light emission, at rates that could be damaging to PMTs, was a significant data quality issue and ultimately resulted in the termination of SR2. Detector conditions were altered as presented in Table 3.1, and the strength of the drift field was reduced, whilst maintaining the strength of the extraction field. The detector response under the new conditions was measured using calibration sources and a new science run, SR3, was initiated.

3.10 Summary

TPC_Health module plots, algorithms and shifter instructions were produced to monitor the health of the LZ detector using PREMweb tools. This played a vital role in the evaluation of data quality during SR1 and SR2, and these tools are now in place for SR3 and future science runs.

Both long-term trends and short irregularities were identified using PREM. Despite the early termination of SR2 due to unstable detector conditions, the recommendations of the PREM team were useful in determining the list of runs suitable for the WIMP search. The SR2 data have not yet been used for a WIMP search analysis, in part due to the complications arising from corrections required for temporal variation. However, it remains a possibility that SR2 runs identified by the PREM team be used for future analyses.

Data quality validation was conducted retroactively for SR1 and SR2, as data required manual processing with PREM. In future, PREM processing will be automated, so that PREM modules are run over all relevant LZap-processed runs by default. Development of this system is in progress, and will allow for offline data quality shifts during future science runs. A specific shifting module has been developed, with several plots sourced from the TPC_Health module, so that shifters will be able to review key histograms for every run and compare to references. This will enable a thorough contemporaneous evaluation of data quality, for example through improved investigations of hotspots.

Chapter 4

The SR1 WIMP Search

Data taking for the LZ Science Run 1 (SR1) WIMP search commenced on 23$^{\text{rd}}$ December 2021 and concluded on 18$^{\text{th}}$ April 2022. Following the removal of periods of detector instability, the total SR1 exposure was 60 livedays with a 5.5-tonne fiducial mass. The campaign produced world-leading limits for spin-independent WIMP-nucleon interactions, with cross-sections above $9.2 \times 10^{-48}\,\text{cm}^2$ rejected at a confidence level of 90% for a WIMP mass of 36 GeV [93]. The SR1 detector characterisation and analysis procedure will be described, with a particular emphasis on accidental-coincidence backgrounds and the data quality cuts designed to remove them.

4.1 Detector Characterisation Using Calibrations

An effective calibration campaign was essential in providing an understanding of the LZ detector response under SR1 conditions. This was required for the tuning of the ER and NR bands, determination of position-dependent light and charge collection efficiencies, modelling of backgrounds, and development of accurate simulations. The calibration strategy employed internal, external and source tube methods of deployment. A selection of sources which were particularly impactful for the WIMP search analysis are described.

4.1.1 ^{83m}Kr and ^{131m}Xe

During internal calibrations, sources are injected directly into the LXe using a carrier gas, leading to uniformity throughout the TPC, and overcoming the self-shielding properties of LXe. Internal calibrations include the use of the metastable monoenergetic sources ^{131m}Xe and ^{83m}Kr.

^{131m}Xe has a half-life of 11.8 days and emits a 164 keV γ ray, whereas ^{83m}Kr decays through two stages with a short half-life. Each ^{83m}Kr decay occurs via a combination of internal conversion, Auger electrons or X-rays, with a first transition of 32.1 keV and a second of 9.4 keV, which have half-lives of 1.83 hours and 154 ns, respectively [133].

The uniform spatial distributions and known energies of these sources are useful properties for defining the detector response. Light collection efficiency varies with position in the TPC, and depends on optical properties of detector materials. The size of S1 pulses must be corrected to account for this, so a corrected S1 area, S1c, is defined. This is derived through application of a corrections map, generated by normalising ^{83m}Kr S1 pulse sizes with respect to the S1 size at the detector centre. This methodology is described in further detail in section 5.2.

The sizes of S2 pulses are similarly corrected to S2c, but the drift time must now be separately taken into account. Drifting electrons can be captured by electronegative impurities in the LXe, leading to artificially reduced S2 sizes. The probability of capture increases with drift time, t, as $\exp(-t/\tau)$, where τ is the electron lifetime. This quantity describes the purity of the xenon and remained between 5000 µs and 8000 µs throughout SR1, significantly higher than the 951 µs maximum drift time. τ was determined using exponential fits of the S2 size for ^{131m}Xe events as a function of drift time. S2 XY corrections were applied in a similar manner to the S1 corrections, by normalising ^{131m}Xe S2 sizes to $(X = 0, Y = 0)$ to account for variations in grid deflections and the strength of the extraction field. Any monoenergetic source that is uniformly distributed in the detector is suitable for producing these corrections (or example α decays from ^{222}Rn that has emanated from detector components), and these were used for validation purposes.

Calibrations are also required for tuning of the NEST models used to predict the detector response to NR and ER interactions. NEST employs semi-empirical models that have been tuned using data from many experiments to reproduce microphysics processes, and maps energy depositions to S1 and S2 signals using functions of detector parameters. Both xenon light and charge yields and detector response parameters were tuned in NEST to reproduce the energies of ^{83m}Kr and ^{131m}Xe decays, as well as NR and ER band medians and widths that matched tritium and deuterium-deuterium calibration data. Chapter 5 presents the process in which these sources were used for the determination of the light and charge gain parameters.

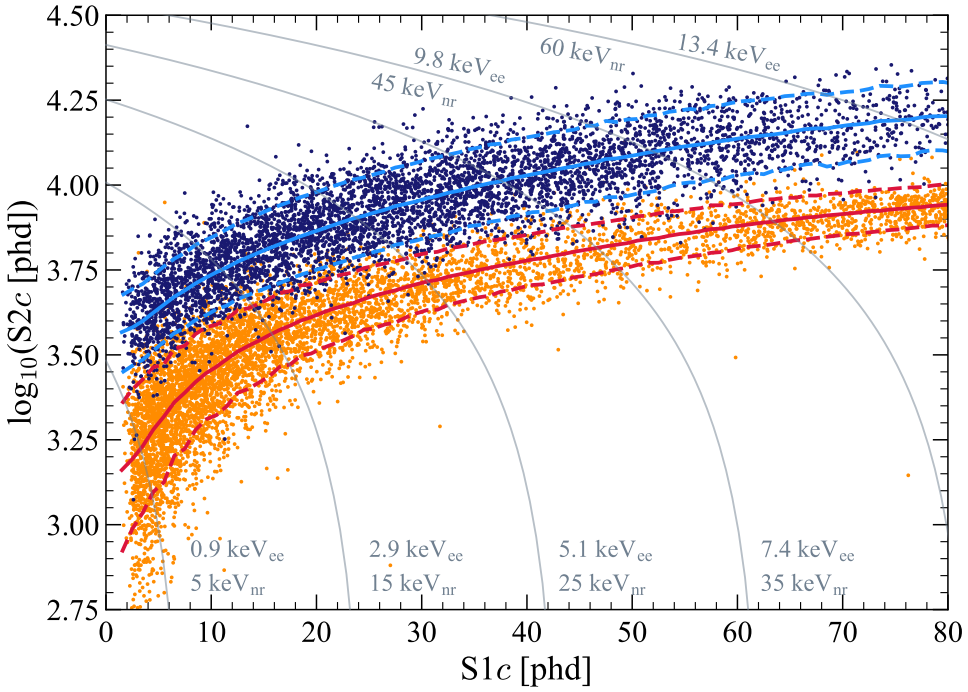


Figure 4.1: Distribution of both tritium (blue) and DD (orange) calibration events in $\log_{10}$ S2c-S1c space. Solid blue and red lines represent the medians of the ER and NR simulated distributions, respectively, while dashed lines show the 10% and 90% quantiles.

4.1.2 Tritium

Tritium is also introduced directly into the TPC in the form of tritiated methane. As a β source with an endpoint of 18.6 keV, tritium was deemed an ideal source to tune the median and width of the ER band using NEST. Tritium has a long half-life of 12.3 years, so is removed through a getter [134].

4.1.3 Deuterium-deuterium

In addition to other external neutron sources, a deuterium-deuterium (DD) generator produces monoenergetic neutrons with an energy of 2.45 MeV via DD fusion. The generator is located in the Davis Cavern outside the water tank. In DD direct mode, the neutrons travel unimpeded through air-filled pipes to the outer cryostat vessel, resulting in a continuum of recoil energies between 0 and 74 keV [135]. As the interaction length of 12 cm is small in relation to the size of the TPC, the generator is aimed near the top of the detector, so that neutrons may scatter a single time before exiting through the gas phase. Parameters of the ER model were propagated to the NEST NR model and were seen to be in good agreement with the DD data, which are shown alongside the tritium events in Figure 4.1, compared to the tuned NEST models.

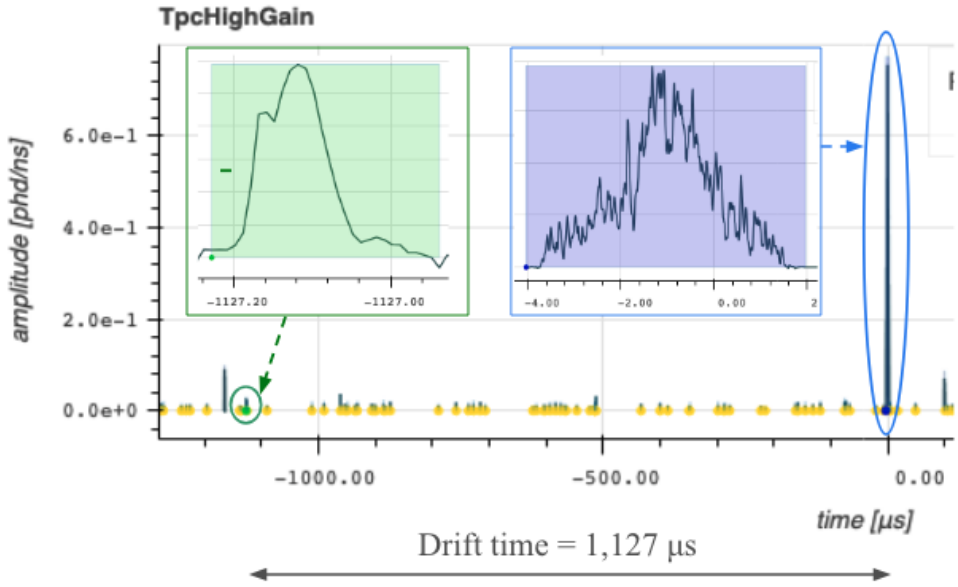


Figure 4.2: Example accidental-coincidence event from SR1 WIMP search data where the separation between S1 and S2 pulses is larger than the maximum drift time of 951 µs. Insets show magnified S1 and S2 waveforms.

4.1.4 AmLi

AmLi sources also produce neutrons, but with a continuum of energies up to approximately 1.5 MeV, and are deployed via three calibration tubes, situated between the inner and outer titanium cryostat vessels [136]. AmLi data were collected with all three sources at the same Z-position for heights of 0 cm, 70 cm and 140 cm above the cathode, following the end of SR1. The wider spatial distribution allows for superior cut acceptance evaluation in comparison to the DD source, as discussed in subsection 4.5.1, although it should be noted that the self-shielding of the LXe leads to a higher concentration of events at high radius.

4.2 Accidental Backgrounds

Understanding and mitigating backgrounds was a significant part of the WIMP search analysis. Accidental-coincidence events are the result of an isolated S1 pulse and an isolated S2 pulse occurring within the maximum possible drift time. An example event from the LZ Event Viewer is shown in Figure 4.2, where the separation between the S1 and S2 is unphysical, as it is larger than the 951 µs maximum drift time. This type of event can be erroneously classified as a Single Scatter. An understanding of the origins of these lone pulses allowed for the development of data quality cuts to efficiently remove accidental events.

4.2.1 Isolated S1s

Isolated S1s can refer to S1-like pulses either without an accompanying S2, or with an associated charge signal which cannot be collected. The latter can occur for events close to the TPC walls, as electrons can become attached to the PTFE resulting in charge loss, or events in the RFR, where electrons cannot be collected. Plate-out of radon daughters on the cathode and bottom grids also results in an S2 signal which is lost due to the RFR.

Some prominent sources of S1-like pulses without an associated charge signal will now be described. Dark counts occur when thermal fluctuations in PMTs promote an electron from the photocathode to the vacuum [137]. This results in pulses distributed uniformly in time that are SPE-like, despite not being instigated by a photon. The process is temperature-dependent, with a low PMT dark rate of ∼40 Hz in the TPC PMTs. Nevertheless, dark counts and SPE pulses can pile-up to mimic S1s meeting the three-fold coincidence requirement.

Cherenkov light can be produced in the quartz window of a PMT when ^{40}K contaminants decay. Light from this PMT can escape and be detected in other channels, or be coincident with, for example, dark counts in other PMTs. For Cherenkov S1s, most of the light is detected in a single PMT, a feature which can be exploited by data quality cuts, but this becomes difficult if the decay occurs in a dead PMT.

An increased SPE rate that exhibits a power-law decay over time following a large S2 is known as a photon train (ph-train), and is postulated to be due to fluorescence of the PTFE that lines the TPC, activated by VUV S2 light [138]. The rate of photons in the ph-train is proportional to the area of the S2 pulse. An increased rate of SPE pulses leads to a higher probability of pile-up into an S1-like pulse.

S1-like pulses with no accompanying S2 can also be a result of light leaking into the TPC, for example, that which was seen in the Skin and manifested as a change in the XY distribution of SPE pulses in the TPC (see subsection 3.9.6).

4.2.2 Isolated S2s

As with isolated S1s, a lone S2 can either have an associated S1 pulse that goes undetected, or can be the consequence of pile-up of electrons from various sources.

An example of the latter scenario, as described in chapter 3, is the emission of electrons from the high-voltage grids, that can occur due to

contamination or defects in the grid wires.

Similar to the ph-trains that can pile-up into isolated S1s, an elevated rate of electrons can also occur in the tail of a large S2. The process by which these "e-trains" are produced is thought to be the capture and release of drifting electrons by slow-moving electronegative impurities in the LXe [138]. This is supported by the increased rate of electrons for events with a high drift time, where there is a higher probability of capture on an impurity. The length of the e-train scales linearly with the size of the progenitor S2, and pile-up of these SE pulses is a significant source of lone S2s.

For some classes of S2-only events, the associated S1 is misclassified or lost. For example, events which occur near the liquid surface have a short drift time, which can result in the S1 being merged into, and classified as part of, the S2 pulse. As the electrons that comprise these S2s will experience very little diffusion, they can often be identified by their shorter, rectangular pulse shape.

Deposits in the gas region above the anode result in prompt scintillation and ionisation of xenon atoms in the gas. Due to the negative bias of the PMTs, the resulting electrons drift downwards. As the field is relatively low compared to below the anode, electroluminesence is only produced in a short burst close to the grid wire. The S2s will therefore be narrower than usual. If the S1 and S2 were correctly identified by the pulse finder, the event would be removed by fiducialisation. However, due to slower recombination in gaseous xenon [139], the S1 pulses can become stretched, leading to misclassification.

Due to the relatively poor collection efficiency of photons compared to electrons in the detector, a population of low-energy isolated S2 events exists below $\sim$ 4000 phd, where the S1 signal falls below the 3 phd threshold. This can be the case for radon daughters plating out on the cathode and gate grids, for example.

4.3 SR1 Core Cuts

A variety of data quality cuts were applied to remove poorly-reconstructed events and anomalous data. High rates of isolated S1s and S2s meant that many of the cuts were designed to specifically target accidental coincidences of these pulses. Cuts were developed on calibration data and sidebands outside the WIMP search ROI in order to mitigate bias, and were optimised to provide a high signal acceptance and background rejection efficiency. The cuts were categorised into four groups: S1-based

cuts, S2-based cuts, physics background cuts and livetime impact cuts, and are described in the following sections.

4.3.1 Livetime Impact Cuts

- **Hotspot Exclusions:** High rates of electrons due to grid emission occur in hotspots and can pile up to form S2s, causing accidental coincidences. To target these, a one-hour window was removed around periods of high S2 and SE pulse rate.

- **Muon holdoff:** The calculated muon flux of $(5.31\pm0.17)\times10^{-9}\,\mu s^{-1}cm^{-2}$ at the Davis Campus produces a rate of $\sim$ 10 muons per liveday passing through all three detectors. Time-coincident signals in the TPC, Skin and OD were used to tag these events, and a hold-off of 20 seconds was applied to remove periods of high-electron rates due to ionisation.

- **Electron/photon-train veto:** e-/ph-trains following large S2s can lead to pileup of SEs or SPEs resulting in classification as S2s and S1s, respectively. The area of the progenitor S2 was used to define the length of a veto window to remove these high-rates, resulting in a loss of livetime of 29.8% for SR1.

- **High S1 Rate Exclusions:** Periods surrounding large spikes in S1 rates, likely due to instrumental effects, were removed to reduce the number of isolated S1 pulses contributing to accidental coincidences.

- **Bad Buffer Cuts:** The DAQ buffer for a PMT channel (as described in section 3.2) can become full post-trigger, and additional data can no longer be written to file. This can result in artificially low pulse areas, so the cut ensured that TPC buffers were live until the end of the prominent S2, and Skin and OD buffers were live during the veto windows required to capture a pulse coincident with the S1.

- **Excess Area Cut:** This cut was designed to target periods of high rate not removed by the muon or e-train veto, for example when a large S2 occurred during a period of DAQ deadtime, resulting in an e-train leaking into the following event. Events with a larger amount of light then expected before and between the S1 and S2 pulses were removed.

- **Sustained rate cut:** Similarly, the sustained rate cut targeted prolonged periods of high SPE rate due to muons or S2s that were not tagged by the muon and e-train vetoes.

- **Burst Noise Cut:** Bursts of noise were observed in PMTs in all three detectors, but most significantly in the OD. These pulses can be tagged in the OD based on their shape, and a high density of these pulses led to the removal of an event.

4.3.2 Physics Background Cuts

- **Single Scatter:** WIMPs are expected to scatter only once in the TPC, so only events identified by the Interaction Finder as containing one prominent S1 and one prominent S2 were used in this analysis.

- **S1 and S2 threshold:** The WIMP search ROI, where the background model has been validated, was defined as $3 < \mathrm{S1}c < 80$ phd, $\mathrm{S2} > 600$ phd, $\log_{10} \mathrm{S2}c < 5$. This is in addition to the three-fold coincidence requirement for S1 classification.

- **Fiducial Volume (FV):** A drift time cut of [86, 936.5] µs, and a drift-time dependent radial cut of up to 5.2 cm from the wall, removed external backgrounds by exploiting the excellent self-shielding provided by LXe. The position reconstruction resolution is poorer for events at high radius due to reduced S2 light collection efficiency and charge loss at the walls. The radial cut was therefore chosen simultaneously with the S2 lower bound to prevent events leaking into the FV, an approach which required the raw uncorrected S2 pulse size. 6 cm circles around the resistor chains which span the height of the TPC were also removed, as these were found to be regions of high activity. The remaining fiducial mass was calculated to be 5.5 tonnes.

- **OD Veto:** A prompt veto targeted γ rays and proton recoils in the OD using a window of [-300,+300] ns around the TPC S1. A delayed veto removed neutrons that scattered in the TPC before being captured by the Gd-LS in the OD, by identifying events with an OD pulse greater than 200 keV within 1200 µs of the S1.

- **Skin Veto:** A prompt veto targeted particles such as cavern γs scattering in both the Skin and TPC, using a veto window of [-500,+500] ns around the S1. A delayed veto searched for γs that

were produced by neutron capture in the OD without depositing energy there. The delayed window was [+500 ns,+1200 µs], with an upper bound to match the OD veto.

4.3.3 S2-based Cuts

- **S2 width vs drift time:** Due to electron diffusion, a relationship exists between the width of an S2 pulse and the drift time. This can be exploited to identify accidental events where the drift time was incorrectly assigned.
- **Narrow S2:** Many accidentals are a product of S2s with very short drift times near the liquid surface, where the S1 is either lost or becomes merged with the S2 during pulse finding. To remove these events, a cut was placed on the RMS width of the S2 pulse with respect to its area.
- **S2 Rise Time:** S2 pulses near the liquid surface also tend to have flat tops, due to random bursts of electron emission. The rise time is defined as the difference between the time at which 1% of photons in the pulse have arrived at the PMT (Area Fraction Time 1 or AFT1) and the time at which 5% of photons have been detected (AFT5). By cutting on this with respect to the S2 pulse area, these events can be removed.
- **S2 early peak:** Accidental gas events with merged S1s and S2s can be targeted by making a cut on the FWHM of an S2 with respect to S2 pulse area, as shown in Figure 4.3.
- **S2 XY quality:** The radial fiducial cut relies on good position reconstruction, where the Mercury algorithm is used to determine the XY position of an event from the PMT hit pattern of the S2 pulse. This can be quantified using the goodness of fit of the Mercury algorithm.
- **S2 TBA:** S2-only gas events in the extraction region, especially those above the anode grid, have a higher top-bottom asymmetry, or TBA, than events from the LXe. Placing a cut on S2 TBA allowed this source of accidental events to be reduced.

4.3.4 S1-based Cuts

- **S1 Prominence Cut:** The Interaction Finder in LZap is tuned to select Single Scatter events, even if sample purity is reduced by

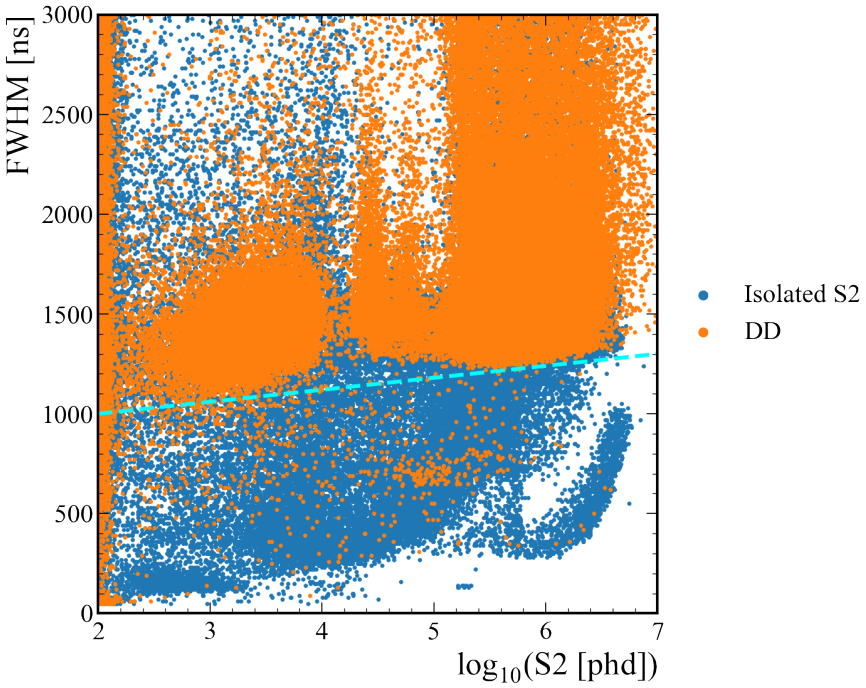


Figure 4.3: The S2 early peak cut removes events below the dashed line. It can be seen that isolated S2 pulses tend to have a smaller FWHM than S2 pulses from a DD calibration under SR1 conditions. Plot courtesy of S. Dey.

contamination from some non-SS. This cut targeted events with two prominent S1 pulses before the S2, due to an additional lone S1. This was achieved by comparing the size of the second largest S1 to the largest S1, and cutting in the parameter space shown in Figure 4.4. For an event to be removed, a minimum time separation of 750 ns between the S1s was required, in order to avoid cutting a single S1 which had been split in two by the LZap pulse-finding algorithm.

- **Stinger Event Cut:** Regions where the grid wires of the gate and anode are misaligned can result in electrons travelling up between the anode wires and entering a region with much lower fields than the extraction region. They can then fall onto the anode wires, where the field is high enough to produce electroluminesence. This results in a delay between the charge pulse and the S1 "stinger", so the cut removed events with a small S1 within 2 µs of an SE or an S2.

- **S1 TBA vs drift time:** This cut focuses on the relationship between S1 TBA and drift time, due to the geometry of the detector and reflectivity of the liquid surface. Events outside 3σ bounds defined by a fit performed on ^{83m}Kr calibration data were removed.

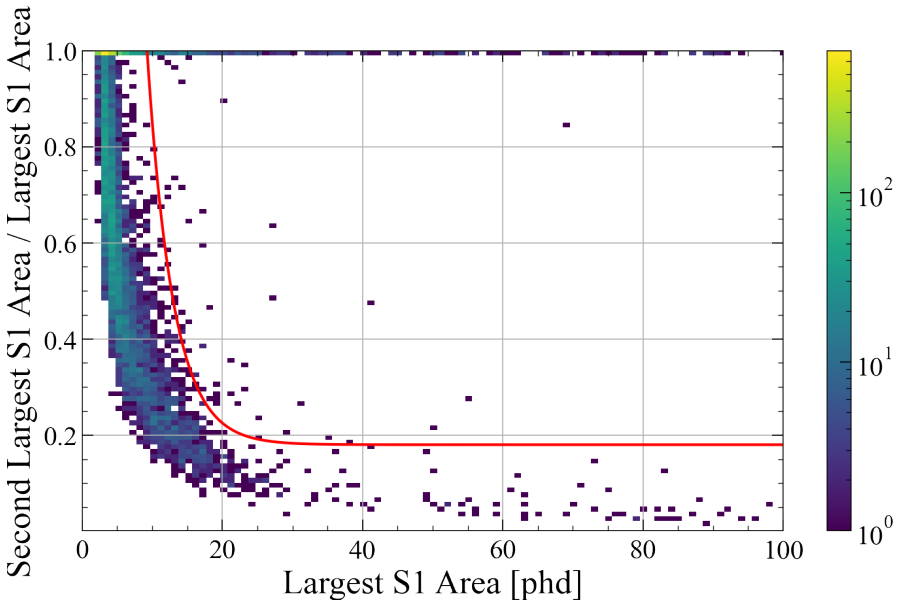


Figure 4.4: Distribution of SR1 WIMP search events in the parameter space which compares the ratio of areas of the two largest S1 pulses in an event to the area of the largest S1. Events to the upper right of the red line, which are likely to be contaminated by an additional lone S1, would be removed by the S1 Prominence cut. This plot was generated using code written by M. Williams.

- **S1 High Single Channel Cut:** Many accidental events have a higher fraction of light in one PMT than expected. An example of this is Cherenkov radiation emitted inside a PMT, where the majority of the light will be detected in that channel. The cut used the "maxChArea" variable to target this behaviour.
- **S1 Shape:** This cut used various metrics for the width, as well as the "prompt fraction time" (the fraction of the total pulse area that is detected within a certain time), to target isolated S1 pulses, which often have a longer and more distorted shape than real S1s.
- **S1 photon timing:** Isolated S1 pulses tend to have a larger spread than scintillation S1s in the times at which each PMT channel observes the maximum number of photons. By eliminating S1s with an RMS of channel peak times larger than 80 ns, this type of event can be removed.

4.4 Accidentals Model

4.4.1 Accidentals Datasets

Although the data quality cuts have been tuned to target both isolated S1s and S2s, the surviving accidentals are an important background and

must be modelled. Single Scatter events with a drift time larger than that allowed by the TPC (Unphysical Drift Time or UDT events) are known to be purely accidental coincidences. Assuming no correlation between S1 and S2 pulses, this dataset could be employed in the creation of an accidentals PDF. However, it was found that following the application of all data quality cuts, only one UDT event survived, so it was not possible to produce a reasonable distribution.

For the purpose of increasing statistics, a new dataset, known as AccidentalChopStitch (ACS), was generated by stitching together S1 and S2 pulses using the following process. Events classified as "Other", in which no good S1-S2 pair was identified by the Interaction Finder, were selected from the WIMP search data. As event-level RQs, such as the trigger times, of the stitched events would not be meaningful, the livetime cuts, and a selection of the physics cuts, were applied at this stage. Of these events, if the trigger was random and the area of the S1 was less than 115 phd, the event was selected as an input S1. If the event was S2-triggered and the S2 area was less than 1.02×10^5 phd, the event was selected as an input S2. The S1 and S2 pulses were then randomly paired, and the stitched events were reprocessed using LZap. A total of 28 million Single Scatter events were produced using this methodology.

4.4.2 Distribution

Statistical comparisons of the S1 and S2 area distributions confirmed reasonable agreement between the ACS and UDT datasets, and demonstrated that the S1s and S2s could indeed be treated as uncorrelated. 15,143 events with a physical drift time remained after the application of all data quality cuts to the ACS data, which was found to be insufficient to produce a smooth PDF. Therefore, the distribution was projected into the S1 and S2 dimensions, made possible by the independence of the S1 and S2 pulses. Kernel density estimation and linear fits were then employed to model the tails of the distributions, and the outer product of these was calculated. The final PDF is shown in Figure 4.5.

4.4.3 Rate Calculation

The PDF was then normalised to the expected number of accidental events in the WIMP search ROI following all data quality cuts. As the number of UDT events was too low to provide a precise event rate, this was calculated by applying the cut acceptance from the large ACS dataset to the initial number of UDT events. The initial counts for both

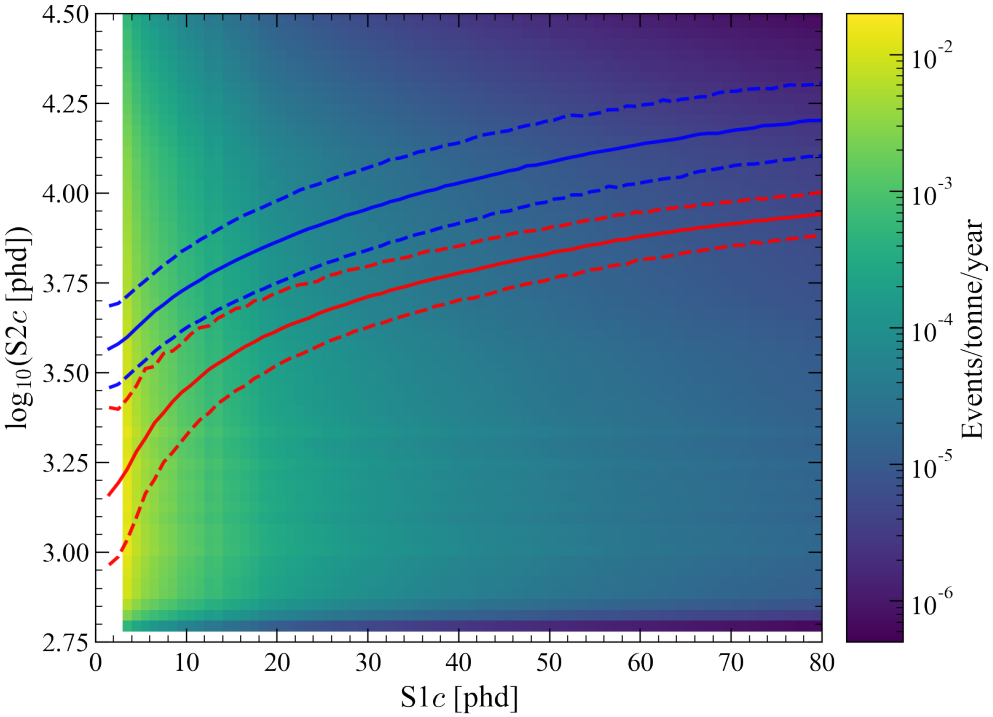


Figure 4.5: Distribution of AccidentalChopStitch events in $\log_{10}$ S2c-S1c space following all data quality cuts and a smoothing procedure, which separately modelled the tails of the S1 and S2 spectra and took the outer product. Solid blue and red lines represent the medians of the ER and NR bands, respectively, while dashed lines show the 10% and 90% quantiles.

the UDT and ACS populations were taken to be the number of events following a set of baseline cuts, including those that were applied to the ACS input and the vertical fiducial volume cut, but before the other cuts. The number of UDT and ACS events surviving both the baseline cuts and all data quality cuts are shown in Table 4.1. The ratio of these two values for the ACS dataset provided the cut acceptance, and applying this to the initial number of UDT events resulted in an expected 1.22 accidental events remaining in the WIMP search ROI, as shown in the final column. This value had an uncertainty of 21%, which includes contributions such as potential differences between the composition of the ACS and UDT datasets, and Poisson fluctuations in the measured number of UDT events. This corresponds to a rejection efficiency of 99.6% of ACS events.

4.5 SR1 Cut Acceptances

4.5.1 AmLi Cut Acceptance

It has been shown that the data quality cuts were effective for removal of accidentals, however it was important to ensure no significant loss in

Table 4.1: The number of events surviving baseline selection cuts and all data quality cuts for both the Unphysical Drift Time and AccidentalChopStitch populations. The final normalised event count of 1.22 was found by applying the ACS cut acceptance to the initial number of UDT events.

	UDT	ACS	ACS Normalised by UDT Rate
Baseline Cuts	310	3,861,123	310
All Cuts	1	15,143	1.22

acceptance of NR events such as WIMPs. An investigation of cut acceptance was carried out by the author, in collaboration with the UCL group, with pure AmLi as the selected source, due to its presence throughout the FV and coverage of the full S2 range.

In order to confirm that selection of AmLi neutrons was successful, it was necessary to compare the recoil energy spectrum from AmLi data to LZLAMA simulations (see chapter 5). The following cuts were applied to all three AmLi sources at the $Z = 70$ cm position:

- SR1 core cuts:
  - Single Scatter
  - Fiducial volume
  - ROI
- OD coincidence cuts:
  - OD pulse area ≥ 200 phd (data) and ≥ 100 phd (simulations), based on spectra of area of the largest pulse in the OD
  - OD coincidence ≥ 5 (data only)
  - S1 start time $<$ OD start time $<$ S1 start time + 150 µs

The simulated spectrum was then scaled by

$$\frac{\text{Livetime from data} \times \text{Activity of source}}{\text{Number of simulated events} \times \text{Percentage of neutron events simulated}} \tag{4.5.1}$$

It was also necessary to adjust for the different activities of the three sources, which were measured at the University of Alabama using a ^{3}He neutron detector [136], so simulations were run separately for each calibration tube, before scaling and summing.

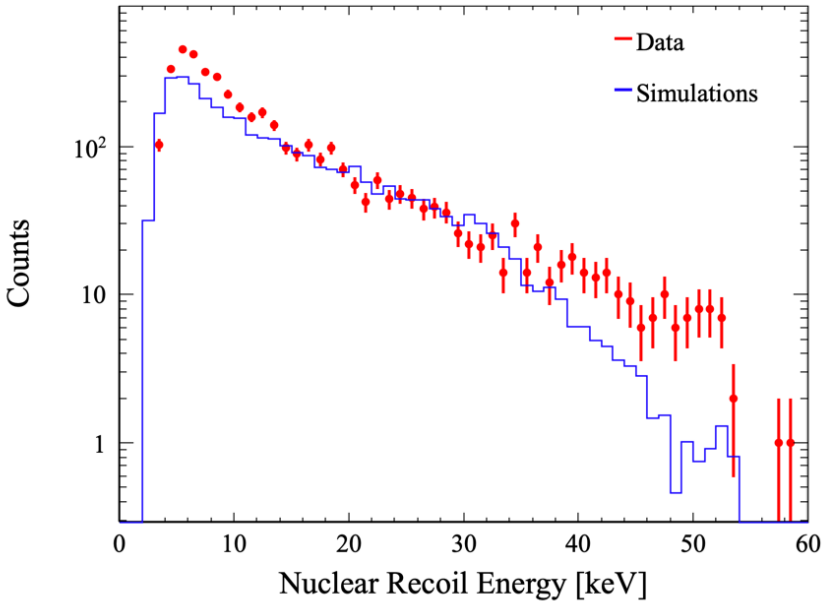


Figure 4.6: Spectrum of recoil energies for three AmLi sources at $Z = 70$ cm for both simulations and data.

The resulting energy spectra are shown in Figure 4.6. Some discrepancies were seen between the data and simulations, particularly at high and low recoil energies. Quantitatively, using the two sample Kolmogorov-Smirnov test [140], the null hypothesis that the two datasets were sampled from populations with identical distributions is rejected at the level 4.5×10^5, meaning that the distributions are highly incompatible. It was thought that the excess of low energy events in the data was due to the misclassification of Multiple Scatter events as Single Scatters by the Interaction Finder. Better agreement was seen between the data and simulations for the combined SS and MS spectra. The disagreement at high recoil energies was likely due to (α,n) reactions from oxygen contamination in the AmLi source not accounted for in the SR1 simulations. For example, the lithium target could become contaminated due to a leak in the capsule seal. This could result in the presence of ^{18}O, which would produce high-energy neutron tails beyond the usual 1.5 MeV endpoint [141]. The effect of the oxygen contamination was later estimated using Geant4-based tools, and an AmLi spectrum which accounted for this was found to be in better agreement with the observed data.

The distribution of AmLi events from calibration data taken with sources at all three heights is shown in Figure 4.7. Of the SR1 Core Cuts discussed, all physics background cuts were either already applied as baseline cuts to select the AmLi events, or are OD and Skin veto cuts that are ordinarily used to identify and reject neutrons. It was found

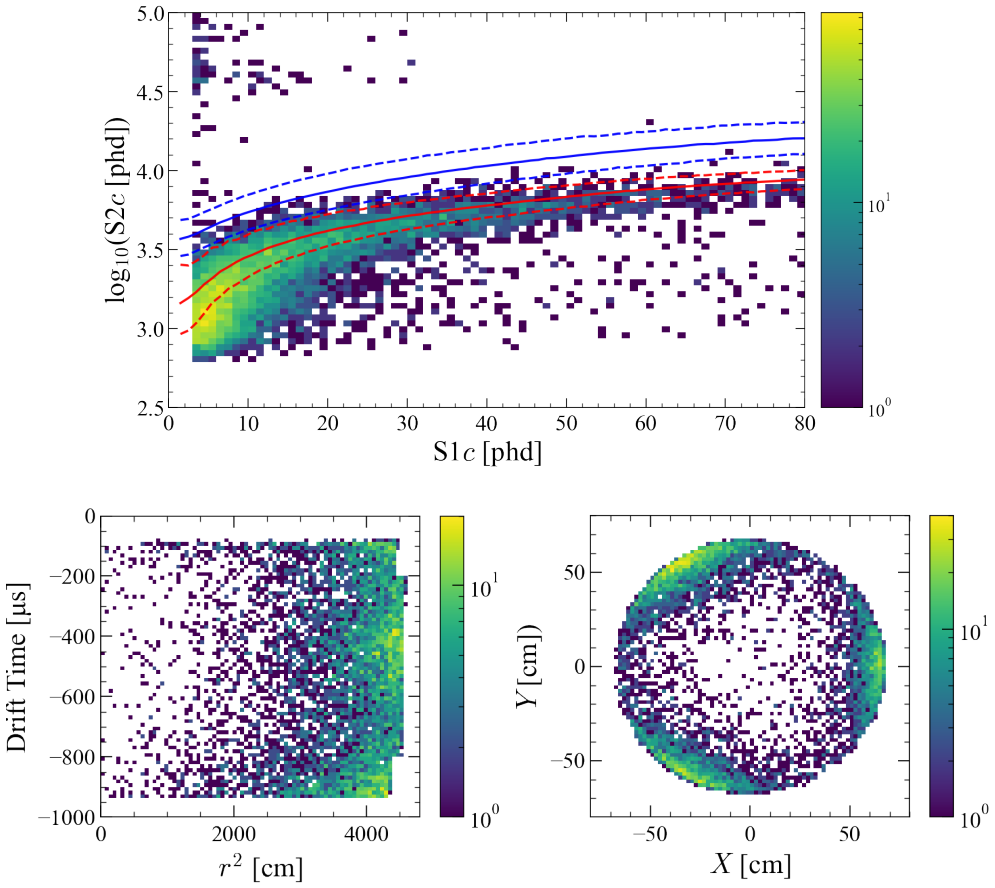


Figure 4.7: Distribution of events in $\log_{10}$ S2c-S1c, drift time-r^2 and XY spaces following AmLi selection cuts, combining datasets with all three sources at Z = 0 cm, 70 cm and 140 cm. Solid blue and red lines represent the medians of the ER and NR bands, respectively, while dashed lines show the 10% and 90% quantiles.

that application of all livetime impact cuts removed 99.7% of events. This was primarily due to the sustained rate cut and e-/ph-train veto, because of the high rate of AmLi events across the entire energy range. As a result of these high rates, the waveform environment was unclean, with pile-up of small pulses, including those from e-/ph-trains, leading to contamination of these events by accidentals. The AmLi acceptance curves were nevertheless useful for validation purposes.

Since the livetime impact cuts were not applicable for this dataset, the acceptance was defined here as

$$\frac{\text{AmLi events surviving baseline cuts and selected cut}}{\text{AmLi events surviving baseline cuts}}. \tag{4.5.2}$$

Acceptance curves were produced for each S1-based cut as a function of S1c (Figure 4.8) and each S2-based cut as a function of $\log_{10}$ S2c (Figure 4.9) individually, as well as for the combined S1-based and combined

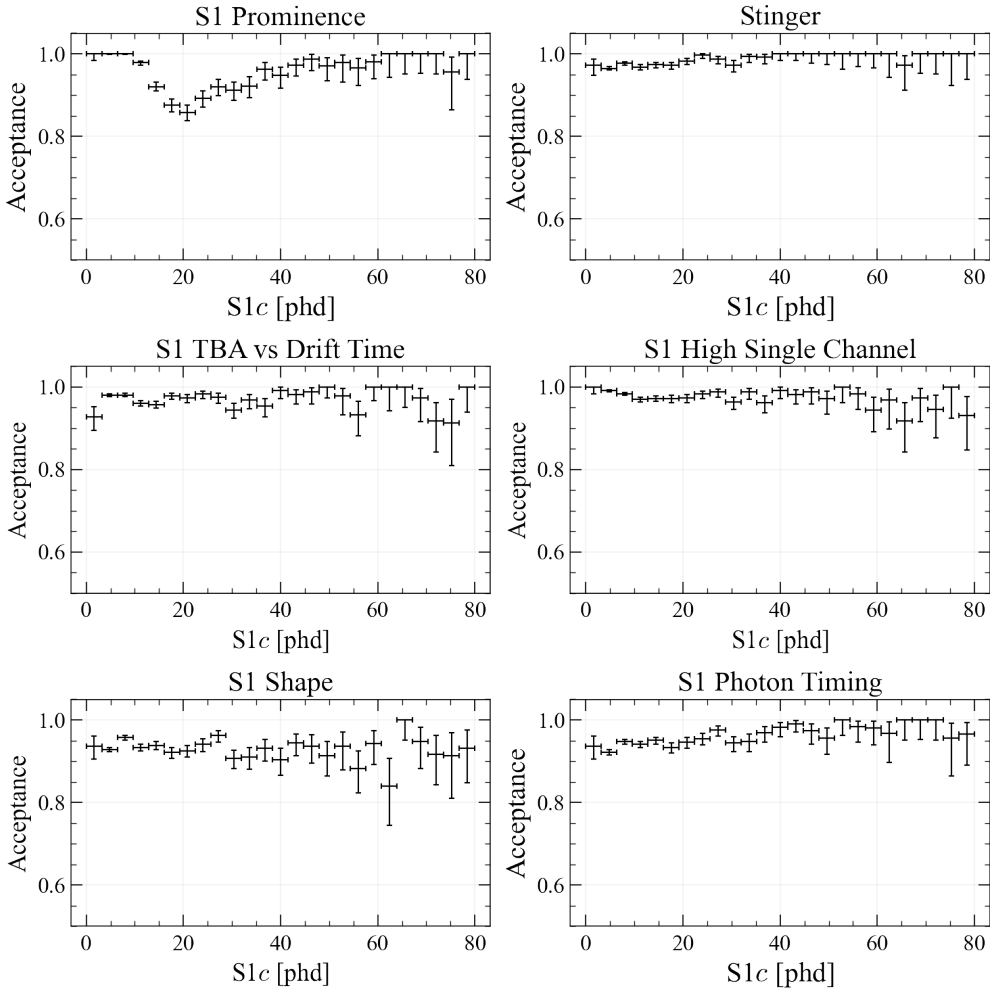


Figure 4.8: Acceptance of AmLi events as a function of S1c for each of the S1-based data quality cuts.

S2-based cuts in Figs. 4.10 and 4.11. Here, acceptance as a function of r^2 and drift time is also shown, as well as distributions of rejected events in $\log_{10}$ S2c-S1c, drift time-r^2 and XY spaces. The error bars were determined using the Clopper-Pearson method, which has a high coverage level and is commonly used for calculating binomial confidence intervals [142].

Many of the cut acceptances were lower than desired for the WIMP search, but due to the increased accidentals component resulting from high AmLi rates, this was expected. Acceptance was generally lower at small radial distances from the detector centre. Events in this region were far from the source tubes, which are situated outside the ICV, so statistics were low. The S1 Prominence cut showed additional interesting behaviour, as a minimum acceptance was observed for an S1c of ~ 20 phd. This can be understood by observing the cut boundary in Figure 4.4. The cut does not remove events where the area of the largest S1 pulse is small. The turning point of the function used as the cut boundary is close to S1c = 20 phd, so the cut is more sensitive to deviations in this

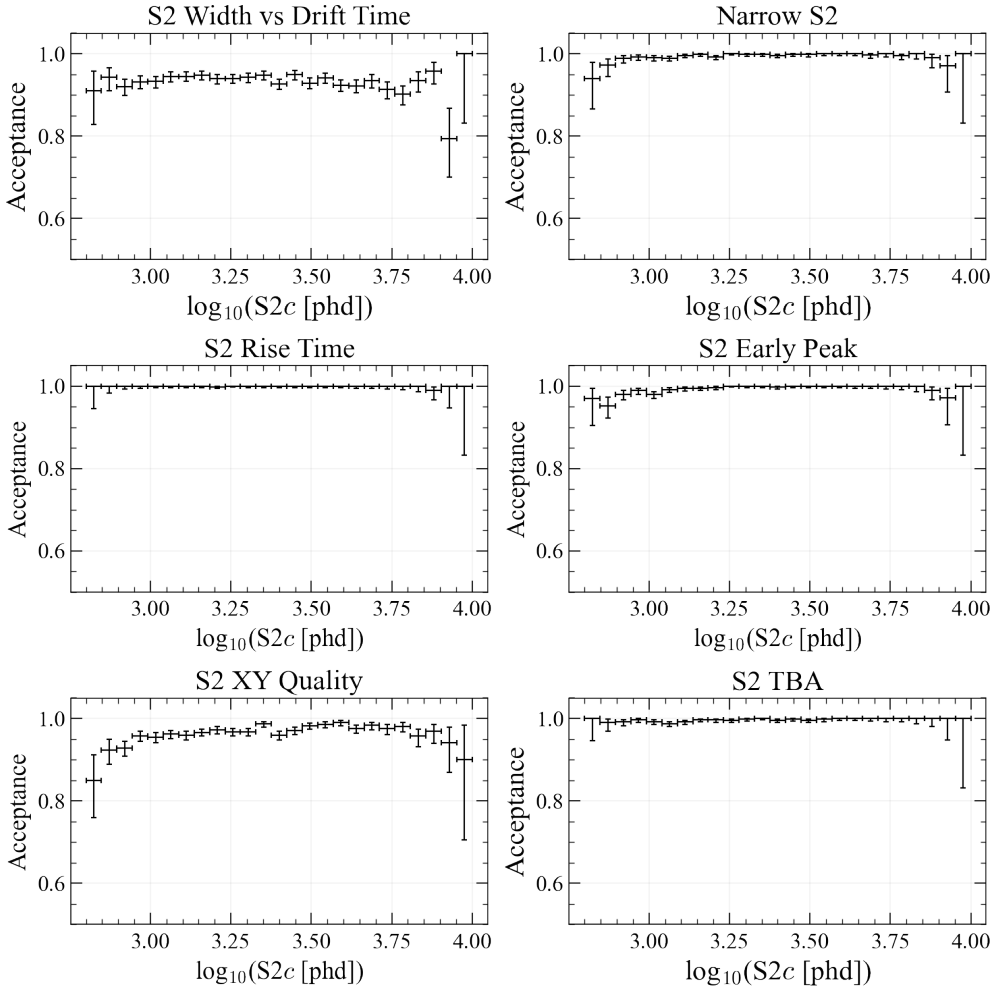


Figure 4.9: Acceptance of AmLi events as a function of $\log_{10}$ S2c for each of the S2-based data quality cuts.

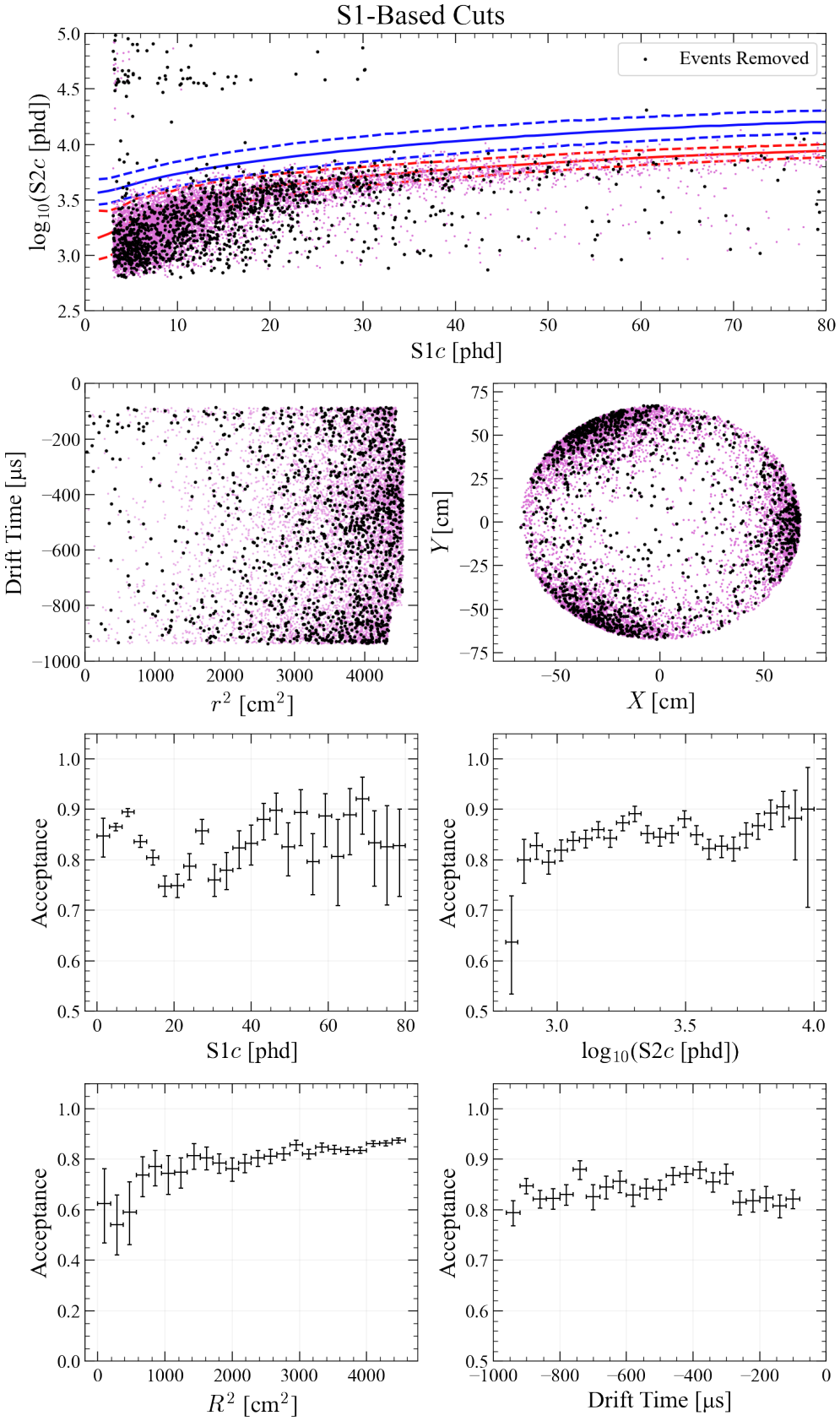


Figure 4.10: Distributions of AmLi events removed by (black) and surviving (pink) all S1-based data quality cuts, as well acceptance as a function of S1c, $\log_{10}$ S2c, r^2 and drift time. Solid blue and red lines represent the medians of the ER and NR bands, respectively, while dashed lines show the 10% and 90% quantiles.

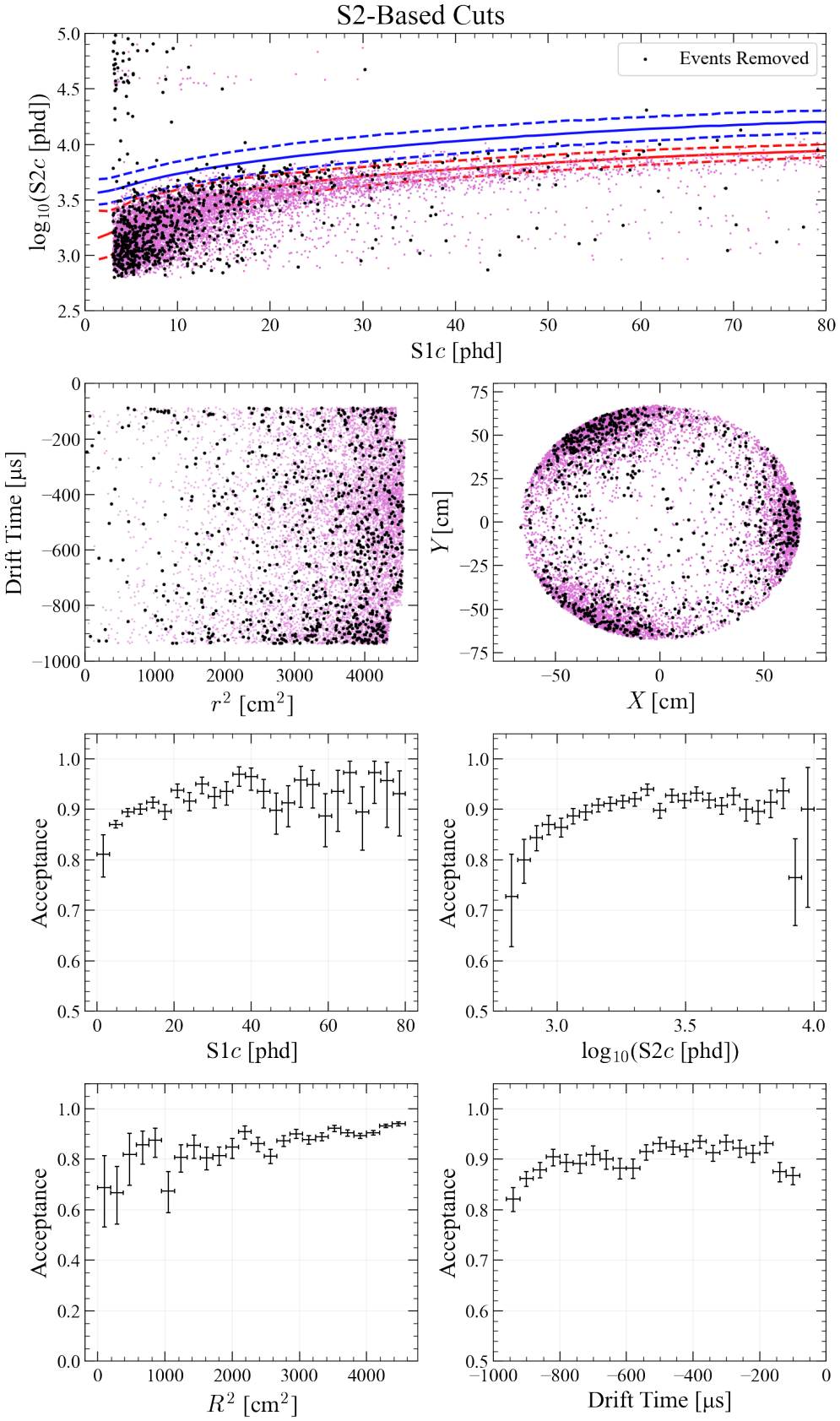


Figure 4.11: Distributions of AmLi events removed by (black) and surviving (pink) all S2-based data quality cuts, as well acceptance as a function of S1c, $\log_{10}$ S2c, r^2 and drift time. Solid blue and red lines represent the medians of the ER and NR bands, respectively, while dashed lines show the 10% and 90% quantiles.

region. For events in the NR band, this corresponds to a $\log_{10}$ S2c value of ~ 3.6, which can explain the dip in the acceptance of all S1-based cuts as a function of $\log_{10}$ S2c seen in Figure 4.10. The acceptance of the S2 Width vs Drift Time cut was seen to increase at high S2 pulse area. This cut was developed by performing fits in the width-drift time parameter space for a ^{83m}Kr calibration dataset, and setting the cut values at 3σ from the mean. The relationship between these variables becomes tighter for S2 pulses with a greater number of electrons, allowing for an improved pass rate. The XY quality cut showed a loss in acceptance at both high and low S2c. Inspections of waveforms showed that rejected events from AmLi data with a low S2c were generally not good events, and included pathologies such as multiple merged S2s and pile-up from e-trains. It was proposed that the dip in acceptance at high S2c could be a result of PMT saturation.

4.5.2 Datasets for Cut Acceptance

The official SR1 WIMP search cut acceptances for the S1-based cuts were determined using tritium calibrations, due to the uniformity of the source in the detector and the minimal accidental S1s following application of the e-/ph-train veto. For the S2-based cut acceptance curves, a purely ER source such as tritium was unsuitable, since the low-energy region of the NR band cannot be populated by higher-area ER S2s. Although events from both AmLi and DD neutron sources cover the full S2 range of the ROI, each source has its drawbacks. As the DD events were highly localised close to the top of the detector, the majority of events were outside the FV and, as discussed, the high rates of the AmLi sources resulted in a poor waveform environment.

The S2-based cut acceptances were therefore determined using an NR ChopStitch dataset, employing a similar process to that which was used for accidental event generation, but choosing an S1 and S2 as close to the desired drift time as possible. The S1 was taken from the tritium dataset, as were some larger S2s, whereas smaller S2s were taken from the AmLi data. The final acceptance for all S2-based cuts was defined as the percentage of events surviving the following sets of cuts, as described in section 4.3:

$$\text{S2-based Acceptance} = \frac{\text{Livetime} + \text{Physics} + \text{S2-based}}{\text{Livetime} + \text{Physics}}, \tag{4.5.3}$$

with an S1-based cut acceptance of:

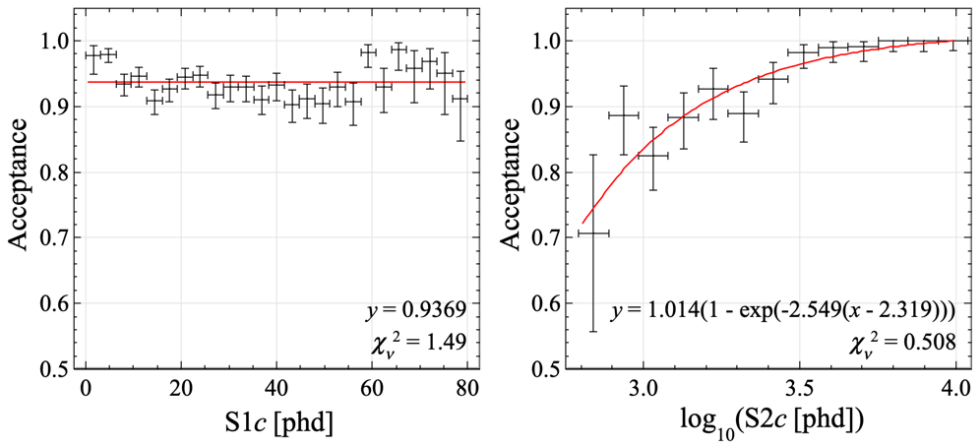


Figure 4.12: Official SR1 WIMP search acceptance curves using a tritium source for S1-based cuts (left) and NR ChopStitch for S2-based cuts (right). The fits shown in red were used in the subsequent analysis.

$$\text{S1-based Acceptance} = \frac{\text{Livetime} + \text{Physics} + \text{S2-based} + \text{S1-based}}{\text{Livetime} + \text{Physics} + \text{S2-based}}. \tag{4.5.4}$$

The final SR1 acceptance curves are shown in Figure 4.12. Fits of these curves were applied to the PDFs of each of the simulated signal and background sources required for limit setting, adjusting the weight of events, rather than removing them entirely. The overall NR acceptance as a function of recoil energy, including the efficiency of the trigger, is shown in Figure 4.13.

4.6 The SR1 WIMP Search Result

4.6.1 Observed Data

After application of all data quality cuts to SR1 WIMP search data, a total of 335 events remained. Their distribution in $\log_{10}$ S2c-S1c space is shown in Figure 4.14, with the majority of events populating the ER band. The spatial distribution of the events is shown in Figure 4.15, demonstrating the removal of backgrounds by the FV cut, as well as the Skin and OD vetoes.

4.6.2 Background Model

A profile likelihood ratio (PLR) was employed in $\log_{10}$ S2c-S1c space to test whether these observed data were in better agreement with a background-only or background and signal hypothesis. This required in-

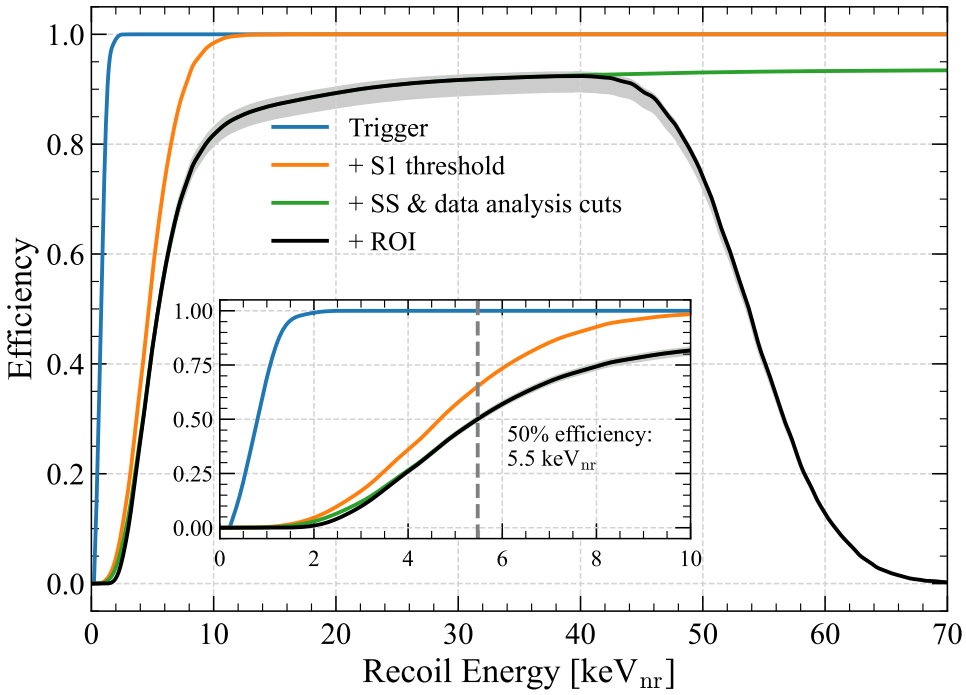


Figure 4.13: Signal efficiency as a function of nuclear recoil energy, demonstrating the cumulative effects of the trigger (blue), S1 three-fold coincidence requirement and 3 phd S1c threshold (orange), other analysis cuts (green), and the WIMP search ROI (black). Figure from Ref. [93].

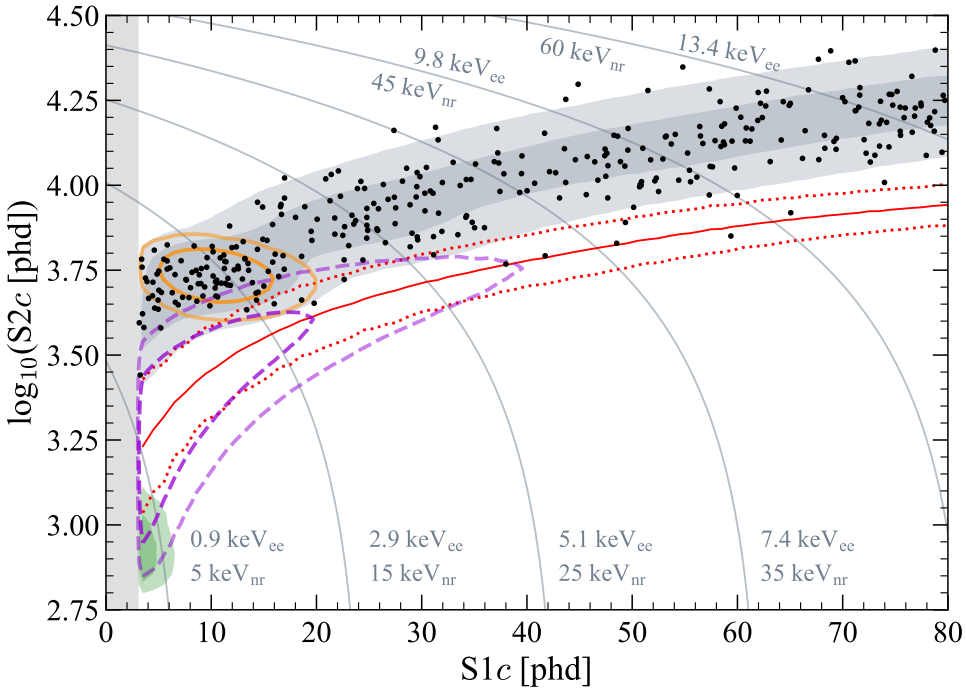


Figure 4.14: Distribution of SR1 WIMP search data in $\log_{10}$S2c-S1c space following all data quality cuts. 1σ and 2σ contours represent the following models: the best-fit background model (grey), the ^{37}Ar background component (orange), the ^{8}B solar neutrino component (green), and a 30 GeV WIMP (purple). Figure from Ref. [93].

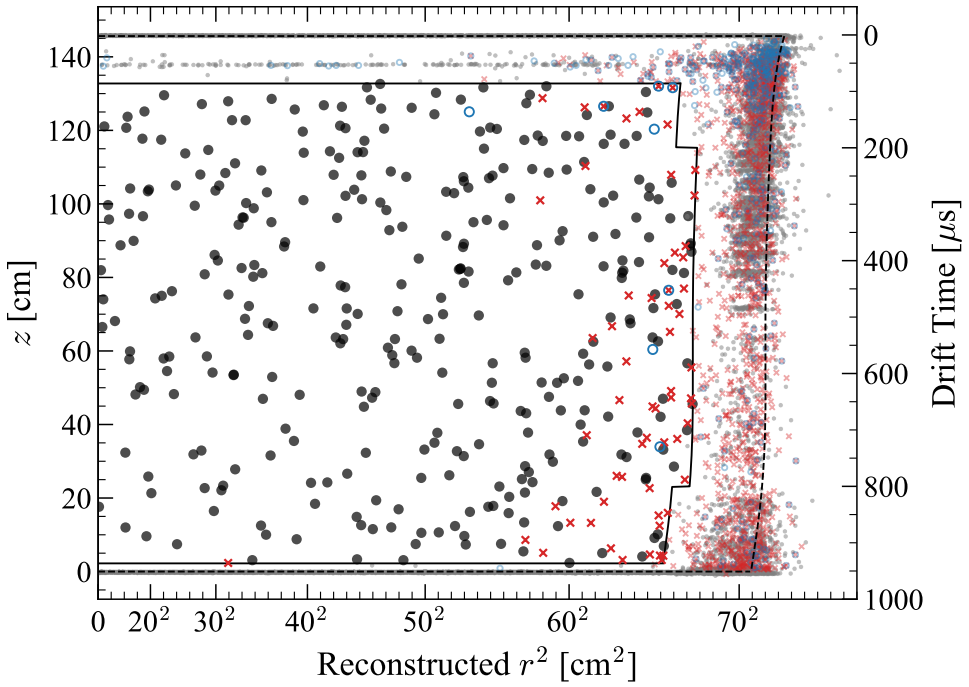


Figure 4.15: Distribution of SR1 WIMP search data in Z-r^2 space following all data quality cuts. The dashed line represents the TPC wall, whereas the solid line indicates the FV. Black and grey points show data inside and outside the FV, respectively. Red crosses represent events removed by the prompt Skin veto and blue circles similarly indicate events vetoed by a prompt signal in the OD. Figure from Ref. [93].

puts of signal PDFs in $\log_{10}$ S2c-S1c space, which were simulated using the Geant4-based BACCARAT package for masses between 9 GeV and 10 TeV for a spin-independent WIMP model. Distributions were also required for each of the components in the SR1 background model, as shown in Table 4.2. The normalisations of each simulated PDF were determined through calculations and analysis of the data.

Of the backgrounds discussed in section 3.3, ERs in the WIMP search ROI were predominantly the result of β decays of contaminants dispersed in the xenon. Contributions from ^{214}Pb and ^{212}Pb, from the ^{222}Rn and ^{220}Rn decay chains, respectively, as well as ^{85}Kr, all have broad energy spectra that are relatively flat across the entire energy range considered, so have been combined into a single background. By fitting S1c peaks of high-energy α sources on either side of ^{214}Pb (^{212}Pb) in the ^{222}Rn (^{220}Rn) decay chain, the rate of ^{214}Pb (^{212}Pb) was found to be 3.26 µBq/kg (0.14 µBq/kg). To determine the ^{85}Kr rate, the concentration of $^{\text{nat}}$Kr in the xenon was measured to be 144 ± 22 ppq (g/g) using a liquid nitrogen cold trap, accounting for both individual xenon bottles and residual xenon in the ICV [125]. γ rays from the detector components and caverns walls also produce a small flat ER spectrum [128], so were combined with the

β components. All cuts were applied to simulations for every component of the Detector ER background, before scaling by energy peak fit results outside the ROI.

A separate flat contribution to the ER background is also expected from solar neutrinos, the rate of which was calculated precisely using Refs. [143]–[145]. The physics backgrounds from decays of ^{136}Xe and ^{124}Xe through double β decay and double electron capture, respectively, also produce ER events, and expected rates were calculated using measurements of isotropic abundances and half-lives from Refs. [130], [132], [146].

^{127}Xe and ^{37}Ar resulting from cosmogenic activation of the xenon must also be accounted for. The rate of ^{127}Xe was constrained by measuring K-shell atomic de-excitations which occur outside the ROI. The number of ^{37}Ar events was estimated using the exposure of the xenon to cosmic rays before underground deployment and the decay time before the start of SR1 [126]. The wide flat constraint on this value accounts for the high uncertainty in the spallation model and the level of activation during transport.

Radiogenic neutrons from detector components contribute to the NR background. The rate prediction was derived by applying the OD neutron-tagging efficiency to OD-tagged Single Scatter events in the FV. The expected rate for each of the other background components was set to 5% of that determined for the WIMP search, due to the 5% chance of an accidental time-coincident signal in the TPC and OD. Statistical inference on the OD-tagged event sample produced 0 as the best fit number of neutron events. The NR background also includes a component from coherent elastic neutrino-nucleus scattering from ^{8}B solar neutrinos, which was determined following Refs. [143]–[145], but this is small due to the majority of such events being below the current S2 threshold.

Each of the backgrounds discussed, together with the accidentals distribution as shown in Figure 4.5, reweighted by the acceptance curves in Figure 4.12, form the SR1 background model. A fit of the model to the data was performed, with uncertainties in each background component included as fit constraints. For every WIMP mass tested, the data were consistent with the background-only hypothesis, as a total of zero WIMP events resulted in the best fit. The best-fit background model and data are shown as a function of reconstructed energy in Figure 4.16.

The only systematic uncertainties considered in this statistical inference were the uncertainties on the estimated background counts. Examples of systematic uncertainties are those introduced through the mea-

Table 4.2: The components of the background model, with the number of events expected within the 60 d × 5.5 t SR1 exposure. Also shown is the result of fitting the model, including a 30 GeV WIMP contribution, to data. Table reproduced from [93].

Source	Expected Events	Fit Result
β decays + Det. ER	215 ± 36	222 ± 16
ν ER	27.1 ± 1.6	27.2 ± 1.6
^{127}Xe	9.2 ± 0.8	9.3 ± 0.8
^{124}Xe	5.0 ± 1.4	5.2 ± 1.4
^{136}Xe	15.1 ± 2.4	15.2 ± 2.4
^{8}B CEνNS	0.14 ± 0.01	0.15 ± 0.01
Accidentals	1.2 ± 0.3	1.2 ± 0.3
Subtotal	273 ± 36	280 ± 16
^{37}Ar	$[0, 288]$	$52.5^{+9.6}_{-8.9}$
Detector neutrons	$0.0^{+0.2}$	$0.0^{+0.2}$
30 GeV WIMP	-	$0.0^{+0.6}$
Total	-	333 ± 17

sured branching ratios to naked β decay for ^{212}Pb and ^{214}Pb, and the half-lives for ^{124}Xe and ^{136}Xe, as well as theoretical values for neutrino fluxes [125]. The calculated values for total livetime and fiducial mass also introduced systematic uncertainties on the background counts. As LZ is a low-background experiment, these model uncertainties do not produce a significant effect on the sensitivity [124]. It can be seen in Table 4.2 that, aside from the β + Det. ER and ^{37}Ar components, the uncertainties resulting from the fit are identical to the input model uncertainties at the quoted precision. This is due to the tight constraints provided by sideband samples and measurements by other experiments, as well as the relatively low impact of these background components. The output systematic uncertainties are shown by the darker blue band on the best-fit background model in Figure 4.16. The impact of various other systematics was evaluated by rerunning the statistical inference, for example by replacing the signal efficiency from Figure 4.13 by the lower 1σ bound of the efficiency band, and the impact on the exclusion limit was shown to be small. External systematic uncertainties, such as those present in the dark matter halo model and nuclear form factors, were not considered for SR1. In addition to the systematic uncertainty, the statistical Poisson fluctuation on the expected number of events is shown as the lighter blue band in Figure 4.16.

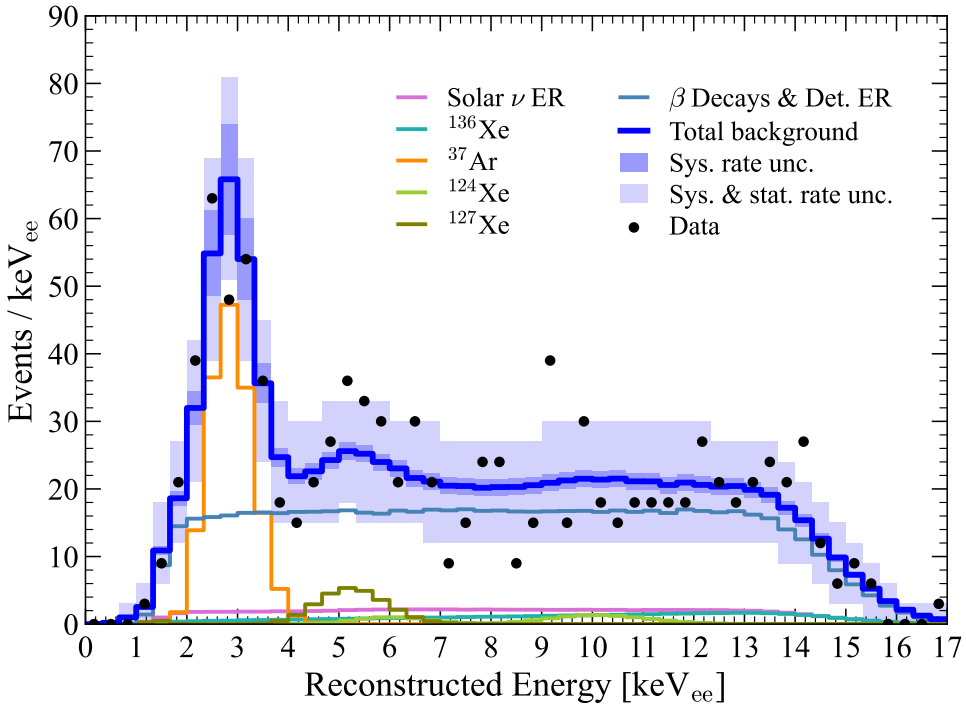


Figure 4.16: The best-fit background model as a function of reconstructed energy. The total of all background components is shown in blue and the black points are the data. The accidentals and ^{8}B solar neutrinos background components are included in the fit but are too small to be visible. Figure from Ref. [93].

4.6.3 SR1 Limits

Fig 4.17 shows the limits of the SI WIMP-nucleon cross-section above which WIMPs are excluded at the 90% confidence level, as a function of WIMP mass. For masses between 13 GeV and 36 GeV, the limit is significantly lower than the median expected limit due to background fluctuations, so has been power constrained following recommendations presented in Ref. [147]. At the time of writing, these are the most stringent limits for spin-independent WIMP-nucleon interactions for WIMP masses greater than 9 GeV, with the optimal sensitivity achieved at a WIMP mass of 36 GeV, where cross-sections of 9.2×10^{-48} cm^2 and above are rejected at the 90% confidence level. This has been achieved using only $\sim$ 6% of the estimated 1000-liveday total exposure of LZ, which will continue to explore new WIMP parameter space and operate other rare physics searches [124].

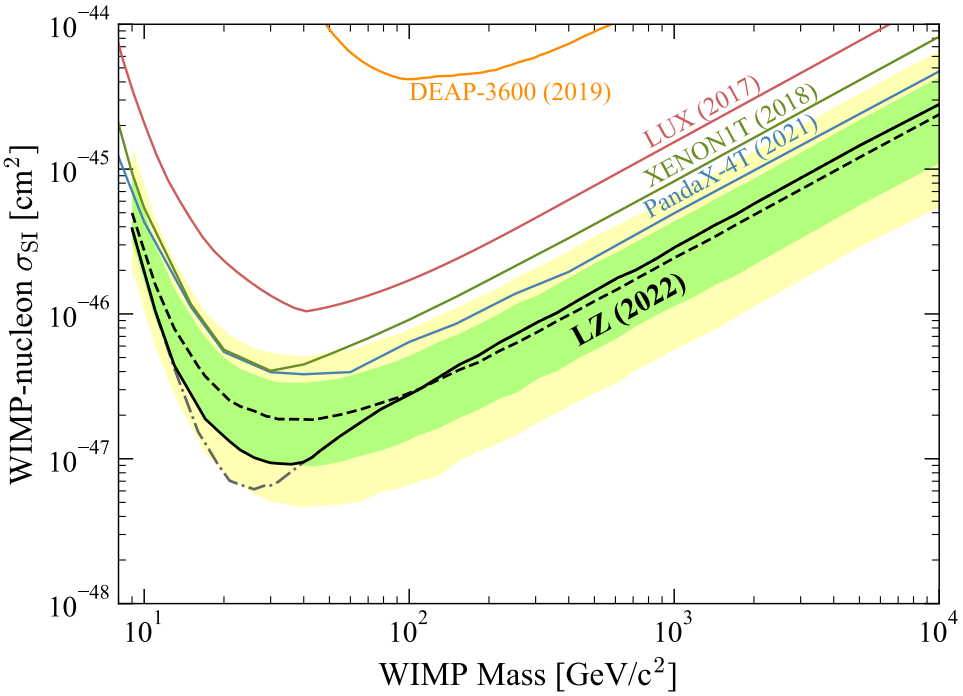


Figure 4.17: The SR1 WIMP search 90% confidence limit (solid black) for the spin-independent WIMP-nucleon cross-section as a function of WIMP mass. The dot-dash line shows this limit before power-constraining. The dashed line shows the median of the sensitivity projection, and the green and yellow bands are the 1σ and 2σ uncertainties on this. Limits showing recent results from other direct detection experiments are also included. Figure from Ref. [93].

Chapter 5

Light Collection Efficiency in Simulations

Accurate optical simulations are valuable for the development of waveform-based analyses in LZ, and it is essential that they are able to reproduce the position-dependent detection efficiency of scintillation photons. The improvement achieved through optimisation of several optical parameters will be the focus of this chapter.

5.1 LZ Simulations

The ability to accurately simulate events in the LZ detector is vital. Simulations were used in the optimisation of the detector design, for example to increase light collection, and in the calculation of the expected number of background events using measurements provided by the radioassay campaigns. As explained in subsection 4.6.2, PDFs produced using simulations of signal and background sources were used as inputs for the profile likelihood fit that produced SR1 exclusion limits. Two simulations chains are utilised by the LZ collaboration, each consisting of various software packages. A schematic of both chains is shown in Figure 5.1, and they will now be described in further detail.

BACCARAT (Basically A Component-Centric Analogue Response to AnyThing) is a Geant4-based package tailored for LZ [129], [148], that is used for both the fast and full simulations chains. It contains custom particle generators, with Geant4 tracking used to follow the resulting particles as they deposit energy in detector components. In the case of the parametric "fast" chain, these energy deposits can be output directly to a ROOT file, whereas for the full chain, additional processing is required.

As the fast chain tracks energy deposits only, rather than every optical photon, it is much less computationally expensive than the full chain,

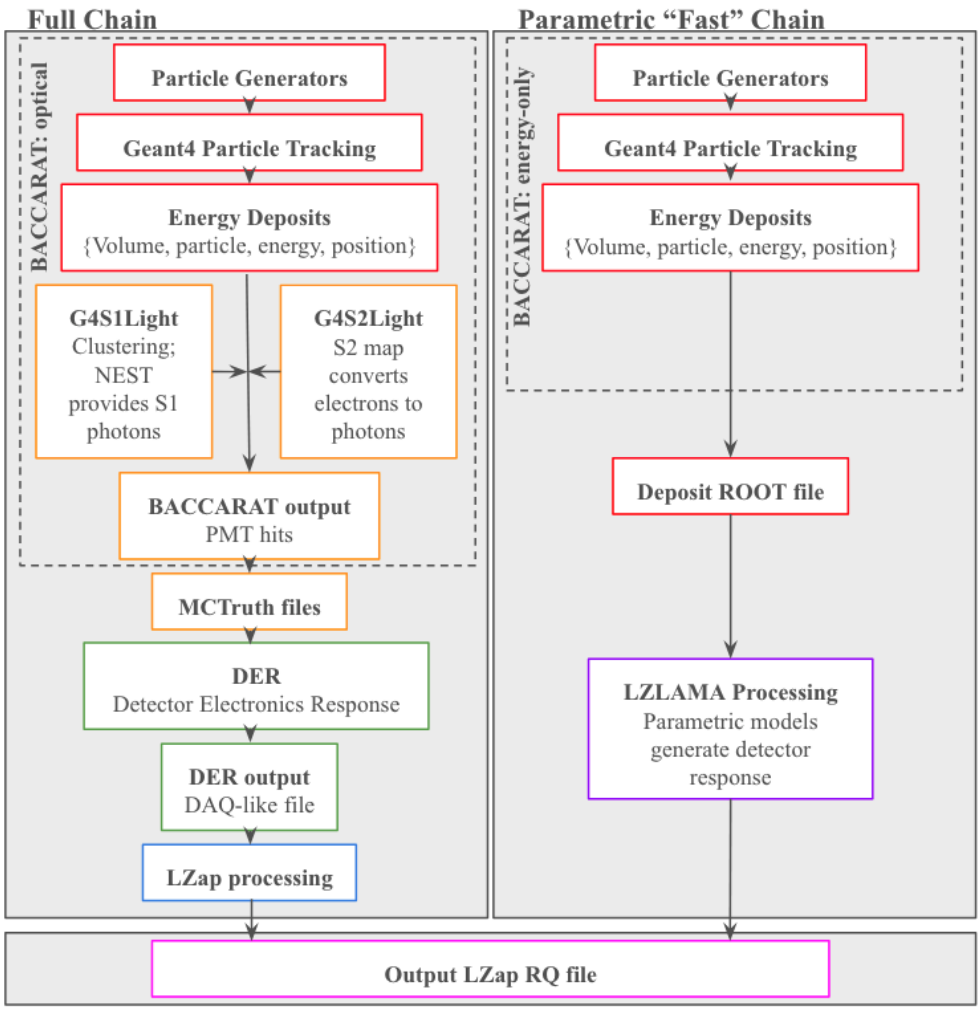


Figure 5.1: Schematic of the software packages used within the full and parametric "fast" LZ simulations chains, based on LZ documentation.

and is used to produce the templates used in limit setting. The BACCARAT output files are processed with the LZ Light Analysis Monte Carlo Application (LZLAMA), which converts energy deposits in sensitive volumes into RQs, including the average number of observed S1 and S2 photons, using parametric models of the detector response. LZLAMA achieves this by performing clustering of energy deposits, before passing them to NEST, which calculates the S1 and S2 response using maps of light collection efficiency and electric field. The arrival time and PMT hit pattern of photons is not modelled, and simulations created in this manner lack features such as electronic and geometric effects. Therefore, in order to study complex event pathologies and develop cuts, it is beneficial to also have full optical simulations available.

In the full-chain mode, BACCARAT converts energy deposits to S1 and S2 signals using G4S1Light and G4S2Light. G4S1Light performs clustering and interfaces with NEST, while G4S2Light uses an S2 map to convert ionisation electrons into photons. Full optical ray tracing is then used to determine which photons are collected by PMTs, resulting in an output file containing PMT hits. This is used to produce Monte Carlo Truth (MCTruth) information about the energy deposit, or "vertex", that produced the relevant photons. The Detector Electronics Response (DER) package converts raw photon PMT hits to digitised waveforms by modelling the front-end electronics of LZ, including the PMTs, cables and DAQ. The resulting ROOT file is in the same format as those which are output from the DAQ, and can be passed through LZap in the same way as raw data, in order to generate RQs from waveforms.

An example contribution of the full simulations chain to LZ data analysis was the validation of an additional inefficiency in the detector, whereby, at large drift times, S2 pulses were more likely to be split in two if only a single template was used in S2 identification. This efficiency loss was quantified using simulations and applied to the S2 acceptance curves. In order for the full simulations to be useful, they must first be tuned to match real data. One important consideration is the light collection efficiency (LCE) for S1 signals.

5.2 Light Collection Efficiency in the LZ Detector

The LCE is defined as the gain for S1 signals in the LZ detector and is known as the g_1 value. This is the average fraction of S1 light detected by TPC PMTs and is normalised to the geometric centre of the detector.

For LZ, this value accounts for both the proportion of photons reaching a PMT, and effects resulting from the PMTs and electronics. Note that this differs from some collaborations, where the term only accounts for the former [149].

LCE is position dependent, particularly in the Z direction, as geometry dictates that a PMT array close to an event will subtend a much greater solid angle than a PMT array far from the event. This results in a higher LCE in the respective arrays for events close to the top or bottom of the detector, as more photons are detected directly, rather than requiring many reflections to reach a PMT. More S1 light is seen in the bottom array than the top, due to reflections from the liquid surface.

In order to ensure that pulse sizes are independent of position, a correction is applied to data. In SR1, this was achieved by generating a correction map that divided the detector into 6859 voxels in X vs Y vs drift time space. ^{83m}Kr events from calibration data were selected, and Gaussian means of their S1 pulse area were found. Dividing these by the mean value in the central bin, and extrapolating to bins with a low event count, provided a correction factor for each voxel. The three-dimensional correction map used in SR1 is shown in Figure 5.2.

These position corrections were important for the process of calculating the g_1 value. Gaussian means of the corrected S1 and S2 pulse areas, S1c and S2c, for monoenergetic sources of known energies from both calibrations and background data, were added to a Doke plot [102], as shown in Figure 5.3. The linear relationship is a result of the anti-correlation between the number of photons and electrons produced in an interaction. The reconstructed energy equation, Equation 2.6.13, can be rearranged to give

$$\frac{\mathrm{S2}c}{E} = -\left(\frac{g_2}{g_1}\right) \cdot \left(\frac{\mathrm{S1}c}{E}\right) + \frac{g_2}{W}, \tag{5.2.1}$$

allowing a g_1 value of 0.1133 ± 0.0030 phd/photon to be derived from the gradient and y-intercept of the fit of the Doke plot. The g_2 value, defined as the effective charge gain for the S2 signal, was similarly determined. Large S2 pulses in the top array can lead to PMT saturation [106], which artificially reduces the size of the S2 signal, so the S2 light in the bottom array only, S2c_{bottom}, is used in the calculation, rather than S2c. This provided a g_2^{bottom} value of 14.75 ± 0.56 phd/electron, which can be converted to a g_2 of 47.1 ± 1.1 phd/electron by dividing by the S2 top-bottom asymmetry, which had a constant value for SR1. As this method was fully data-driven, the simulations were tuned to match this g_1 value of 0.1133 ± 0.0030 phd/photon.

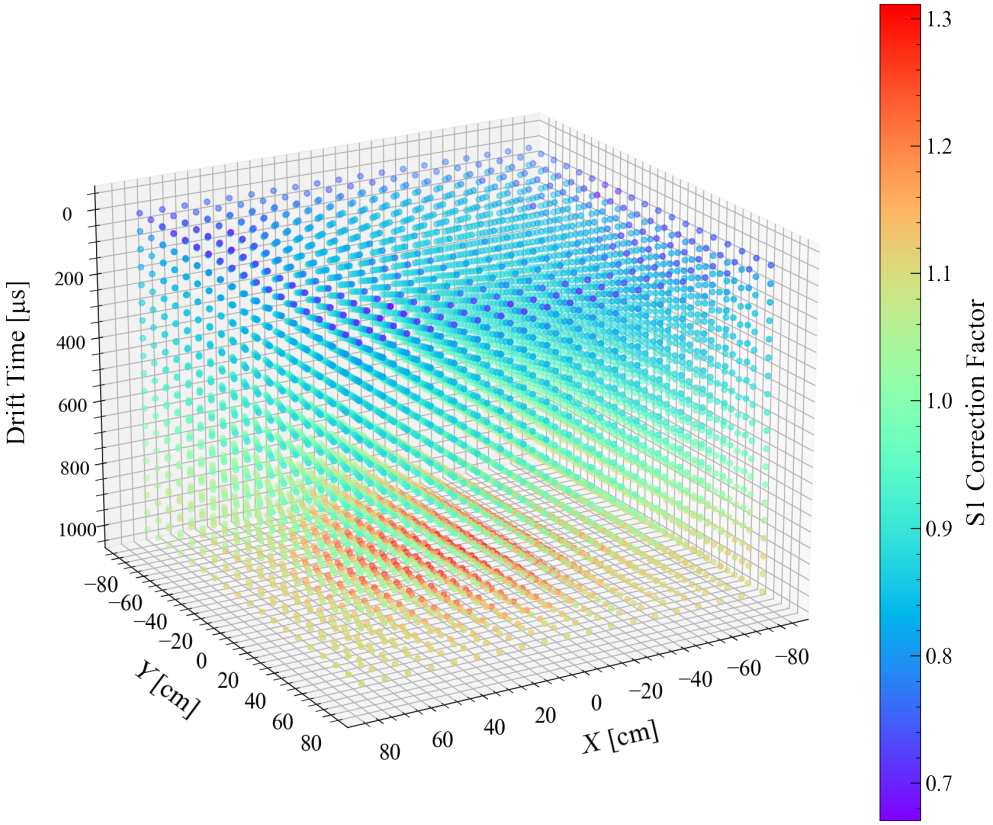


Figure 5.2: Three-dimensional correction map generated using ^{83m}Kr calibration data under SR1 conditions, with code adapted from G. Rischbieter and M. Murdy. Each data point represents the S1 correction factor in one of the 6859 voxels in X vs Y vs drift time space.

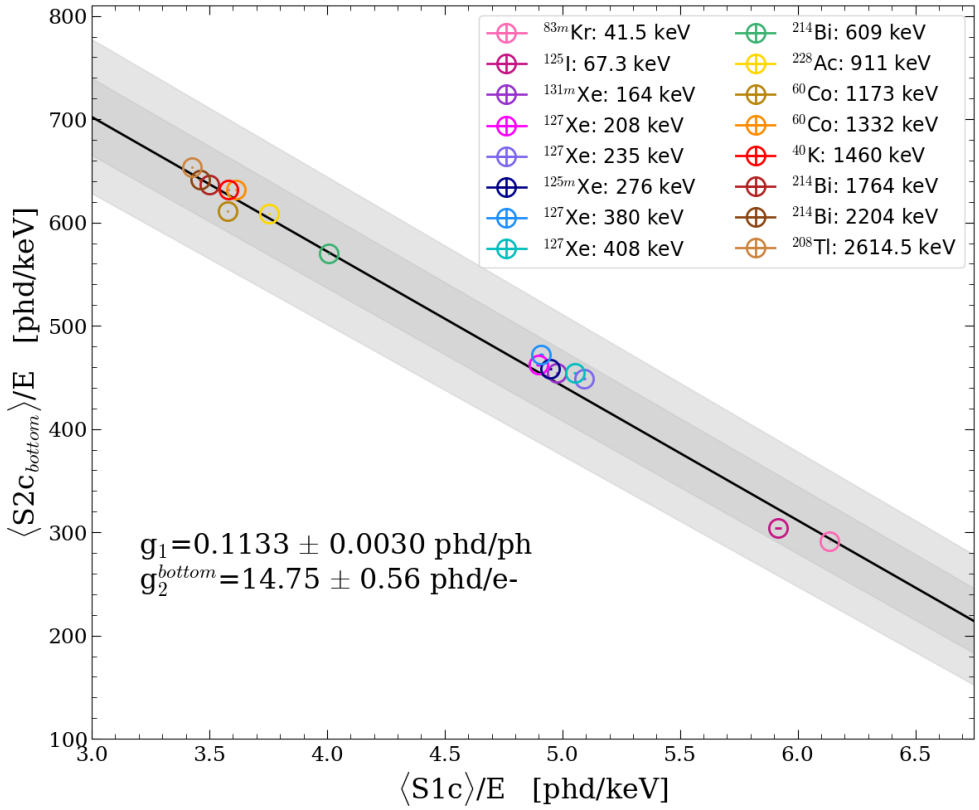


Figure 5.3: Doke plot, courtesy of G. Blockinger, showing the anti-correlation of S1 and S2 signals. The fit of the data points from several monoenergetic sources was used to determine the light and charge gain.

This value was in very good agreement with the results of an alternative method, in which NEST was used to produce ER and NR bands. A raster scan was performed over pairs of g_1 and g_2 values, minimising deviations between the medians of the simulated bands and bands from tritium and DD calibrations, as well as discrepancies between the reconstructed energies and the true decay energies of ^{83m}Kr, ^{131m}Xe and ^{127}Xe sources. This provided a g_1 value of 0.1149 ± 0.0021 phd/photon and a g_2 value of 46.38 ± 1.51 phd/electron.

5.3 Light Collection Efficiency in LZ Simulations

In order to investigate the LCE in simulations and carry out comparisons to data, the BACCARAT package was used in the full optical mode to simulate 10 million photons at various heights throughout the LXe. This is known as a "photon bomb". The photons were constrained to originate within the boundaries of -44.44 mm $< X, Y <$ 44.44 mm, which correspond to the most central bin in the S1 position correction map. Optical tracking was used to determine which photons reached a

PMT photocathode.

The PMT quantum efficiency (QE) was then applied in the same manner as is ordinarily implemented in the DER. The value provided by Hamamatsu for each PMT, in the range of 25-40%, was adjusted to account for several effects. This included the relative improvement in QE of $(17.9 \pm 5.2)\%$ when the PMT is cooled from room temperature [118], and the collection efficiency of the first dynode. At the operating voltages of LZ, 85% is the collection efficiency reported by the manufacturer, but this was likely only measured for visible light. Calculations accounting for the DPE effect at VUV wavelengths, where two photons can overcome the inefficiency, provided a value of 89% [118]. For each photon that reached a PMT photocathode, a random number was drawn from a uniform distribution and, if less than the quantum efficiency of that PMT, a count was added to the appropriate height bin in a histogram.

Scaling by the total number of photons generated provided a value for LCE in each height bin. The same process can also be applied to photons reaching PMTs in the top or bottom array only, to produce corresponding curves for the top and bottom array LCE. LCE curves produced using the default BACCARAT settings are shown in Figure 5.4. The LCE curves can then be fit with a polynomial function in order to determine the value at the detector centre. This allows for a direct comparison to the data g_1 value calculated from the Doke plot.

5.4 Optical Parameters

Many optical parameters are available to tune in the BACCARAT code, and three were found to have the most significant impact on both the absolute LCE at the detector centre, and the variation of LCE with height. These were the reflectivity at the LXe-PTFE interface, the LXe absorption length and the LXe Rayleigh scattering length. The PTFE reflectivity is described using the UNIFIED model in Geant4 [150] and is modelled as entirely diffuse, when in reality there is a specular component which is dominant at high angles of incidence [151]. In future, implementation of a more accurate model in BACCARAT could improve simulations.

The expected ranges of values for each optical parameter at the mean LXe scintillation wavelength are taken from Ref. [149], and are shown in Table 5.1. The default values previously used in BACCARAT simulations were a LXe-PTFE reflectivity of 97.3%, a LXe absorption length of 100 m and a LXe Rayleigh scattering length of 0.3 m. Other parameters,

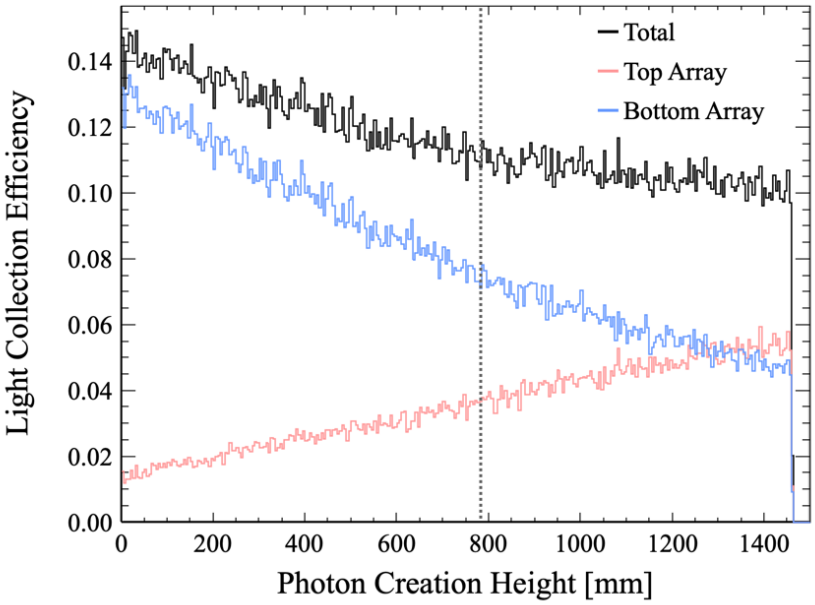


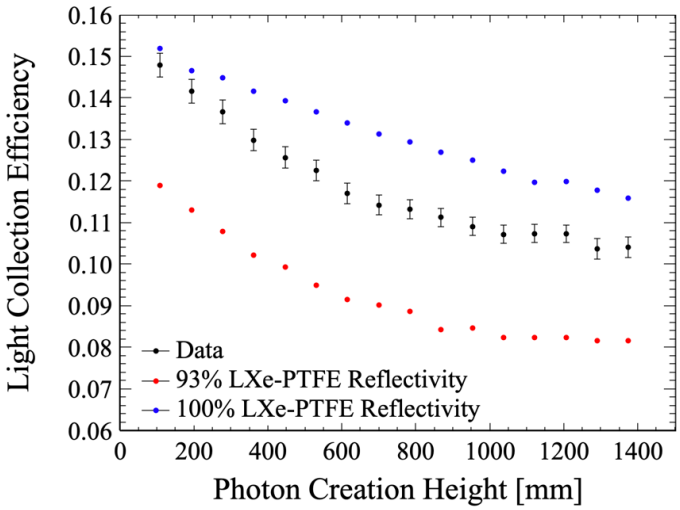
Figure 5.4: Percentage of simulated photons detected as a function of height above the cathode grid. The S1 correction map is normalised to the detector centre at $Z = 784$ mm, indicated by the black dotted line, so the LCE value at this point is directly comparable to g_1 from data.

Table 5.1: Expected ranges from Ref. [149] of optical parameters at the mean LXe scintillation wavelength.

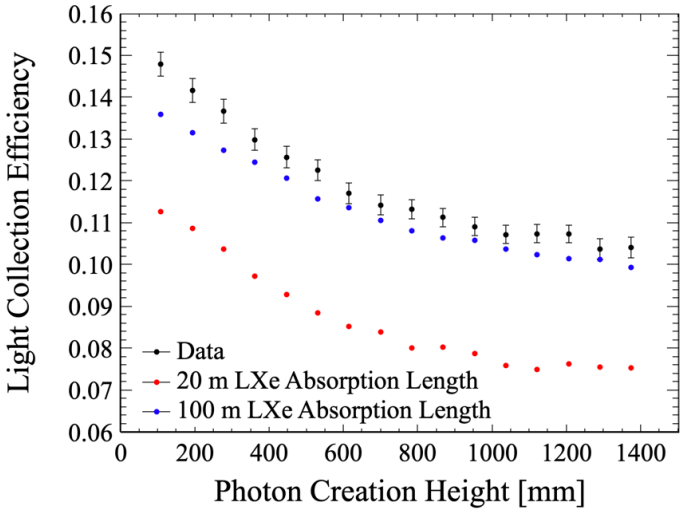
Parameter	Range
LXe-PTFE Reflectivity [%]	93 - 100
LXe Absorption Length [m]	20 - 100
LXe Rayleigh Scattering Length [m]	0.2 - 0.5

such as the reflectivity of the stainless steel grids, were found to have an impact on the overall LCE, but as the dependence on height was minimal and the true value is unknown, these were not tuned separately.

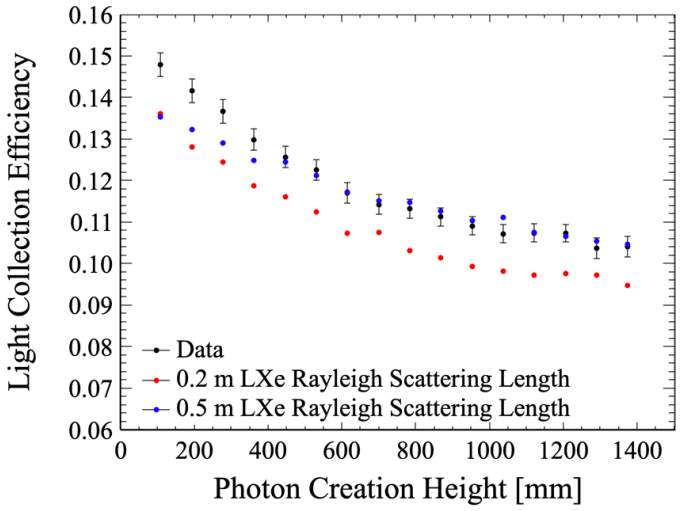
The impact of each of the optical parameters is demonstrated in Figure 5.5, which shows the LCE curves for the most extreme parameter values from the expected ranges. For a direct comparison to data, the S1 correction factors, determined using ^{83m}Kr calibrations, multiplied by the g_1 value are also shown. The simulated data have been rebinned to match the binning of the correction map. It can be seen that the LXe-PTFE reflectivity and LXe absorption length both have a significant impact on the total amount of light collected, whereas the LXe Rayleigh scattering length has a more obvious effect on the shape of the distribution.



(a) LXe-PTFE reflectivity



(b) LXe absorption length



(c) LXe Rayleigh scattering length

Figure 5.5: LCE curves for photon bomb simulations with extreme parameter values compared to S1 correction factors multiplied by g_1 from data. Poisson error bars were calculated for the simulations, but are too small to be seen. Error bars for the data combine the systematic error on g_1 with errors on the Gaussian means used to determine the correction factors.

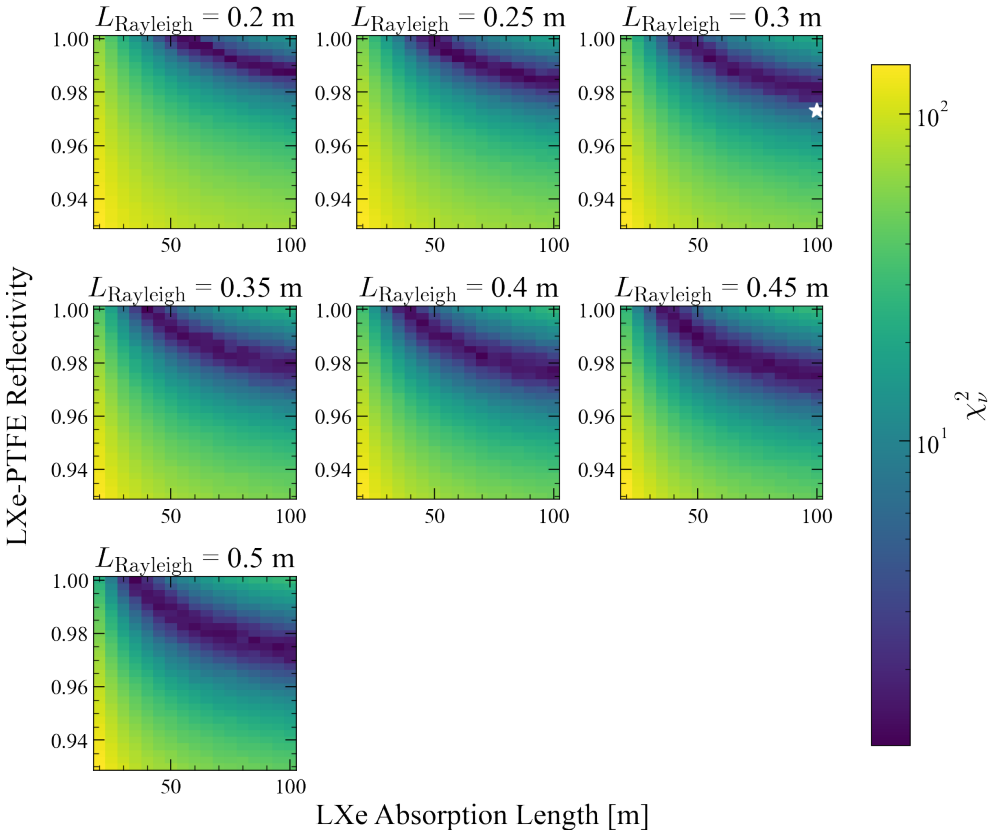


Figure 5.6: Subplots showing χ^2_ν values resulting from a comparison to data, for all combinations of LXe-PTFE reflectivity and LXe absorption, for each LXe Rayleigh scattering length tested in the grid search. The default BACCARAT parameters are indicated by the white star.

5.5 Parameter Optimisation

5.5.1 Method and Discussion

In order to tune the LCE in simulations to match what is observed in data, a grid search was performed by varying all three parameters at once within the given ranges. A χ^2 comparison was then used to quantify deviations between the number of surviving photons in each height bin in the simulations to the expected number of photons if the g_1 value and S1 correction factors taken from data were applied to the original number of photons simulated in that bin. This analysis was originally performed only on the total LCE from the sum of the top and bottom arrays. Each two-dimensional subplot in Figure 5.6 shows the LXe-PTFE reflectivity against the LXe absorption length for one of the LXe Rayleigh scattering length values tested.

Degeneracy can be seen between the LXe-PTFE reflectivity and LXe absorption length, as each subplot has a valley of χ^2_ν values. It can be seen that the previous default parameters are not contained within this valley, demonstrating the potential for improvement. Intensity, I, of light

in LXe is described by:

$$I(x) = I_0 e^{-x/\lambda_{att}}, \tag{5.5.1}$$

where I_0 is initial intensity, x is distance travelled, and λ_{att} is attenuation length, which depends on the absorption length and Rayleigh scattering length through $1/\lambda_{att} = 1/\lambda_{abs} + 1/\lambda_{scat}$ [152]. As LCE depends on the total number of scintillation photons detected by PMTs, only photon loss due to absorption will be considered. The intensity is also reduced each time a reflection from the PTFE surface occurs. Assuming that photons interact with only the LXe and PTFE materials, intensity of light following n reflections can be written as:

$$I_n = I_{n-1} e^{-d_n/\lambda_{abs}} R = I_{n-2} e^{-(d_n + d_{n-1})/\lambda_{abs}} R^2 = ... = I_0 e^{-L/\lambda_{abs}} R^n, \tag{5.5.2}$$

where d_n is the distance travelled between the $(n-1)^{\text{th}}$ and n^{th} reflections, R is the LXe-PTFE reflectivity and L is the total distance. A greater number of reflections n typically results in a larger total distance L, and both variables also depend on the position of photon origin. For a fixed detection probability for a single photon, R is related to λ_{abs} through:

$$R \propto e^{L/(\lambda_{abs} n)}, \tag{5.5.3}$$

which describes the exponential shape of the valley in Figure 5.6. A shorter Rayleigh scattering length would generally result in larger values of both n and L. The smallest χ^2_ν values were found for a Rayleigh scattering length of 0.2 m, which was the lower bound of the expected range. Tests of lower values outside this range resulted in even smaller χ^2_ν values.

Absorption of scintillation photons in pure LXe is negligible, and any absorption is thus due to impurities, in particular water and oxygen. It is conceivable that the LXe absorption length could exceed 100 m due to the high purity of the xenon, in which electronegative impurities were required to be suppressed to the level of $\sim$ 0.1 ppb [153]. The grid search was therefore extended up to a LXe absorption length of 1000 m for a scattering length of 0.2 m, as shown in Figure 5.7. The PTFE reflectivity producing the minimum χ^2_ν value was determined for each absorption length, and these data points were fit with an exponential function, which flattens as the absorption length becomes much greater than the height of the TPC. Since the LXe-PTFE reflectivity and LXe

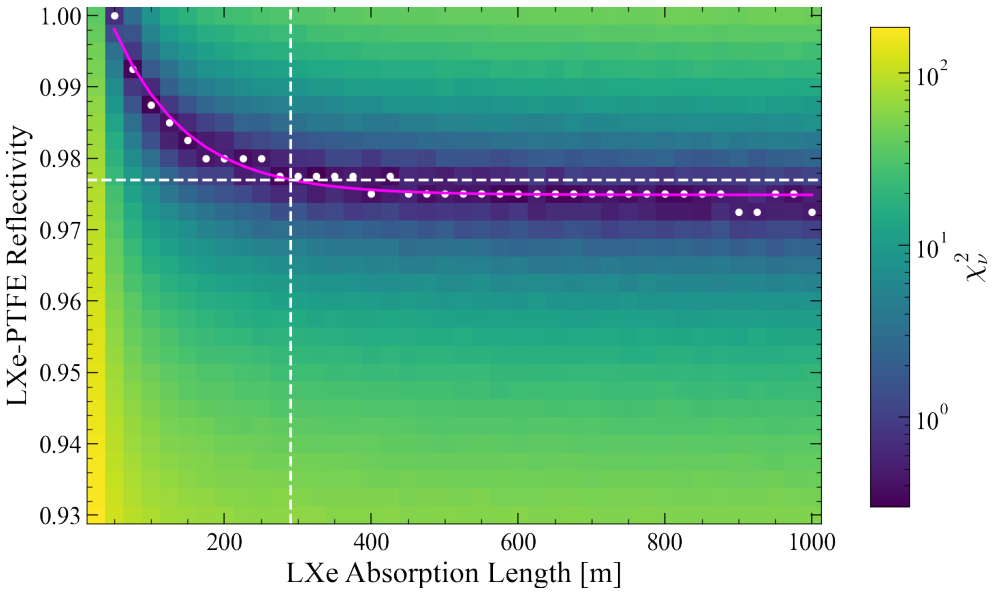


Figure 5.7: χ^2_ν values resulting from a comparison to data for a LXe Rayleigh scattering length of 0.2 m. The range of values for LXe absorption length has been extended up to 1000 m. The white data points show the PTFE reflectivity producing the minimum χ^2_ν for each value of LXe absorption length, and have been fit with an exponential function. The white dashed lines indicate that the reflectivity of 97.7% measured at Coimbra-LIP corresponds to an absorption length of 290 m.

absorption length were degenerate, sensible values were selected for the pair of parameter values. Using the experimental setup described in Ref. [154], a dedicated chamber at Coimbra-LIP was used to perform measurements of the reflectivity in LXe of the PTFE used in LZ. This was determined to have a value of 97.7%, which, from the fit of the minimum valley, corresponded to a LXe absorption length of 290 m.

The LCE curve for a photon bomb simulated using these tuned parameters is shown compared to the default parameters in Figure 5.8, and a better match to data is demonstrated at all heights. The relative LCE was also determined for simulated ^{83m}Kr events, produced by using the relevant event generator in BACCARAT, which implements the relative branching fractions from Ref. [155] for the different decay modes, before G4SLight provides the yields using NEST models. The DER and LZap were then run over the MCTruth output files. The Gaussian means of S1 pulse areas were determined relative to the central bin, in order to provide a direct comparison to the S1 correction map. As the simulated events originated at a range of radii, it was decided that only events in the range -40 cm $< X, Y <$ 40 cm, in which the correction factors are relatively uniform, would be selected for both simulations and data. The non-uniformity is most evident near the bottom of the detector, where the LCE is higher at small radii, as seen in the two-dimensional correction

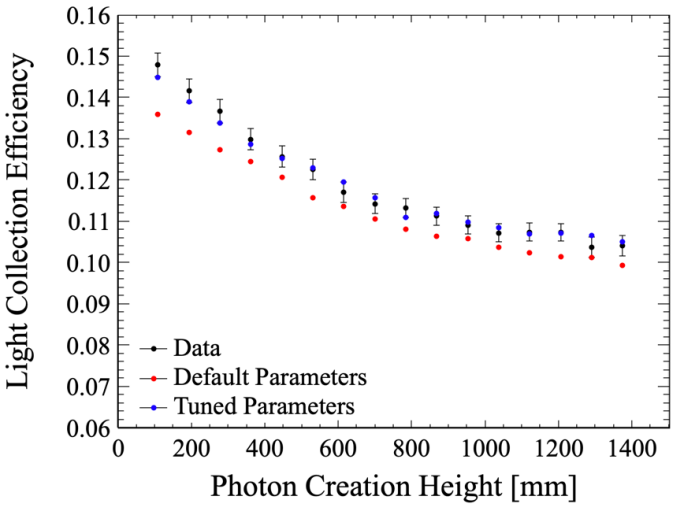


Figure 5.8: LCE curve for a photon bomb simulation using default and tuned optical parameters, compared to S1 correction factors multiplied by g_1 from data.

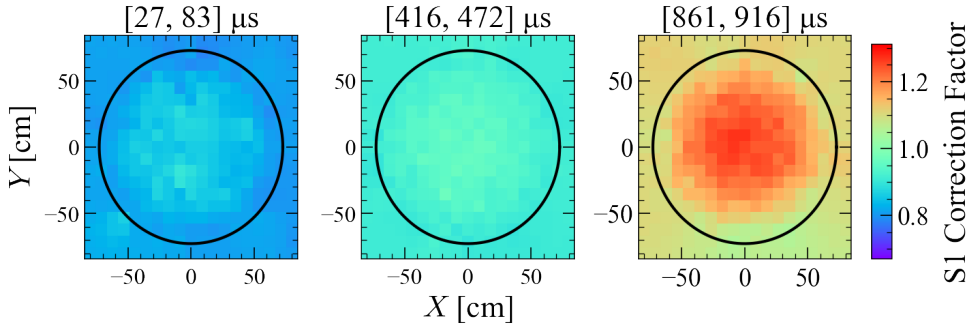


Figure 5.9: The S1 correction map generated using ^{83m}Kr calibration data, as shown in Figure 5.2, but in XY space for three drift time slices, to show the non-uniformity with radius near the bottom of the detector. The black circles show the position of the TPC wall.

maps shown in Figure 5.9. The results of this comparison are shown in Figure 5.10. Again, an improved match to data can be seen, particularly near the bottom of the detector.

5.5.2 S1 Area Discrepancy

Although an improvement was observed in the relative LCE for ^{83m}Kr, there was a significant discrepancy in the absolute S1 area, as shown in Figure 5.11. The distributions were fit with skew Gaussian functions of the form $f(x) = 2\phi(x)\Phi(\alpha x)$, where $\phi(x)$ is the standard Gaussian distribution and $\Phi(\alpha x)$ is the cumulative distribution [156]

$$\Phi(\alpha x) = \int_{-\infty}^{\alpha x} \phi(t)dt, \tag{5.5.4}$$

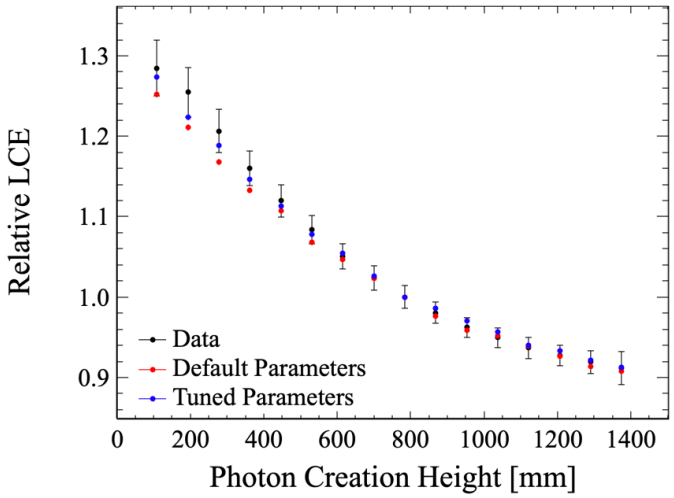


Figure 5.10: Mean S1 area relative to the detector centre for a ^{83m}Kr source using simulations with both default and tuned parameters, compared to S1 correction factors from data. The error bars on the data combine the statistical error on the Gaussian means with the standard error due to the variation of correction factors across XY space.

in which α is the shape parameter which determines the skewness. Although the χ^2_ν values resulting from the fits are relatively high, it can clearly be seen that the peak of the distribution is $\sim$ 50 phd lower for the tuned simulations than for data. Several studies were carried out to investigate the cause of this behaviour.

First, a comparison was made between the number of raw S1 photons generated in BACCARAT and a recent version of NEST, 2.3.12. The results can be seen in Figure 5.12, where the two peaks correspond to the 32.1 keV and 9.4 keV transitions from the ^{83m}Kr decay. The means of the BACCARAT peaks were lower than the values determined from NEST and, following further investigation, it was found that the calculation of ^{83m}Kr yields in G4S1Light used an old model, from NEST version 2.0.1. It was therefore identified that BACCARAT would need to be updated to use the new NEST ^{83m}Kr model, which was predominantly based on analysis of ^{83m}Kr S1 signals in the PIXeY detector [133]. As the accuracy of the fast chain, which was used to produce the signal and background PDFs, had been the priority for SR1, this study was an important step in the validation of the full chain simulations.

In the meantime, S1 area comparisons were made for another monoenergetic source, ^{131m}Xe, for which the model was up to date. Figure 5.13a shows that a discrepancy between simulations and data was still present. The simulations tuning had been carried out without accounting for secondary PMT effects, which are usually applied in the DER. These include

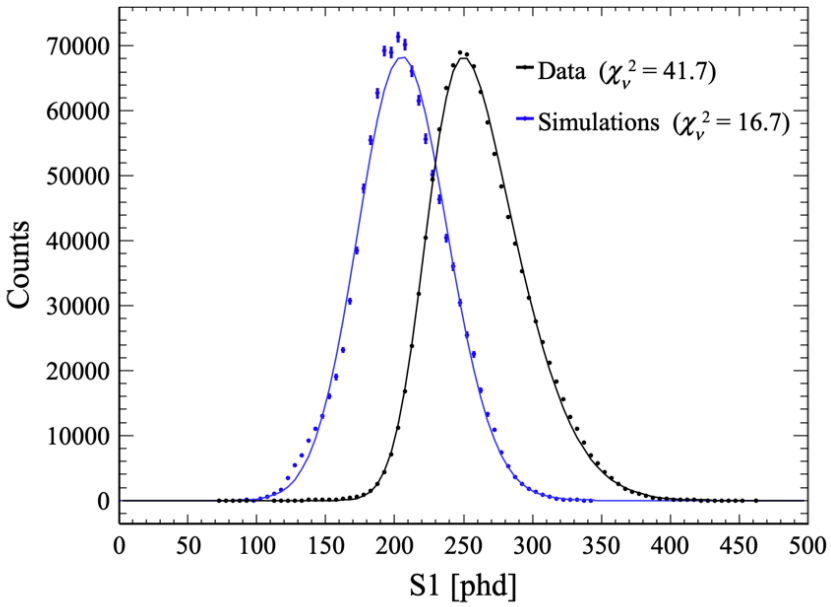


Figure 5.11: Total S1 area for a ^{83m}Kr source, produced using full chain simulations with tuned parameters, and data. The distributions have been fit with skew Gaussian functions.

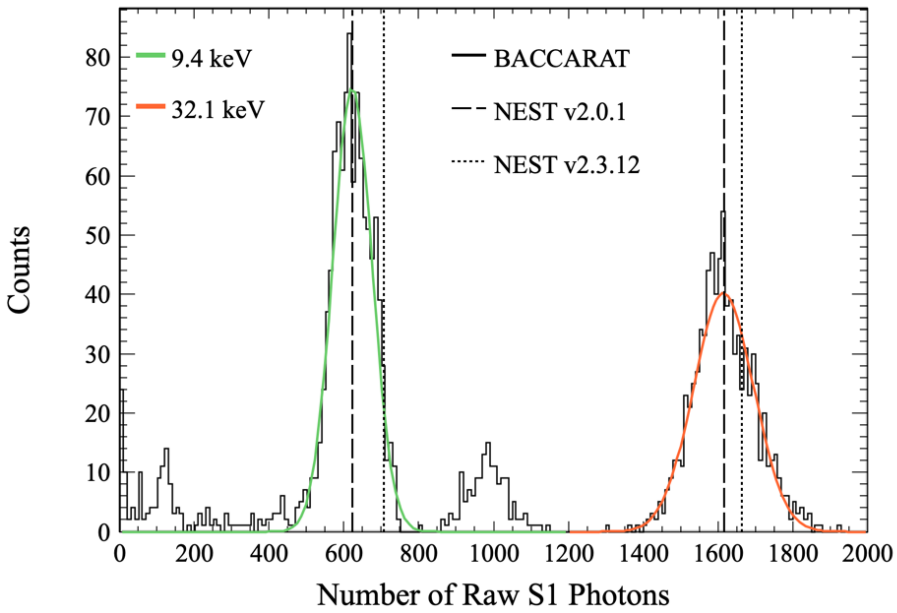


Figure 5.12: The histogram shows the number of raw S1 photons produced for a ^{83m}Kr source in BACCARAT, where the peaks for the 9.4 keV and 32.1 keV transitions have been fit with Gaussian functions. The means of Gaussian fits from ^{83m}Kr data generated with two versions of NEST are indicated with the vertical dotted and dashed lines.

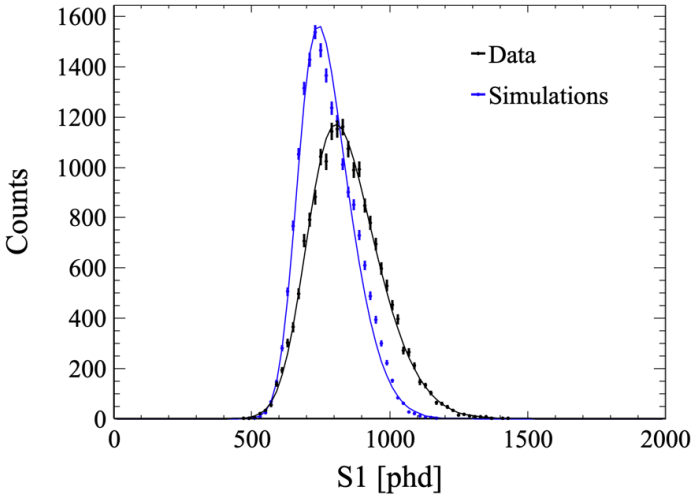
first dynode hits, the process in which a photon is transmitted through the PMT photocathode and induces a photoelectron at the first dynode [157]. As electron multiplication only begins at the second dynode, the total PMT gain is reduced by the gain of the first dynode. The PMT transit time is decreased by the difference in time taken for a photon to travel from the photocathode to the first dynode, compared to an electron. Photoelectrons can also backscatter elastically or inelastically off the first dynode or the structures that surround it, decelerating and stopping before they reach the photocathode, and then beginning the typical electron multiplication process from the first dynode [158]. This results in a smaller pulse, as well as a time delay. These factors all lead to smaller S1 pulse sizes, which had not been accounted for when tuning the simulations in BACCARAT alone.

As shown in Figure 5.13b, when these secondary effects are turned off in the DER, the mean of the simulated S1 pulse areas is brought much closer to the data, although the standard deviation is still slightly lower. As these effects were tuned for a higher gain than that at which the LZ PMTs are currently operated, their impact is likely not as significant in reality. It was therefore decided that the DER would be run with secondary PMT effects turned off until further tuning could be conducted.

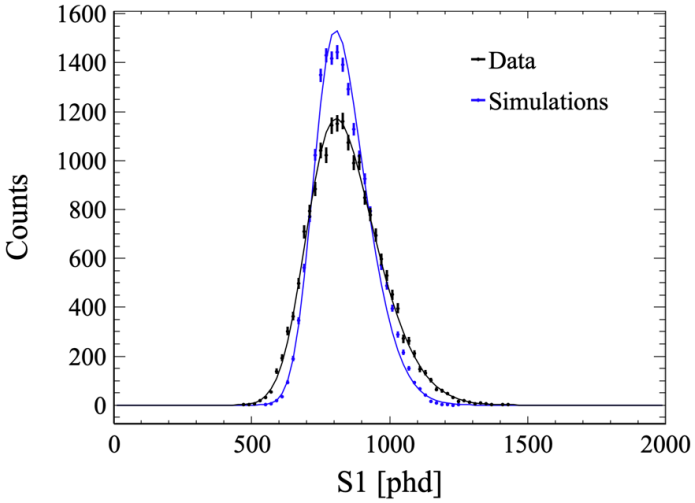
A final check was carried out, in which a mock source in the DER was used to simulate single photons with VUV wavelengths. Their areas were then compared to those of SPEs from SR1 WIMP search data in Figure 5.14. The peak value remains slightly lower for the DER simulations, which may be due to a discrepancy in the SPE calibration between the DER and LZap. The distribution is again narrower for the simulations.

5.5.3 Results

Gaussian means were determined for S1 pulse areas in slices of height throughout the TPC, using tuned optical parameters in BACCARAT, with no secondary PMT effects in the DER, and a ^{131m}Xe source that is correctly modelled in NEST. Figure 5.15 shows that there is much better agreement with data throughout the detector.



(a)



(b)

Figure 5.13: Total S1 area for a ^{131m}Xe source produced using tuned full chain simulations compared to data. In (a), secondary PMT effects are turned on in the DER, and in (b) they are turned off.

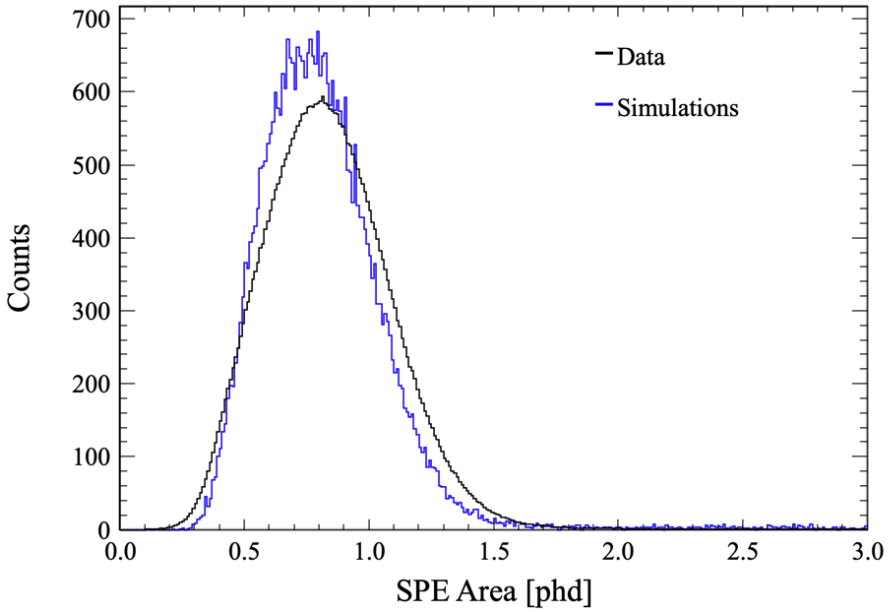


Figure 5.14: The pulse area of single photons generated in the DER, compared to SPEs from WIMP search data.

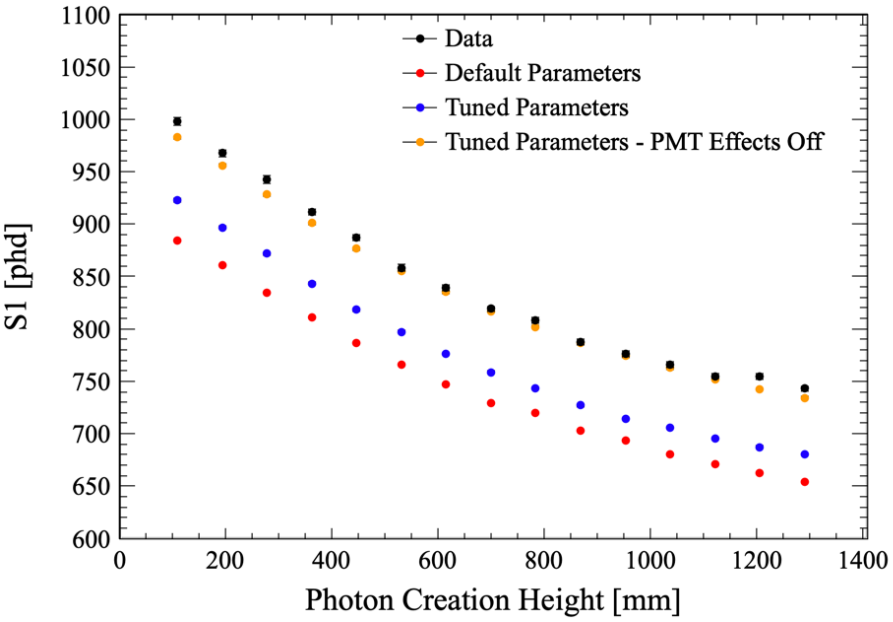


Figure 5.15: Comparison of mean S1 area for ^{131m}Xe in slices of height in the TPC for simulations with default parameters, tuned parameters, and tuned parameters with secondary PMT effects turned off, as well as data. Error bars from uncertainties on the Gaussian means are too small to be seen.

5.6 Array-dependent Parameter Optimisation

5.6.1 Method and Discussion

The tuned simulations were able to replicate the total LCE in the LZ detector well. However, when plotting the LCE separately for the top and bottom arrays as in Figure 5.16, it can be seen that the agreement worsens significantly near the top of the detector. The correction map generation code was adapted to produce separate maps for the top and bottom arrays to facilitate this comparison. The gradients have become steeper compared to the default parameters, which is thought to be primarily a result of decreasing the Rayleigh scattering length. If photons travel shorter distances before deflection, the difference in relative LCE between the near and far PMT array will be enhanced.

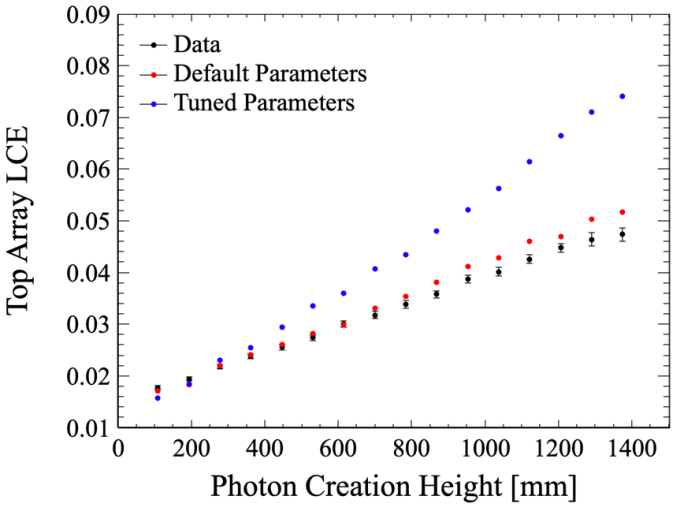
It was therefore decided that top and bottom array LCE should be considered separately in the χ^2 comparison, before summing the terms. The histograms resulting from this grid search are shown in Figure 5.17, and it can be observed that the LXe-PTFE reflectivity and the LXe absorption length remain degenerate, as expected. The most significant difference is seen in the LXe Rayleigh scattering length, which now produces a minimum χ^2_ν value when it lies between 0.40 m and 0.45 m.

Figure 5.18 shows the minimum χ^2_ν value for each two-dimensional histogram against the Rayleigh scattering length. Spline interpolation between the data points provided an optimal scattering length of 0.43 m. Photon bomb simulations were run for this scattering length, with LXe absorption lengths of up to 1000 m, and again the minimum valley was fit with an exponential function, as shown in Figure 5.19. In this instance, a PTFE reflectivity of 97.3% was chosen, which applies a 0.4% penalty to the value measured at Coimbra-LIP to account for gaps in the PTFE surface, due to pins made of polyether ether ketone (PEEK). The corresponding LXe absorption length was determined to be 121 m.

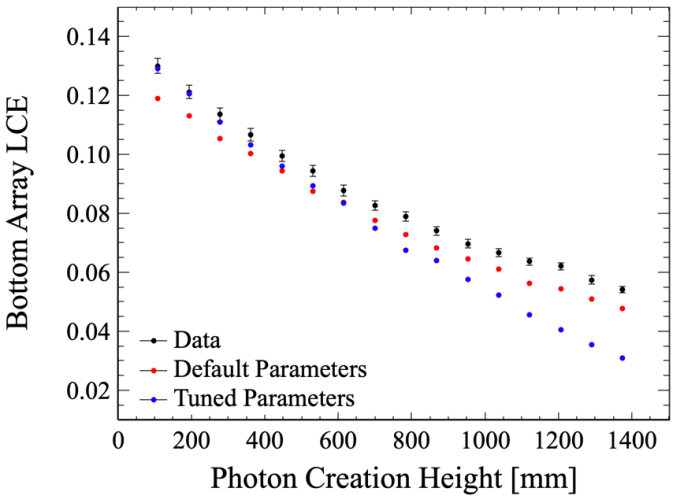
5.6.2 Results

The top array, bottom array and total LCE curves for photon bombs using the optimal set of parameters are shown in Figure 5.20, demonstrating improved results compared to the default parameters. A good agreement to data is seen in all regions of the TPC, except for the bottom array LCE for photons generated close to the bottom of the detector.

This discrepancy was explored by varying additional optical parame-



(a)



(b)

Figure 5.16: LCE curves for the top (a) and bottom (b) array only, for photon bomb BACCARAT simulations with default and tuned parameters, compared to data.

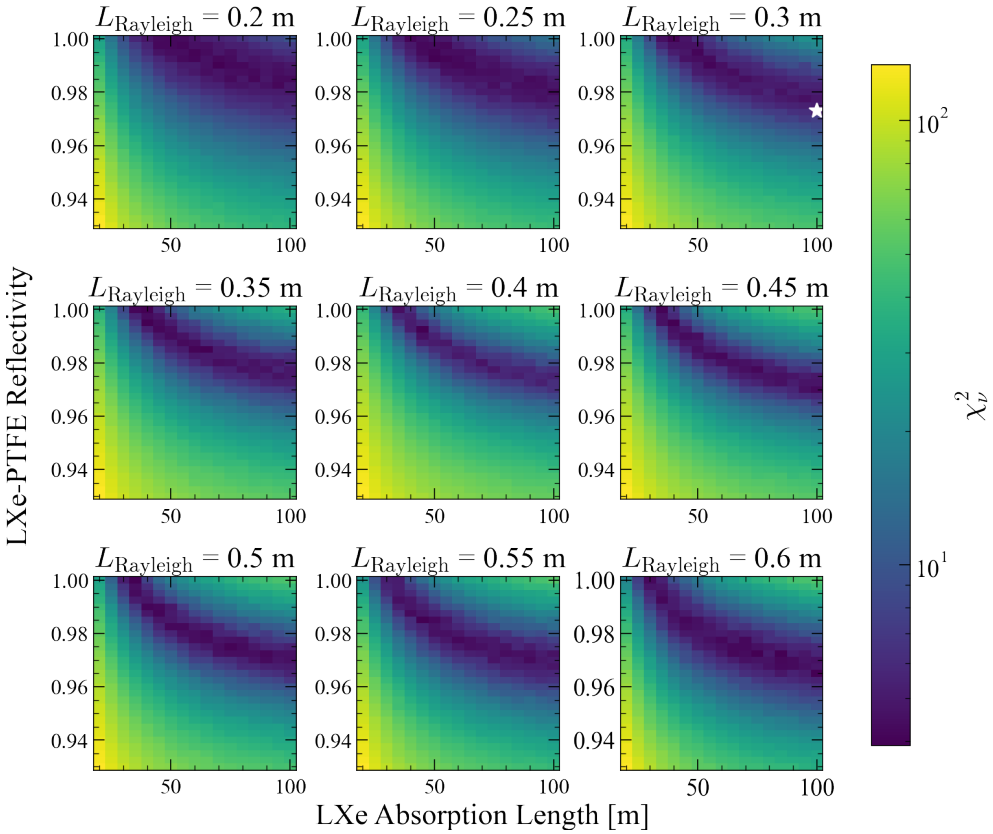


Figure 5.17: Subplots showing χ^2_ν values resulting from a comparison to data, considering top and bottom array LCE separately, for all combinations of LXe-PTFE reflectivity and LXe absorption length, for each LXe Rayleigh scattering length tested in the grid search. The default BACCARAT parameters are indicated by the white star. The parameters resulting from the previous tuning are beyond the x-axis range of these plots.

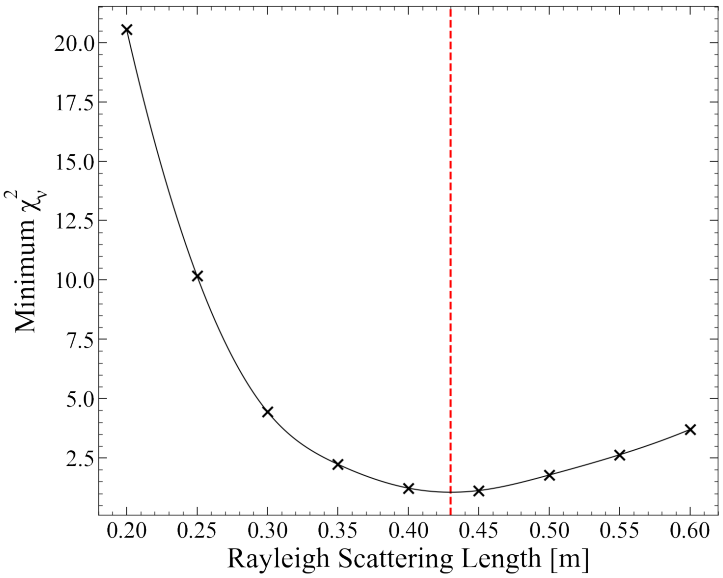


Figure 5.18: The minimum χ^2_ν value from each of the histograms in Figure 5.17, which correspond to a different LXe scattering length. The black line shows the interpolation between the points and the red dashed line indicates the position of the minimum point on this line.

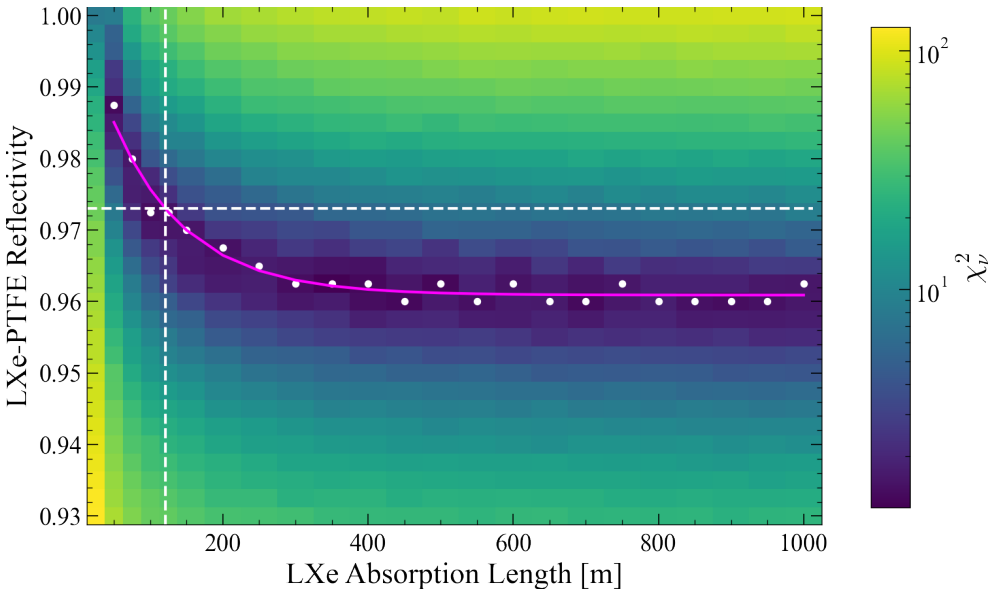


Figure 5.19: χ^2_ν values resulting from a comparison to data for a LXe Rayleigh scattering length of 0.43 m. The range of values for LXe absorption length has been extended up to 1000 m. The white data points show the PTFE reflectivity producing the minimum χ^2_ν for each value of LXe absorption length, and have been fit with an exponential function. The white dashed lines indicate a reflectivity of 97.3% and an absorption length of 121 m.

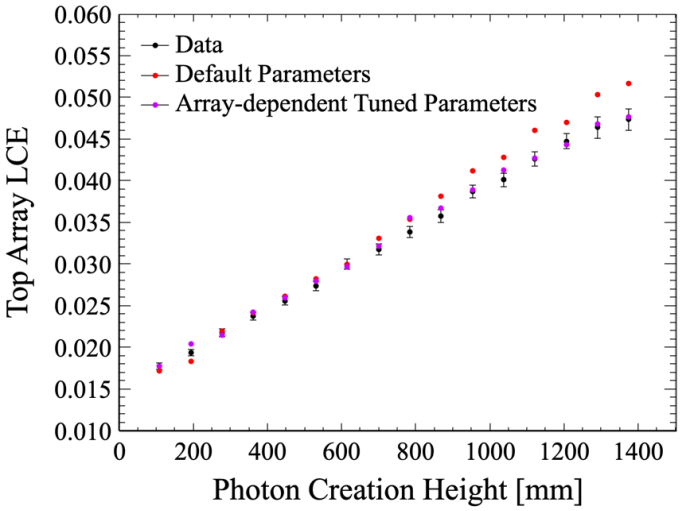
ters, however further study is required to gain a full understanding of the issue. Possible solutions, that would have a position-dependent impact, include adjustments to the refractive index of the quartz PMT windows, the refractive index of the gaseous xenon, and the reflectivity of the grids in LXe compared to GXe.

Figs. 5.21 and 5.22 show WIMP search data from the start of SR1 in S2c_{bottom}-S1c space to highlight distributions of events resulting from activation of the xenon. The lowest-energy peak corresponds to the 164 keV decay of ^{131m}Xe, while the two remaining peaks are from the 208 keV and 235 keV K- and L-shell electron capture decays of ^{127}Xe [159]. Also shown are the 1σ contours for BACCARAT simulations of these isotopes using default parameters and parameters tuned using both the top and bottom array LCE. At the time of running, the simulations had not been tuned to reproduce SE signals in data, so S2c_{bottom} values have been scaled to account for the difference in SE size and the variation in S2 pulse area with drift time. As expected, improved alignment is seen for events near the top and middle of the detector, while some discrepancy remains near the bottom of the TPC. When considering data from all vertical positions, good overall agreement is observed.

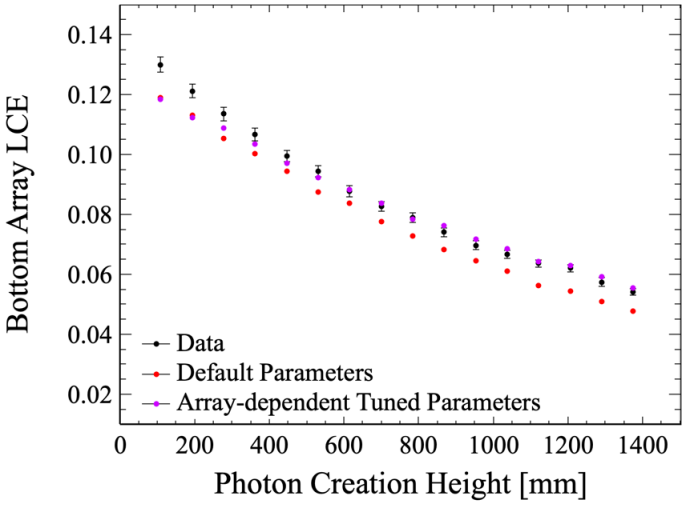
5.7 S1 Pulse Length

Investigations were carried out to determine whether varying the optical parameters would affect the length of the S1 pulse, due to altered photon transit times. Three metrics for the pulse length were considered: (AFT95 - AFT5), FWHM, and (AFT95 - pulse peak time), which describes the length of the pulse tail. These measurements make use of the Area Fraction Time (AFT) quantities, which are defined as the time at which the specified percentage of photons in a pulse have arrived at the PMT. The histograms shown in Figure 5.23 compare simulations of a ^{131m}Xe source, with default and tuned parameters, to data. The 10 ns binning corresponds to the resolution at which the DAQ samples waveforms. Good agreement is seen between simulations and data for the tail length, but the (AFT95 - AFT5) and, in particular, FWHM metrics indicate that pulses are typically shorter in the simulations. For all three metrics, the agreement with data has slightly worsened for the tuned parameters compared to the default, which motivates future analyses, as discussed in section 5.8.

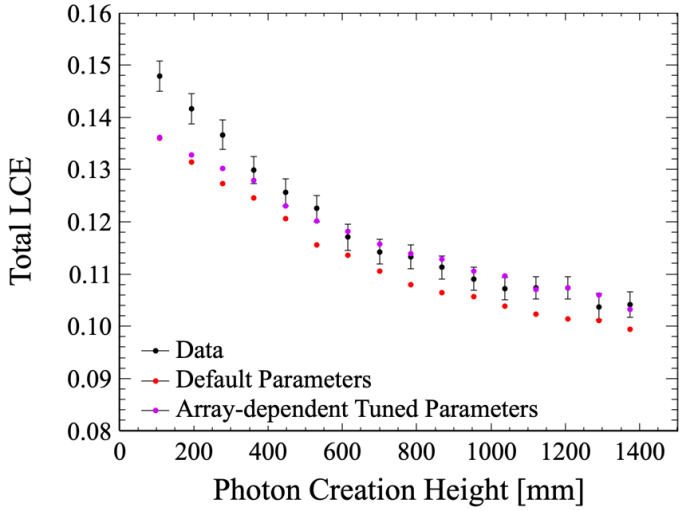
Figure 5.24 shows S1 pulses from 100 ^{131m}Xe events in both tuned simulations and data, which have been aligned with AFT25 at zero, summed



(a)

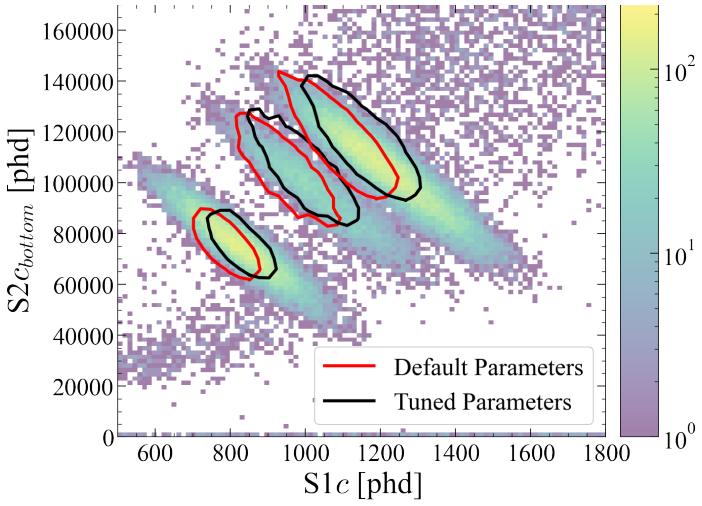


(b)

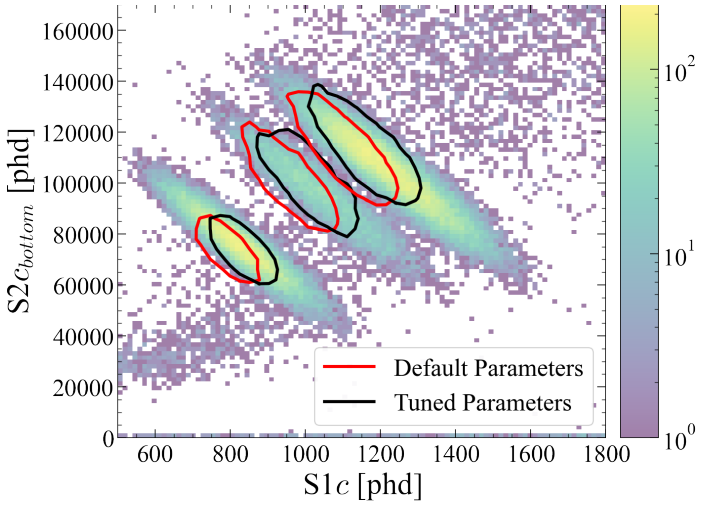


(c)

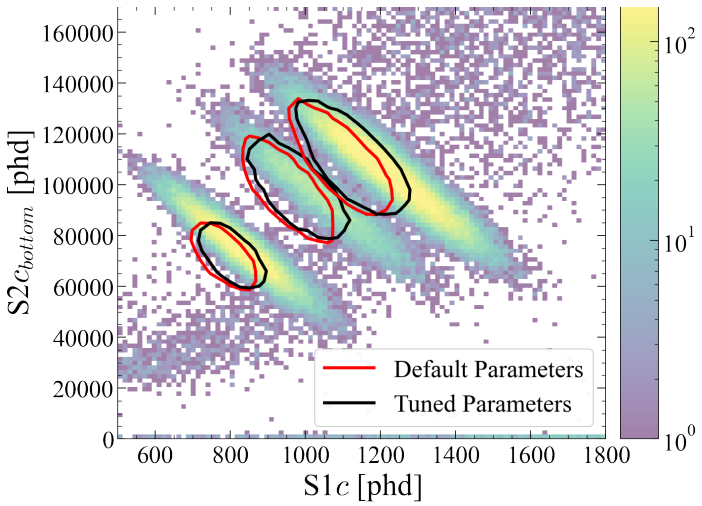
Figure 5.20: LCE curves for the top array (a), bottom array (b), and total light collected (c), for photon bomb BACCARAT simulations with default parameters and parameters tuned by considering the top and bottom array LCE separately.



(a)



(b)



(c)

Figure 5.21: Two-dimensional histograms show the distributions of ^{131m}Xe and ^{127}Xe events from early SR1 WIMP search data at the top (a), middle (b), and bottom (c) of the detector. Overlaid are 1σ contours for the same sources, simulated using BACCARAT with default parameters (red) and parameters tuned by considering the top and bottom array LCE separately (black).

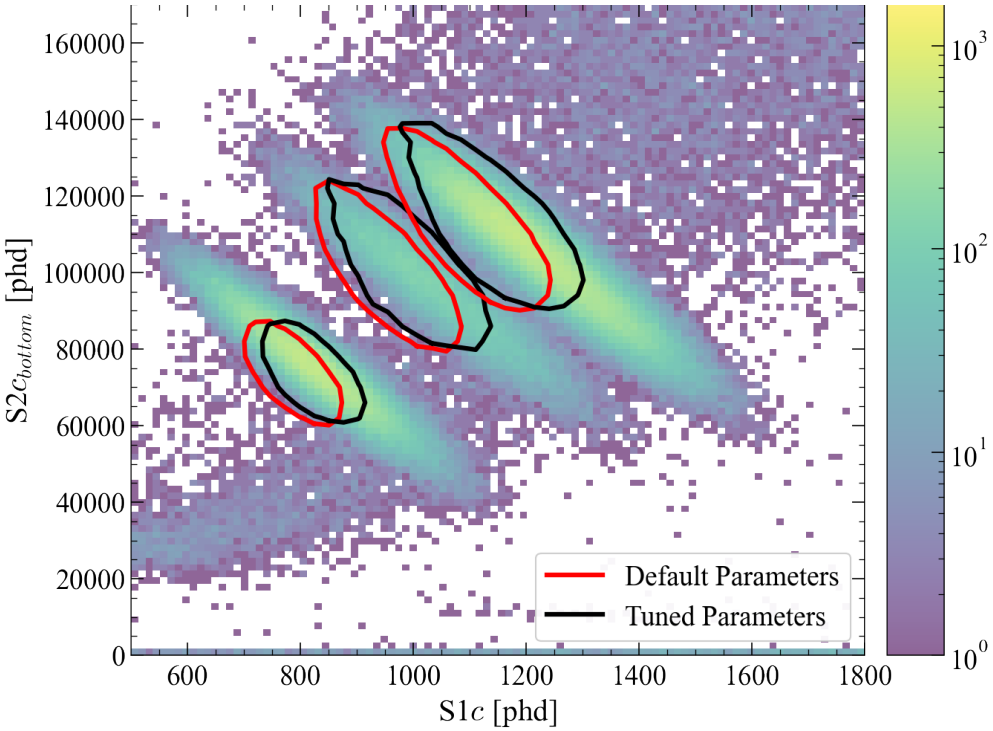
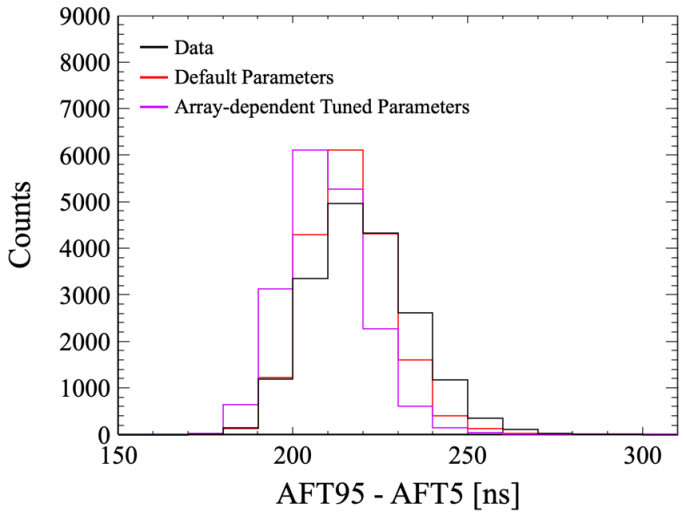


Figure 5.22: Distributions of ^{131m}Xe and ^{127}Xe events from early SR1 WIMP search data at all heights in the detector. Overlaid are 1σ contours for the same sources, simulated using BACCARAT with default parameters (red) and parameters tuned by considering the top and bottom array LCE separately (black).

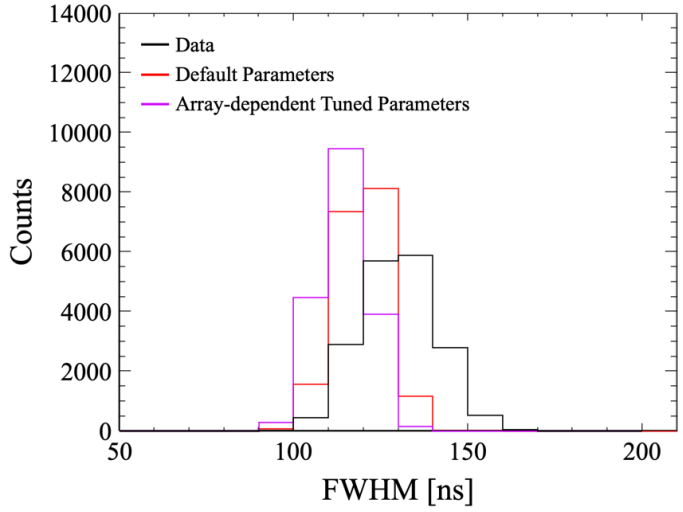
and normalised. Although there is generally good agreement, this confirms that S1 pulses are slightly narrower in the simulations, compared to in data. The Rayleigh scattering length is likely the parameter that has the most significant impact on photon arrival times, due to its effect on the overall path length.

5.8 Summary and Discussion

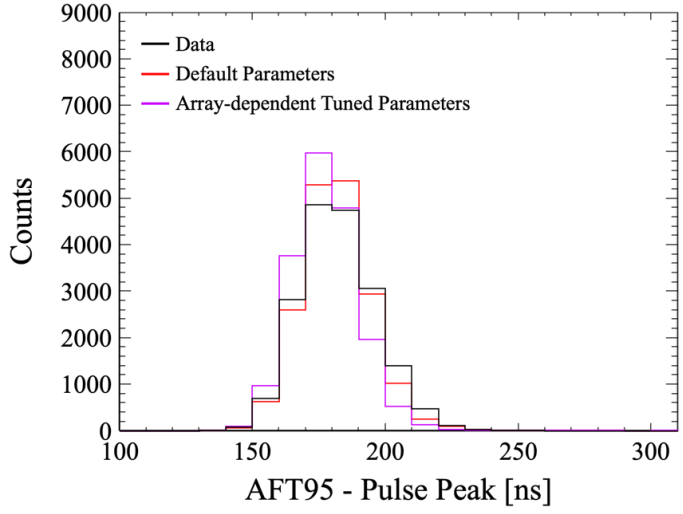
Accurate optical simulations are vital for analyses that depend on waveforms. The position-dependent LCE has been tuned to replicate the data by conducting a grid search, in which the LXe-PTFE reflectivity, LXe absorption length and LXe Rayleigh scattering length were varied. Several issues were identified, including changes to the ^{83m}Kr model and secondary PMT effects. It was found that tuning the simulations to match only the total LCE resulted in significant discrepancies in the LCE for the top and bottom PMT arrays. The grid search was therefore repeated, considering the top and bottom photon counts separately. This resulted in good agreement between the LCE in both PMT arrays throughout the detector, apart from in the region close to the bottom of the TPC. This warrants further investigation, and several possible solutions have been



(a)



(b)



(c)

Figure 5.23: Histograms showing three metrics describing the S1 pulse length: AFT95 - AFT5 (a), FWHM (b), and AFT95 - pulse peak time (c). A ^{131m}Xe source generated from simulations using default parameters and array-dependent tuned parameters is compared to data.

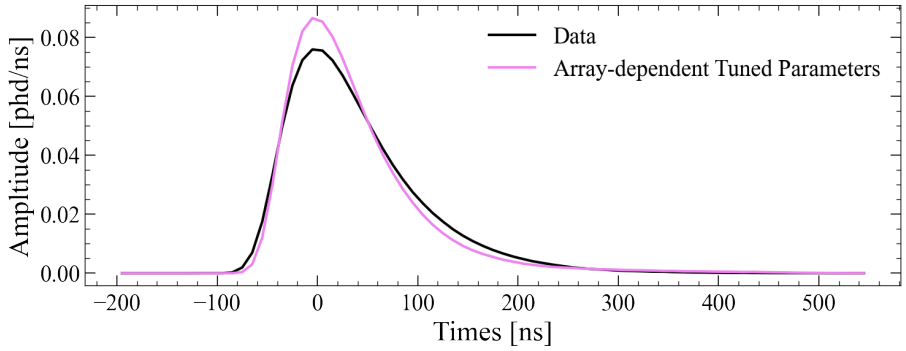


Figure 5.24: S1 pulses from 100 ^{131m}Xe events aligned at AFT25, summed and normalised, for simulations using array-dependent tuned parameters, compared to data.

suggested. Radial variation of LCE is not expected to be a large effect, and simple checks were conducted to ensure that it behaved as expected. In future, the χ^2 comparison could be extended to include XY position as well as drift time. As photons originating close to the TPC wall are more likely to undergo reflections, this is likely to break the degeneracy between LXe-PTFE reflectivity and LXe absorption length, but would require increased computation time. The S1 pulse length was also explored, with the spread in arrival times of S1 photons found to be slightly narrower in the simulations compared to data.

It would be beneficial to conduct further tuning of the simulations with photon transit time contributing additional terms to the χ^2 test. Comparisons to timing in the data are more complex than LCE, as photon creation times are unknown, prohibiting a direct calculation of the transit time. Sources such as ^{131m}Xe can be simulated, in order to compare the shape of the S1, but running the DER and LZap in addition to BACCARAT for many combinations of parameters, to allow for waveform analysis, is computationally expensive. The arrival time of the first photon in an S1 pulse cannot be determined reliably, so metrics for the width, such as those previously described, must be relied upon to determine the spread of photon detection times.

Despite the challenges described, simulations that accurately model the timing of S1 photons, as well as the LCE, are important for various analyses. In particular, simulations were used to investigate the timing offset effects that resulted in the development of the S1 likelihood cut discussed in chapter 6. This cut would further benefit from testing and tuning on fully accurate simulations, which provide photon counts and arrival times for S1 pulses.

Chapter 6

The S1 Likelihood Cut

Of the background signals described in chapter 3, one of the most significant for the WIMP search is the accidental coincidence of S1 and S2 pulses, which can be misclassified as a Single Scatter event. The focus of this chapter is the development of three data quality cuts, which use arrival times of S1 photons to target accidentals. S1 waveform templates, which depend on both the drift time and PMT array in which photons are collected, were generated using large samples of events from data. These were used to calculate the likelihood that an event is a true Single Scatter with a real S1 pulse. The methodology of the cuts, as well as the resulting signal acceptance and background rejection efficiency, will be discussed.

6.1 Photon Timing Offset

Many of the SR1 data quality cuts described in section 4.3 are designed to remove accidental-coincidence events by targeting the S1 pulse. One method for doing so is assessing whether the S1 looks as expected for the drift time assigned to the event. Features of the S1 pulse itself can also indicate whether it is likely to be a result of pile-up of SPEs or PMT dark counts, that meets the three-fold coincidence requirement. Both the former general case and the latter specific pathology can be targeted using the arrival time of photons in the S1 pulse.

An offset between hit times in the top and bottom PMT arrays has been observed in data. This effect is dependent on the drift time of an event, and is most significant near the top and bottom of the detector, where an offset of up to ~ 45 ns between pulse peaks can be seen. The offset between waveforms is demonstrated in Figure 6.1. Here, monoenergetic ^{218}Po α events were retrieved from SR1 WIMP search data by using cuts in $\log_{10}$ S2c-S1c space to select the population of events with

a decay energy of 6.115 MeV following Equation 2.6.13. The S1 waveforms from 1000 events from drift time slices of $\sim$ 55 µs near the top, middle and bottom of the detector were aligned at AFT25, the time at which 25% of the S1 photons have been detected, prior to summation and normalisation. The "lz-code" Python package (recently developed by A. Baker) was used to separate the waveforms into signals detected in the top and bottom PMT arrays. Waveforms are sampled by the DAQ in time bins of 10 ns. To achieve a better resolution on the peak time than this 10 ns sampling, an exponentially-modified Gaussian function was fit to the waveform, and the maximum was found. Only when photons originate very close to the top of the detector does the waveform observed by the top array become comparable in amplitude to that observed by the bottom array, due to reflection from the liquid surface. The top array PMTs detect light sooner than the bottom array for events with short drift times close to the top of the detector, and vice versa. The top-minus-bottom peak time offsets are shown against drift time on an event-by-event basis in Figure 6.2, for ^{218}Po events from SR1 data.

Calculating the expected time of flight for a photon traversing the height of the detector in a straight line, using the PMT-to-PMT distance of 1708.3 mm and the speed of light in LXe of 0.6c, a maximum timing offset on the order of 10 ns would be expected, not accounting for reflections. It was postulated that the discrepancy between this and the observed values of $\sim$ 45 ns may be due to Rayleigh scattering in the LXe, as a small scattering length increases the distance a photon is required to travel to reach a PMT at the far end of the detector. A timing offset was also seen in simulated data, so, to investigate this, a photon bomb simulation was run using BACCARAT with an initial LXe Rayleigh scattering length of 0.43 m, resulting from the tuning described in chapter 5. The difference in time between a photon reaching a PMT photocathode and its creation are shown in Figure 6.3 for photons originating at three different heights at the top, middle and bottom of the SR1 fiducial volume. An offset in the means of the distributions for the top and bottom PMT arrays of 29.8 ns at the bottom of the fiducial volume and 24.6 ns at the top is observed. The Rayleigh scattering length was then changed to a large value of 10 m, and the offsets reduced to the expected value of $\sim$ 10 ns. This dependence on the Rayleigh scattering length suggests that there is potential to use the timing offset as a metric for further tuning of BACCARAT optical parameters.

This timing offset can also be exploited when targeting accidental events, where the drift time is not meaningful, so the strong relationship

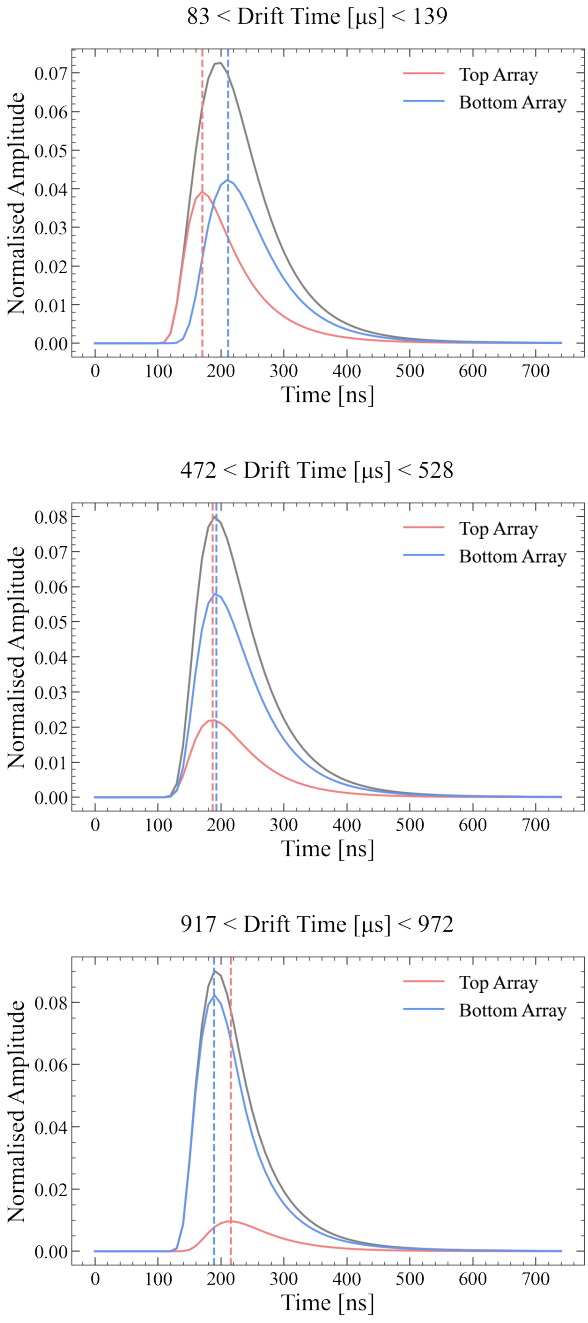


Figure 6.1: Summed and normalised S1 waveforms for ^{218}Po events in drift time slices near the top, middle and bottom of the TPC. Signals detected in the top and bottom PMT arrays are shown separately in red and blue, with their peaks marked with dashed lines to show the offset between them. The total pulse is shown in grey.

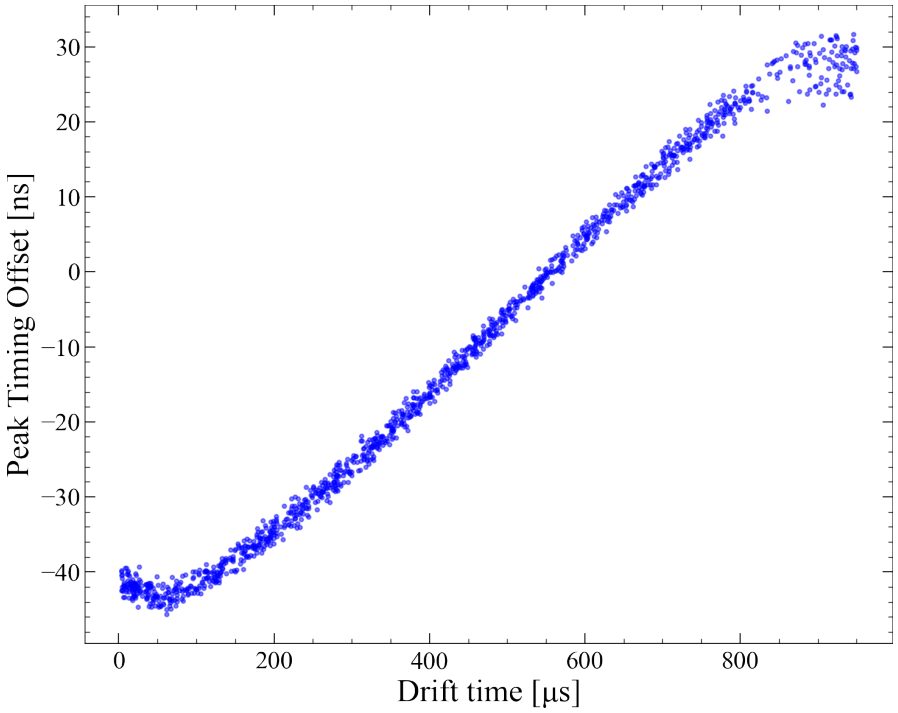
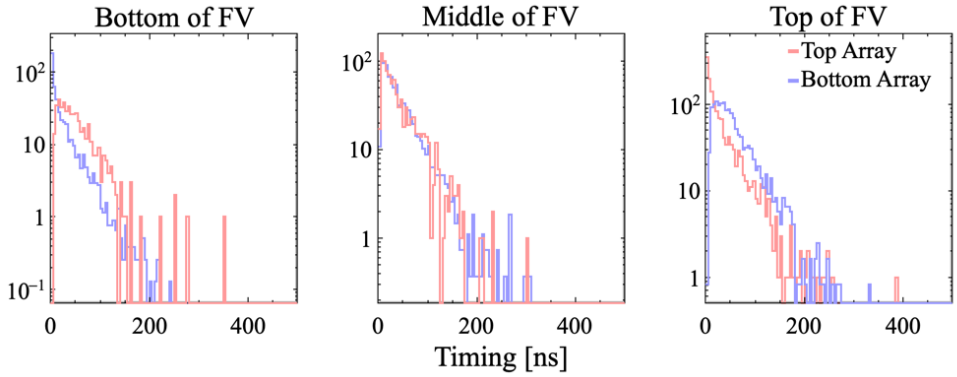


Figure 6.2: The top-minus-bottom photon timing offset for ^{218}Po S1 pulses in SR1 data.

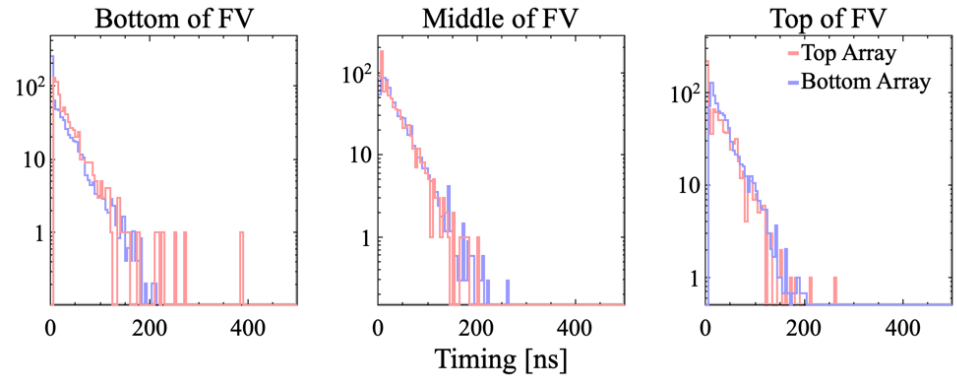
between the two variables would not be seen. Therefore, it was decided that a data quality cut, depending on both the drift time and the PMT array in which the photon is detected, would be developed. As S1-like pulses resulting from pile-up of SPEs or dark counts are often irregular in shape, the arrival times of photons within a pulse could also be used to discriminate against this specific source of isolated S1s. Therefore, several test statistics would be calculated, with the aim of using photon arrival times to determine the likelihood that an event is an accidental, each focusing on different properties and pathologies.

6.2 S1 Template Construction

To formulate a hypothesis that a pulse is a real S1 at the correct drift time, several templates were produced for the S1 waveform. The templates are entirely data-driven, using a high-energy ^{218}Po source, providing high statistics for the number of photons in each time bin. The waveforms were again aligned, summed, normalised and divided into contributions from the top and bottom arrays. The waveforms were trimmed to 2000 ns and the bottom array signal was then artificially attached to the end of top array waveform in order to contain all the information within one PDF, whilst maintaining the timing offset between the two arrays. Example templates for several drift time slices are shown in Fig-



(a)



(b)

Figure 6.3: The difference in time between photon generation and arrival at a PMT photocathode for photons simulated using BACCARAT, for photon creation heights at the bottom, middle and top of the FV. In (a), the LXe Rayleigh scattering length is set to its tuned value of 0.43 m, whereas in (b) it is set to 10 m. Counts have been normalised for easier comparison of distributions.

ure 6.4.

6.3 Likelihood Calculation

In order to evaluate whether an S1 pulse is consistent with a certain template, a likelihood fit was performed. For each event tested, photon arrival times are found using a peak-finding algorithm in LZap, which considers each PMT channel separately. A data sample larger than a trigger threshold constitutes the start of a peak. When the data value subsides below the trigger threshold, the search for the next peak begins. The algorithm also searches for subpeaks, if the peak is longer than four samples. Higher precision algorithms, for example using waveform deconvolution, would not be limited by the 10 ns sampling and could thus be a useful future development to the cut.

Fig. 6.5 shows top and bottom waveforms for an example tritium event, where the number of photons in each 10 ns time bin, as determined by the LZap algorithm, are also shown. An S1 template has been normalised by the total number of S1 photons in the tritium event and the top and bottom components are also overlaid on the plot. Although the top and bottom components are shown separately here for ease of observation, they are considered to be one distribution for the purpose of the likelihood calculation, as shown by the unphysical high times on the x-axis for the bottom waveform, which has been artificially attached to the end of the top waveform. The aim is to compare the observed number of photons in the tritium pulse to the expected number of photons according to the S1 template, in each time bin, using a likelihood calculation.

For each test statistic, the number of photons in each time bin was assumed to be Poisson distributed, allowing the likelihood to be calculated using

$$\mathcal{L}_b(\mathbf{p}) = \prod_{b=1}^{N} \frac{f_b^{n_b} e^{-f_b}}{n_b!}, \tag{6.3.1}$$

where f_b is the expected number of photons and n_b is the observed number of photons in bin b. As the time defined as the start of a pulse can be unreliable, the start time of the test pulse relative to the template, t_0, was allowed to float in order to maximise the likelihood. For two of the test statistics considered, it was desirable to repeat this process for every S1 template, in order to identify which drift time range produces the highest likelihood.

The logic of the S1 likelihood calculation is therefore as follows:

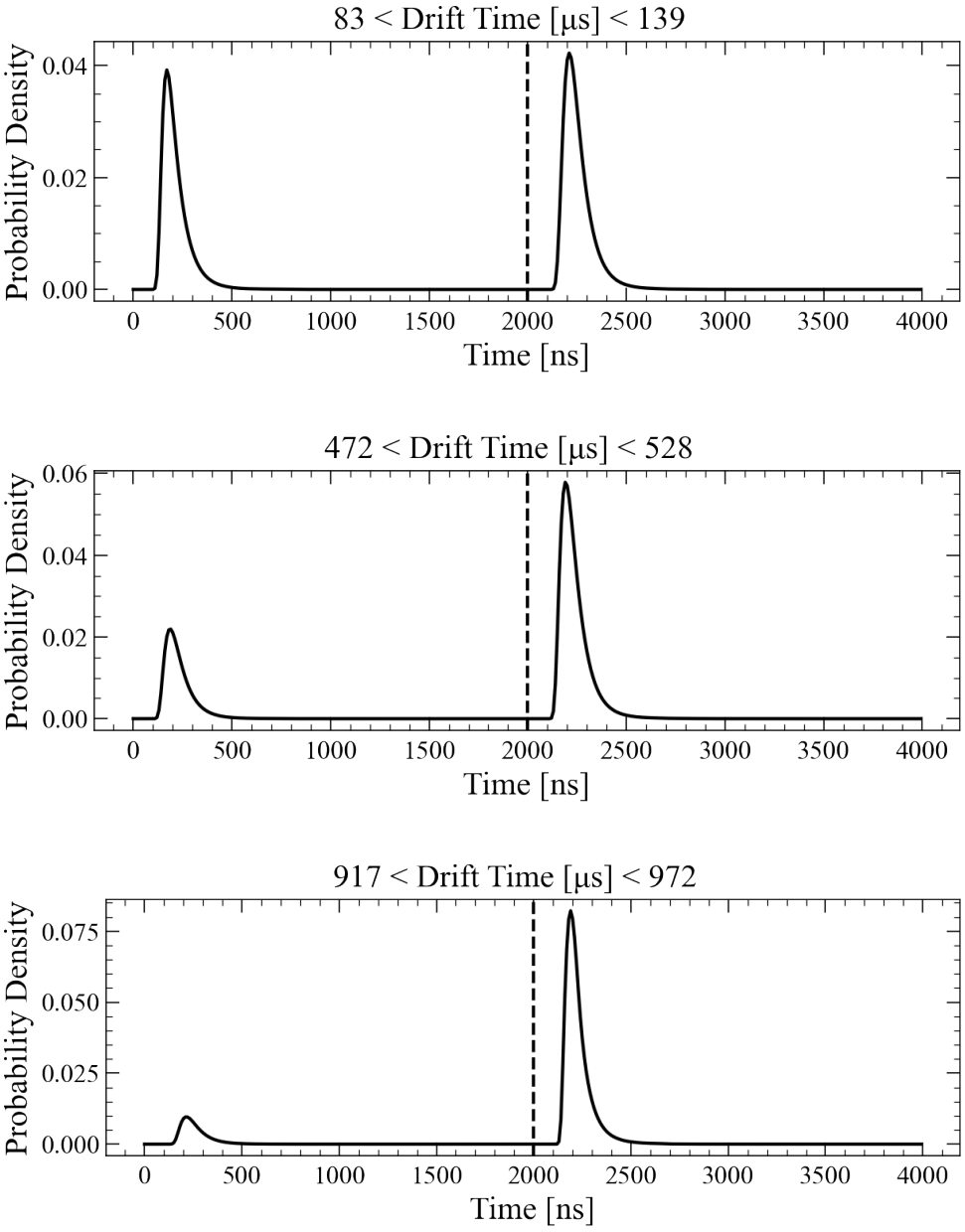
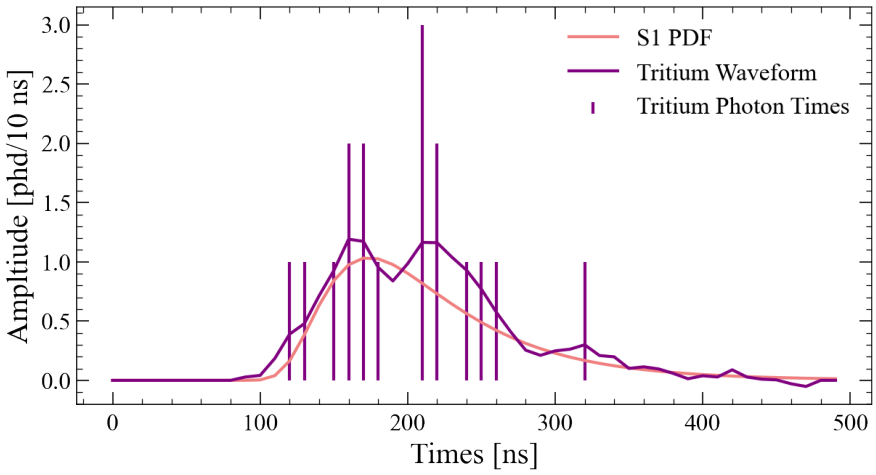
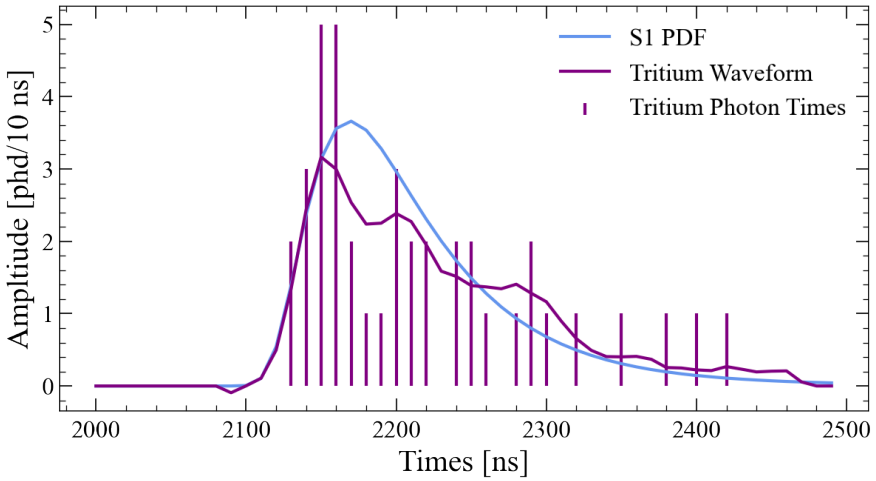


Figure 6.4: Three of the S1 templates generated using ^{218}Po events in drift time slices near the top, middle and bottom of the detector. The 0 - 2000 ns range corresponds to the top array PDF and the 2000 - 4000 ns range shows the bottom array PDF after joining the two together.



(a) Top Array



(b) Bottom Array

Figure 6.5: Top and bottom S1 waveforms for a tritium event are shown in purple, together with the number of tritium photons in each 10 ns time bin. Top and bottom S1 templates, normalised by the total number of tritium S1 photons, have been overlaid, to illustrate the comparison which is to be made between expected and observed number of photons. This information will all be contained within one PDF, with 2000 ns artificially added to the bottom template as in Figure 6.4, but the two arrays are shown separately here to provide clarity.

- Loop over templates for every drift time slice to maximise likelihood
    - For each template, loop over the offset between pulse start time and the start of the template (t_0) to maximise likelihood
        - For each t_0, loop over the 10 ns time bins, calculating each term contributing to the likelihood using Eq. (6.3.1).

Each test statistic was constructed using a ratio of likelihoods, and will now be described.

6.3.1 S1 Goodness of Fit

A general goodness of fit (GOF) test was employed to determine the likelihood that a pulse is a real S1 at the correct drift time, without the requirement for a specific alternative hypothesis. This was achieved by using the likelihood ratio as defined in Ref. [160]:

$$q = -2\ln\left[\prod_{b=1}^{M}\frac{f_b^{n_b}e^{-f_b}}{n_b!}\Big/\frac{n_b^{n_b}e^{-n_b}}{n_b!}\right] = -2\ln\left[\prod_{b=1}^{M}\left(\frac{f_b}{n_b}\right)^{n_b}e^{n_b-f_b}\right]. \quad (6.3.2)$$

The numerator of the likelihood ratio is the Poisson likelihood as described in Eq. (6.3.1), and the denominator is identical, except that the expected number of photons is replaced by the observed number of photons. The latter can be thought of as the expected likelihood if the template was a perfect match to the data:

$$q = -2\ln\left[\frac{H_0}{H_1}\right] = -2\ln\left[\frac{\mathcal{L}(\text{S1 template at correct drift time})}{\mathcal{L}(\text{“Ideal” template})}\right]. \quad (6.3.3)$$

This allows for the construction of a χ^2 statistic, which has the benefit of being agnostic to the background signal, but suffers from a resulting reduction in sensitivity. A high value of q indicates that the S1 template at the assigned drift time does not fit the data well, allowing these events to be rejected as likely accidentals.

6.3.2 S1 Template Comparison

The next test statistic makes use of the photon timing offset described, by comparing the S1 template that maximises the likelihood with the S1 template corresponding to the assigned drift time of the event:

$$q = -2\ln\left[\frac{H_0}{H_1}\right] = -2\ln\left[\frac{\mathcal{L}(\text{S1 template at assigned drift time})}{\mathcal{L}(\text{S1 template maximising likelihood})}\right]. \tag{6.3.4}$$

An incorrect drift time can result in both an unexpected timing offset between the top and bottom waveforms, and an unexpected split in the total number of photons in the top and bottom arrays. This is closely related to the existing S1 TBA vs drift time cut that was implemented in SR1, but now also takes the position-dependent timing characteristics of the event into consideration. A high value of q indicates that the assigned drift time does not provide the best match to the data, so it is likely that the S1 belongs to an accidental-coincidence event.

S1 pulses from regions outside the fiducial volume, such as the gas layer at the top of the TPC, can contribute to accidental-coincidence events. These S1s can have unusual pulse shapes. For example, S1s generated in the gas region are often wider than usual due to slow recombination in GXe. The inclusion of S1 templates outside the fiducial volume allows populations such as gas events to be included in the alternative hypothesis. If these were not included, both H_0 and H_1 would have a low likelihood for these events.

6.3.3 Shape-based Pile-up Discrimination

The final test statistic targets "fake" S1 pulses due to pile-up of SPEs by comparing the likelihood of the S1 template best matching the data with the likelihood corresponding to a flat rate of photons. As this cut is specifically targeting the population of "fake" S1s, rather than accidentals in general, it does not take the drift time into account, hence why every S1 template is tested to maximise the likelihood. It was found that for this cut, using the summed top and bottom waveforms was more effective than separating them, so the PMT array is also not taken into account here.

As the expected number of photons in each time bin must be the same, following from Equation 6.3.1 the general form of the flat rate likelihood is

$$\mathcal{L}(\mathbf{p}) \propto f^N e^{-Bf}, \tag{6.3.5}$$

where N is the total number of observed S1 photons, B is the total number of bins, and a scale factor accounting for the number of photons in a single bin is not included here. To maximise this likelihood, the

derivative is taken with respect to the expected number of photons in each bin, f:

$$f^{N-1}e^{-Bf}(N - fB) = 0, \tag{6.3.6}$$

providing a maximum at $f = N/B$. The expected number of photons for the flat rate hypothesis is therefore found by dividing the total number of photons in the S1 evenly over every time bin. The likelihood ratio is then defined as:

$$q = -2\ln\left[\frac{H_0}{H_1}\right] = -2\ln\left[\frac{\mathcal{L}(\text{S1 template maximising likelihood})}{\mathcal{L}(\text{Flat rate template})}\right]. \tag{6.3.7}$$

This allows pulses formed from pile-up, where the observed number of photons is in better agreement with the flat template than the S1 template, to be rejected based on the high value of q.

6.4 Development on SR1 Data

6.4.1 Cut Value Determination

Initial development of the cut was performed on SR1 data, using tritium calibration data as the signal sample and AccidentalChopStitch (ACS) events as the background. As explained in chapter 4, tritium is a β source that is injected directly into the LXe. The uniformity of the source in the detector and the clean waveform environment meant that tritium SS events were here selected as a proxy for the WIMP signal. ACS events are produced by artificially stitching together unpaired S1 and S2 pulses to model the distribution of the accidental-coincidence backgrounds that are targeted by this cut.

Each likelihood ratio has a different dependence on the number of photons in the S1, so fits were performed on these distributions for the tritium data, with a linear and exponential fit used for the shape-based and GOF likelihood ratios, respectively. The template comparison likelihood ratio does not have a dependence on the number of photons, because, when the assigned drift time provides the best match to data, $H_0/H_1 = 1$ exactly, for any number of photons. This is demonstrated in Figure 6.6, where it can be seen that the likelihood ratio follows the same distribution for tritium events with 5, 35 and 65 S1 photons. The parameters from the fits to tritium data were used to produce a one-dimensional distribution of each test statistic. Initially, cut values were

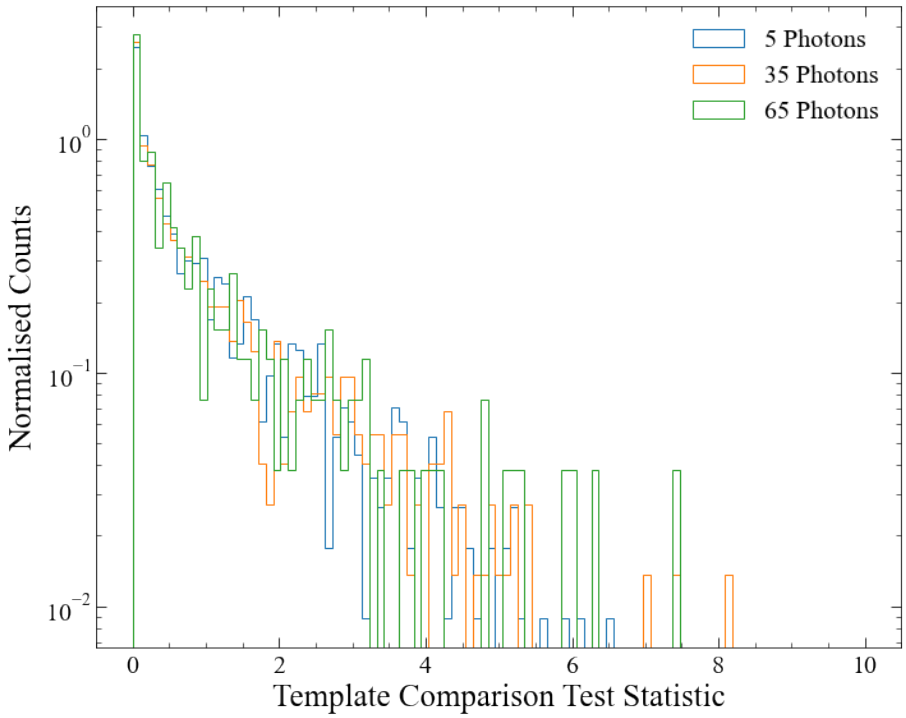


Figure 6.6: Distribution of test statistics for the template comparison likelihood ratio, for tritium events containing 5, 35 and 65 S1 photons, without other S1-based cuts applied

selected to achieve an overall 95% tritium signal acceptance for each test statistic. The background rejection efficiency was then calculated from the ACS data.

This process was carried out twice. In the first scenario, every SR1 data quality cut was applied to both datasets, apart from:

- The S1*c* lower bound, in order to maintain high statistics at low energies, where a high rate of accidentals is expected
- The S1-based accidentals cuts, which target the same events as the likelihood cuts

Then, in the second scenario, all data quality cuts, including the S1-based cuts, were applied, apart from the S1 lower bound, and the fits and cut values were redetermined. This was to demonstrate the power of the cut in addition to the existing cuts.

For the former, each likelihood ratio is shown in Figs. 6.7–6.9, first as a function of number of photons, then as a flat distribution, following normalisation by the fit to tritium data. It can be seen that a significant part of the ACS sample does not follow the functions defined by the tritium and can therefore be removed using a linear cut. In each case,

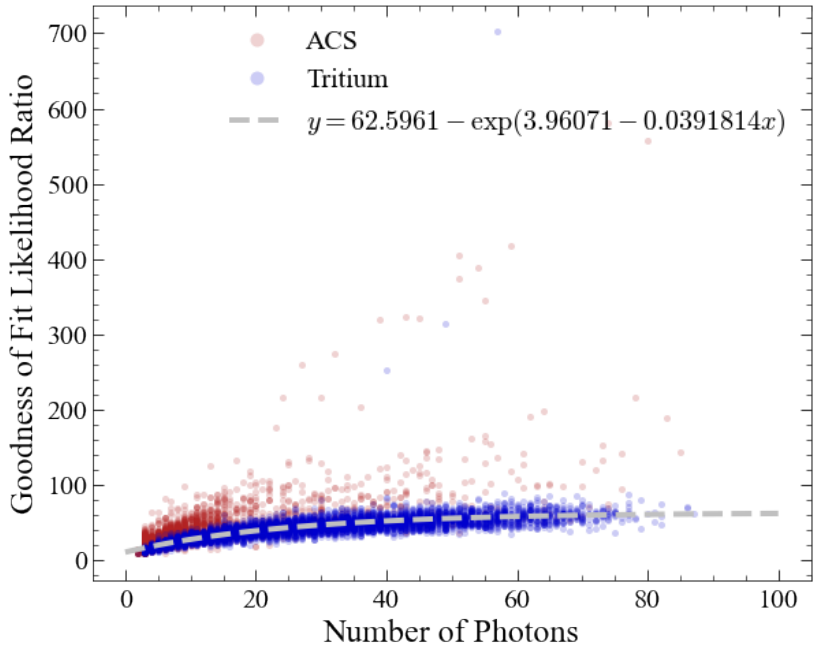
the cut value leading to a signal acceptance of 95% has been noted, and results in an overall ACS rejection efficiency of 33.7%, 33.1% and 29.9% for the GOF, template comparison and shape-based cuts, respectively. The lower efficiency of the shape-based cut is a result of true S1 pulses in the ACS sample, meaning that it is not purely a background sample for this cut. Testing on a photon-train dominated population, where pile-up is common, yielded an improved efficiency. As the aim was to determine the combined effect of all three cuts on the entire accidentals background, the ACS dataset will be the denominator used for efficiencies quoted throughout this investigation.

The one-dimensional test statistic distributions for the second scenario, following application of all other cuts, are shown in Figure 6.10. Fewer outlying ACS events can now be seen, due to the effectiveness of existing S1-based cuts. Nevertheless, a background rejection efficiency of 10.9%, 12.5% and 26.1% for the GOF, template comparison and shape-based cuts, respectively, was achieved, demonstrating that the shape-based cut is most powerful for removing backgrounds not targeted by other cuts.

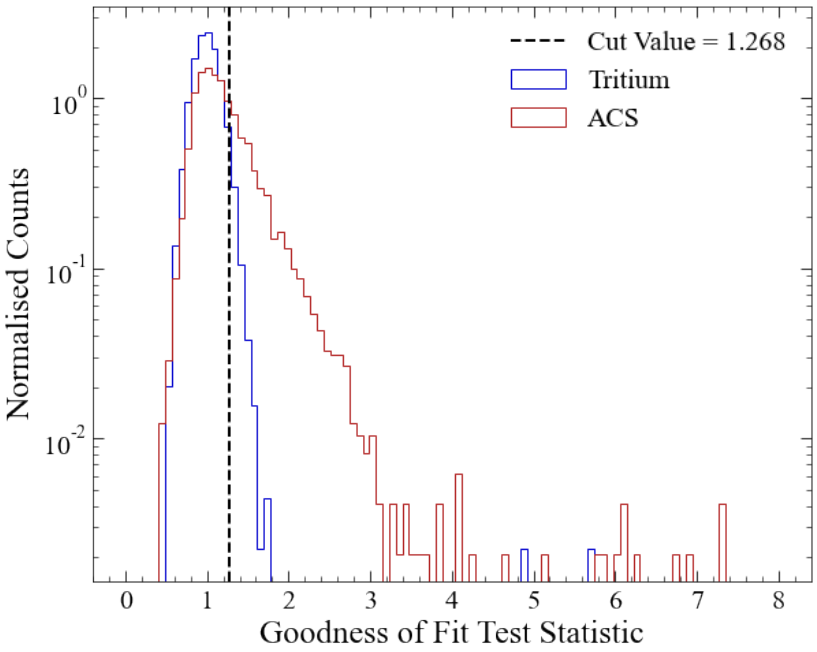
6.4.2 Dependence on Pulse Area

There was some concern that there may be a dependence of the shape of the S1 templates on pulse area. This was investigated using an AmLi source. As discussed in chapter 4, AmLi is a neutron source that was deployed in three calibration tubes at three heights, producing a wider spatial distribution than DD calibrations. As demonstrated in subsection 4.5.1, the majority of the livetime impact cuts could not be applied to AmLi data, due to the high rate of events across the ROI. This resulted in a waveform environment that is less clean than for tritium. S1 pulses from an AmLi source were binned in ranges of 20 phd, and the resulting summed waveforms were compared. There were no significant differences observed, apart from for the lowest bin (0 - 20 phd), which is shown in Figure 6.11 compared to pulses in the 60 - 80 phd range. Similar behaviour was observed for a tritium source.

It was found that good alignment between pulses was difficult to achieve for the lowest energy bin (see Figure 6.12), where the low number of photons leads to separated peaks, rather than a smooth distribution with a consistent AFT25. This led to a preference for the ^{218}Po source for the templates, which has a high number of photons in the S1 pulse. Note that this did not prevent the cut from being applied on low-energy events, as the treatment of t_0 as a free parameter handles any alignment

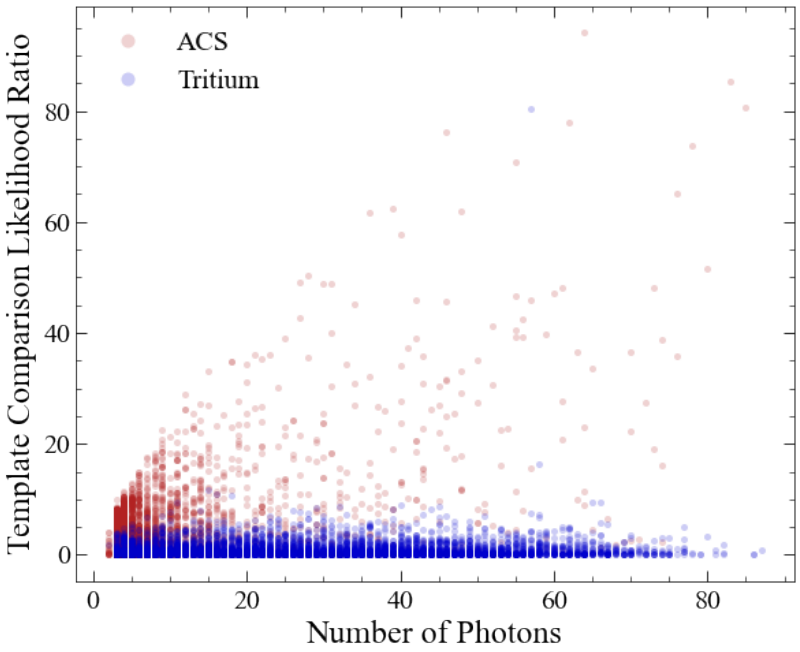


(a)

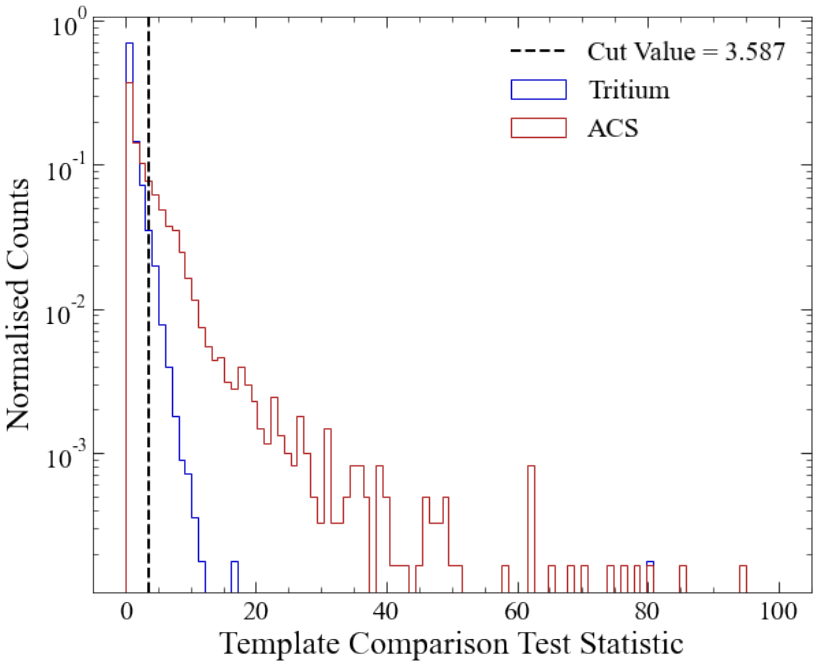


(b)

Figure 6.7: Distribution of test statistics for the GOF likelihood ratio, for tritium and ACS signals without other S1-based cuts applied, against number of S1 photons in (a), and following normalisation by the exponential fit to tritium data in (b). The cut value corresponding to 95% signal acceptance is shown.

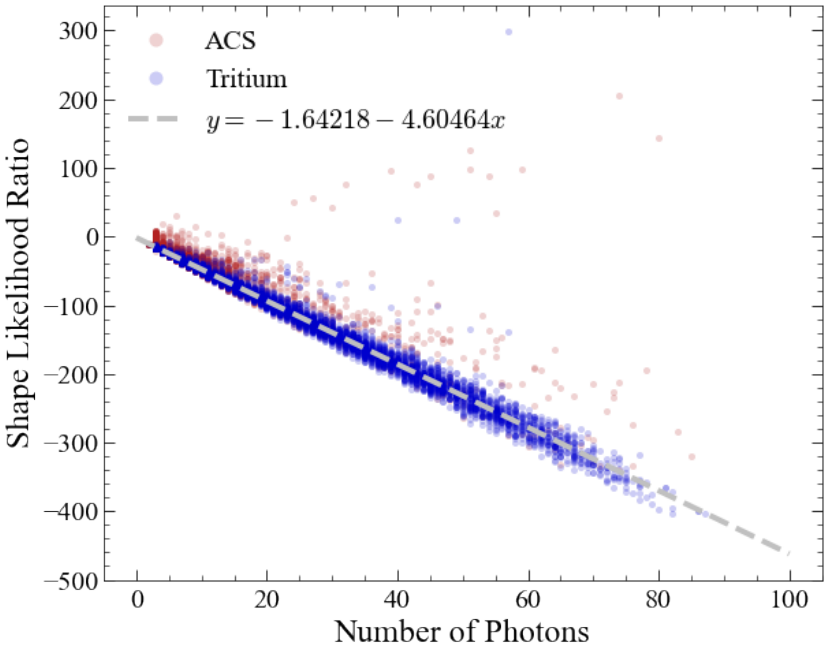


(a)

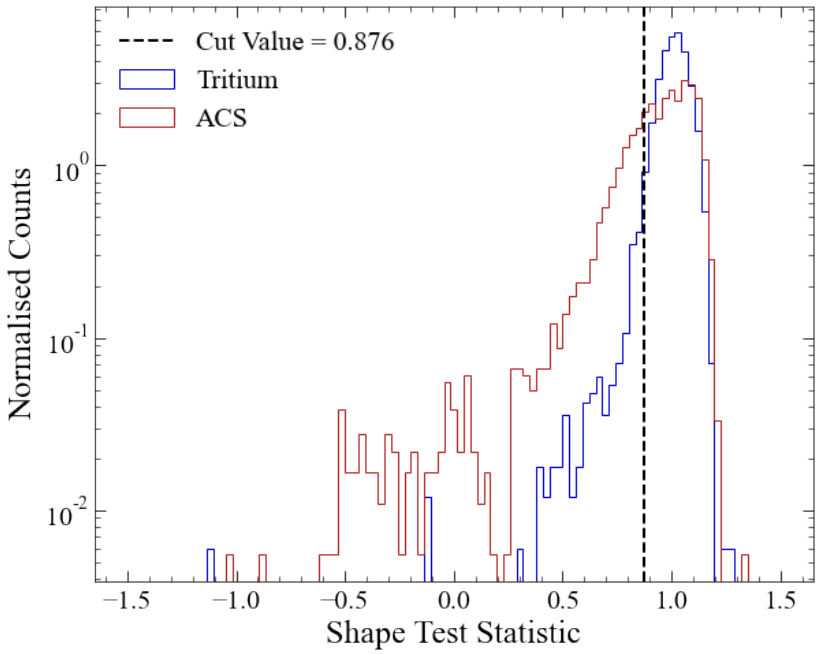


(b)

Figure 6.8: Distribution of test statistics for the template comparison likelihood ratio, for tritium and ACS signals without other S1-based cuts applied, against number of S1 photons in (a), and as a one-dimensional distribution, not requiring normalisation by a fit, in (b). The cut value corresponding to 95% signal acceptance is shown.



(a)



(b)

Figure 6.9: Distribution of test statistics for the shape-based likelihood ratio, for tritium and ACS signals without other S1-based cuts applied, against number of S1 photons in (a), and following normalisation by the linear fit to tritium data in (b). The cut value corresponding to 95% signal acceptance is shown.

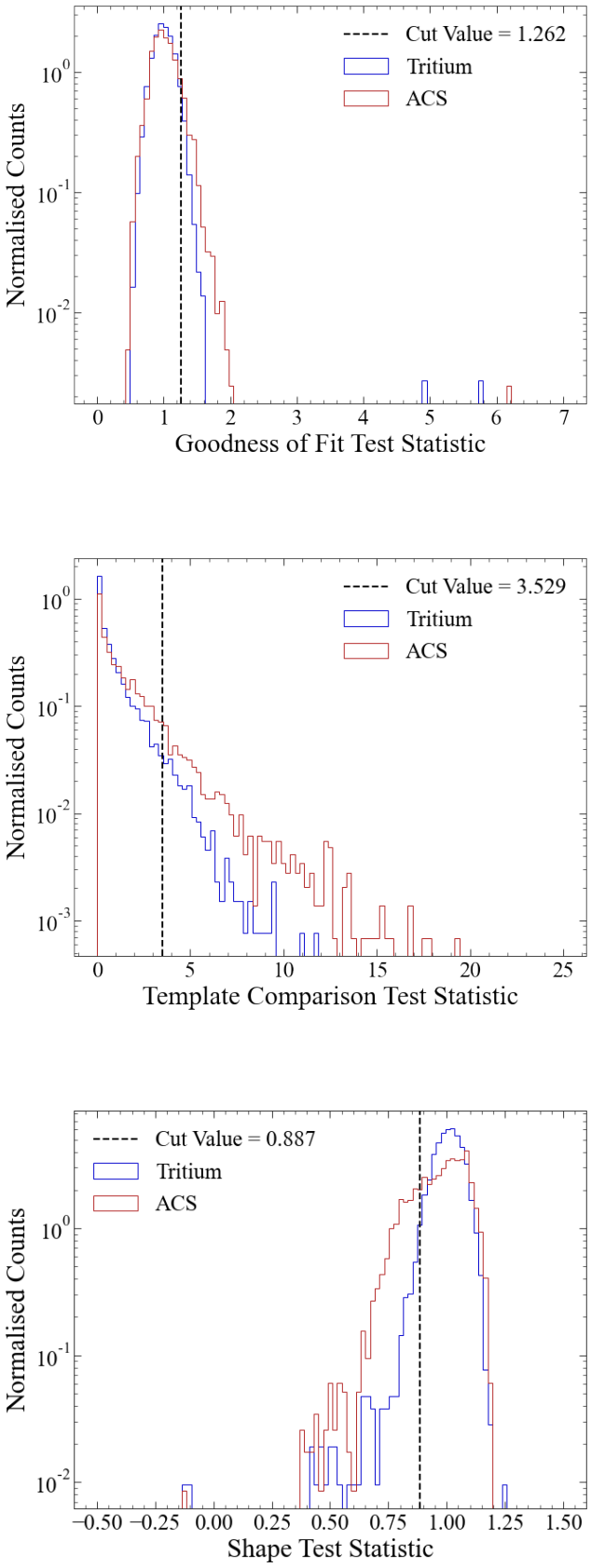


Figure 6.10: One-dimensional distributions of GOF, template comparison and shape-based test statistics, for tritium and ACS events with all other data quality cuts applied, following normalisation by tritium fits.

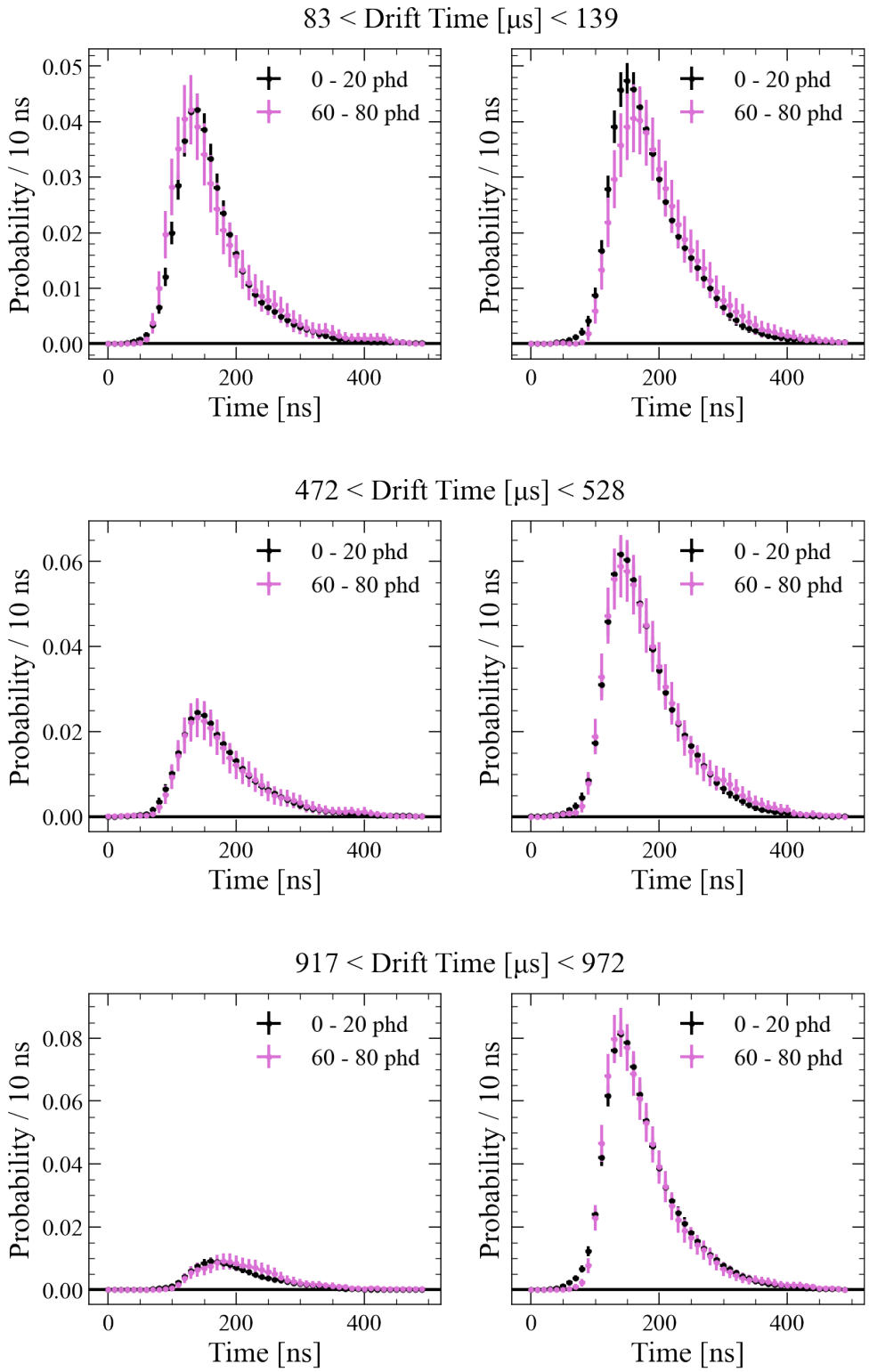


Figure 6.11: S1 templates generated using AmLi events in the ranges of 0 - 20 phd and 60 - 80 phd, in drift time slices near the top, middle and bottom of the detector for the top array (left) and bottom array (right).

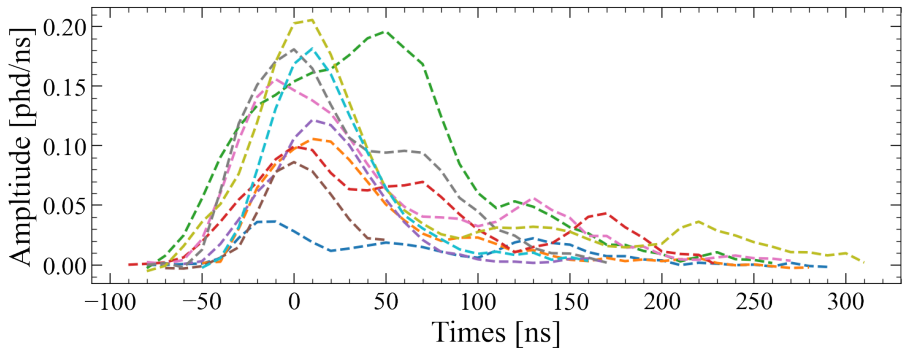


Figure 6.12: Each dashed line is taken from a different AmLi event, and shows the waveform for an S1 pulse with area less than 10 phd, aligned with the AFT25 value at $t = 0$.

issues.

6.4.3 Dependence on Interaction Type

Another concern was that using an α source to create the S1 templates may cause an issue for detecting the NR WIMP signal. As discussed in chapter 2, during the scintillation process, the molecular Xe excimer decays with both a fast and slow component, corresponding to singlet and triplet states, the ratio of which depends on interaction type [104]. Pulse shape discrimination uses this effect to discriminate between ER and NR events. This is not the intended purpose of the S1 Likelihood cut, which distinguishes between real NR or ER events and accidentals. Therefore, it was important to quantify the significance of the interaction type of the events used to produce the templates.

Tritium, AmLi and ^{218}Po sources were used to produce templates in each drift time slice. For tritium and AmLi, only pulses with an area greater than 20 phd were selected for the templates, due to the alignment issues discussed. Examples from near the top, middle and bottom of the detector are shown in Figure 6.13. Some differences in the pulse shape can be observed, for example, near the top of the detector, the NR AmLi source produces a narrower peak than the tritium and ^{218}Po sources, which is in agreement with results from Ref. [104].

To quantify the effect of these differences on the likelihood, combinations of ^{218}Po and AmLi templates and tritium and AmLi signal sources were tested. Care was taken to ensure no overlap in the AmLi samples used as signal sources and in template creation. The likelihood ratio distributions against number photons are compared for each statistical test in Figs. 6.14–6.16, with the fit parameters also shown. All parameters are

either in agreement within the fit uncertainty, or very close, apart from the y-intercept of the linear fits in Figure 6.16, which are consistently higher for the AmLi source compared to tritium.

When following the entire methodology to produce a cut efficiency for a signal acceptance of 95%, it was found that there was very little difference between using AmLi or ^{218}Po templates, with $\leq 1.3\%$ change in efficiency for all three cuts. However, the choice of tritium or AmLi as the source on which to perform the fit and tuning led to significant differences in cut efficiency. The tritium performed better, with a smaller standard deviation of the one-dimensional distributions of test statistics. This can be attributed to the fact that the AmLi dataset contains a higher number of events with a small number of photons, where it becomes more difficult to distinguish between signal and background. Note that this did not cause an issue when using AmLi to produce the templates, as only events with an area larger than 20 phd were used, to mitigate the alignment issues described previously. It was therefore decided that the use of ^{218}Po templates and tritium signal sources would continue, due to the high number of photons, and even coverage of the full S1c range, respectively.

6.5 Application on SR3 Data

6.5.1 Cut Value Determination

Following development of the cut on SR1 data, SR3 datasets became available to use for testing and tuning. The ^{218}Po selection criteria in $\log_{10}$ S2c-S1c space were adjusted to allow for the new detector conditions, and new templates were generated. The drift time slices remained 55 µs in size, but two additional templates were included to allow for the higher maximum drift time resulting from the lower drift field. The S1 templates were also trimmed to 750 ns each for the top and bottom PMT arrays, as this was found to contain the pulse lengths of the majority of legitimate S1 pulses, without including excess baseline. This had the impact of altering the range of shape-based test statistic values, due to the dependence on the total number of bins, B, in the flat rate likelihood in Eq. (6.3.5).

The relationships between each likelihood ratio and the number of S1 photons were re-generated using data from a tritium calibration under SR3 conditions. A set of baseline cuts were applied to select the tritium events, while other SR3 core cuts underwent development. Fit parameters were recalculated and the likelihood ratio values were again

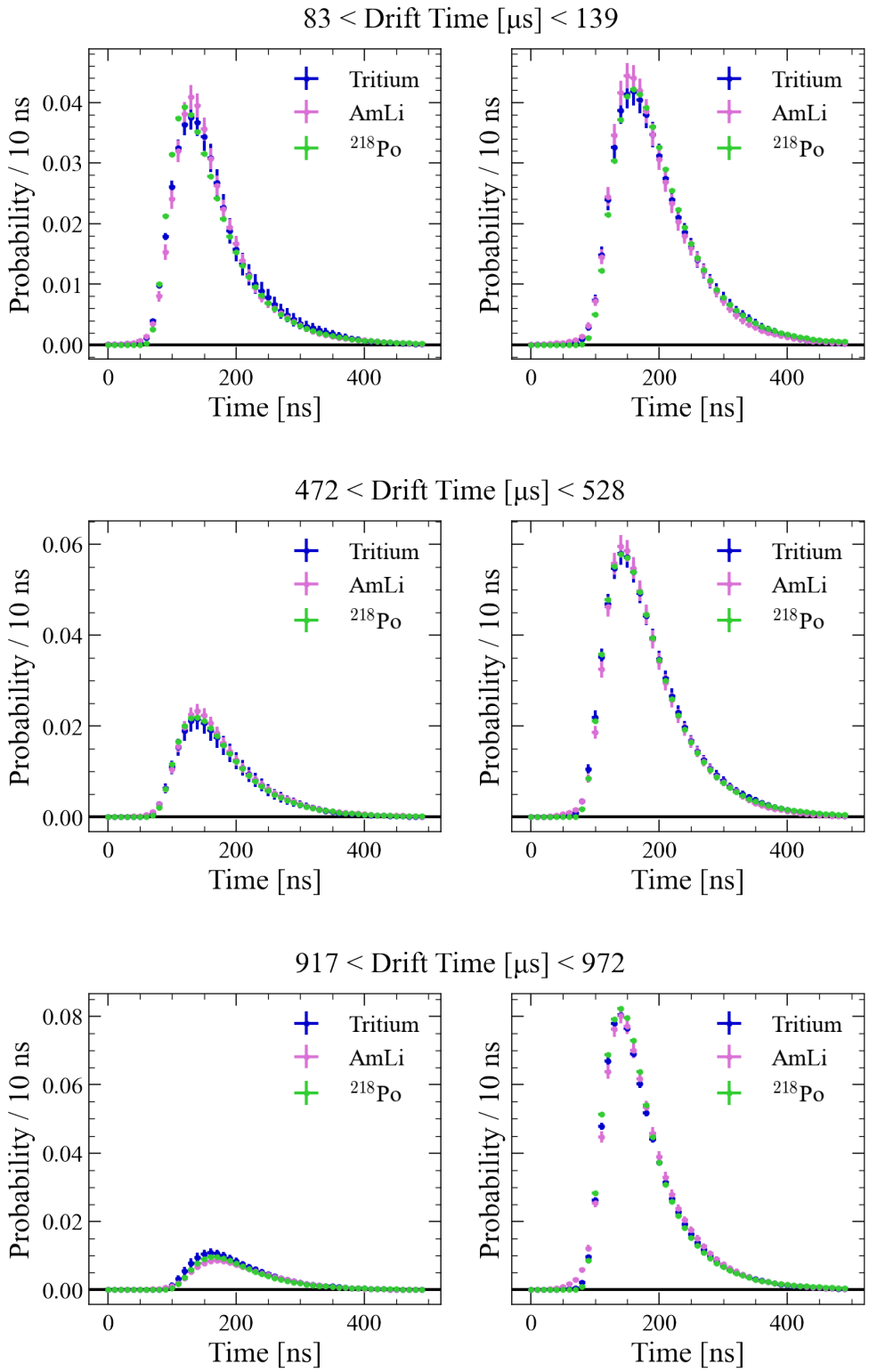


Figure 6.13: S1 templates made using tritium, AmLi and ^{218}Po events from drift time slices near the top, middle and bottom of the detector, for signals in the top array (left) and bottom array (right).

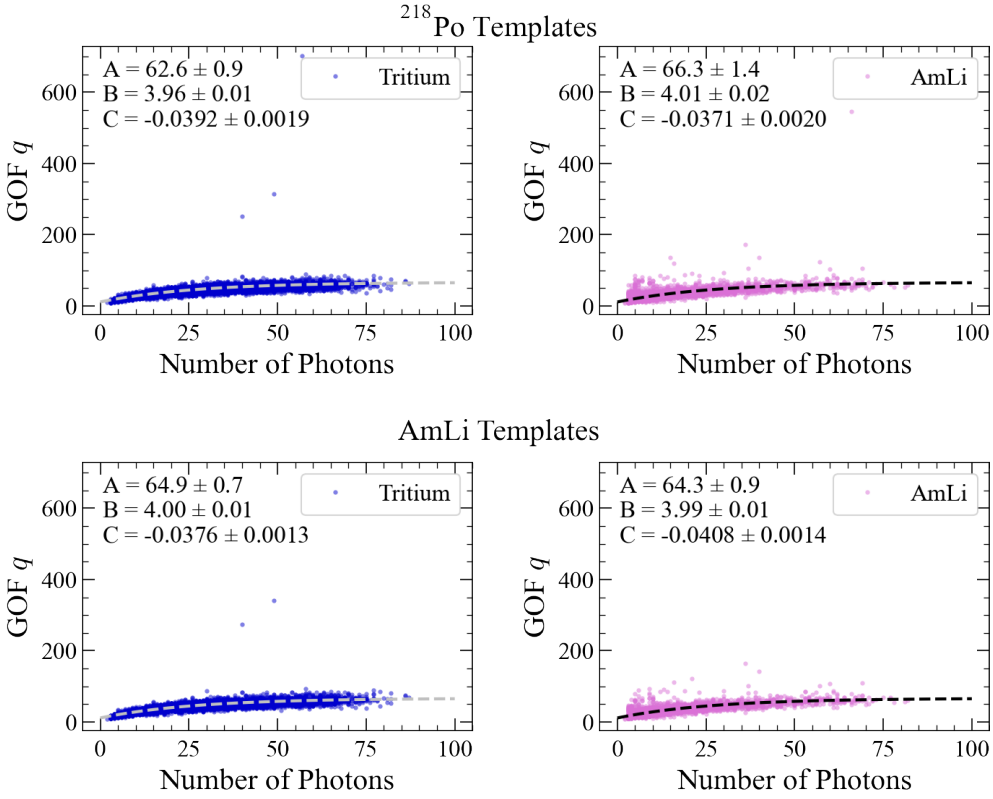


Figure 6.14: Relationship between the GOF likelihood ratio and the number of S1 photons for a tritium source (left) and an AmLi source (right) using ^{218}Po templates (top) and AmLi templates (bottom). The fit parameters for the function $f(x) = A - \exp(B + Cx)$ are shown.

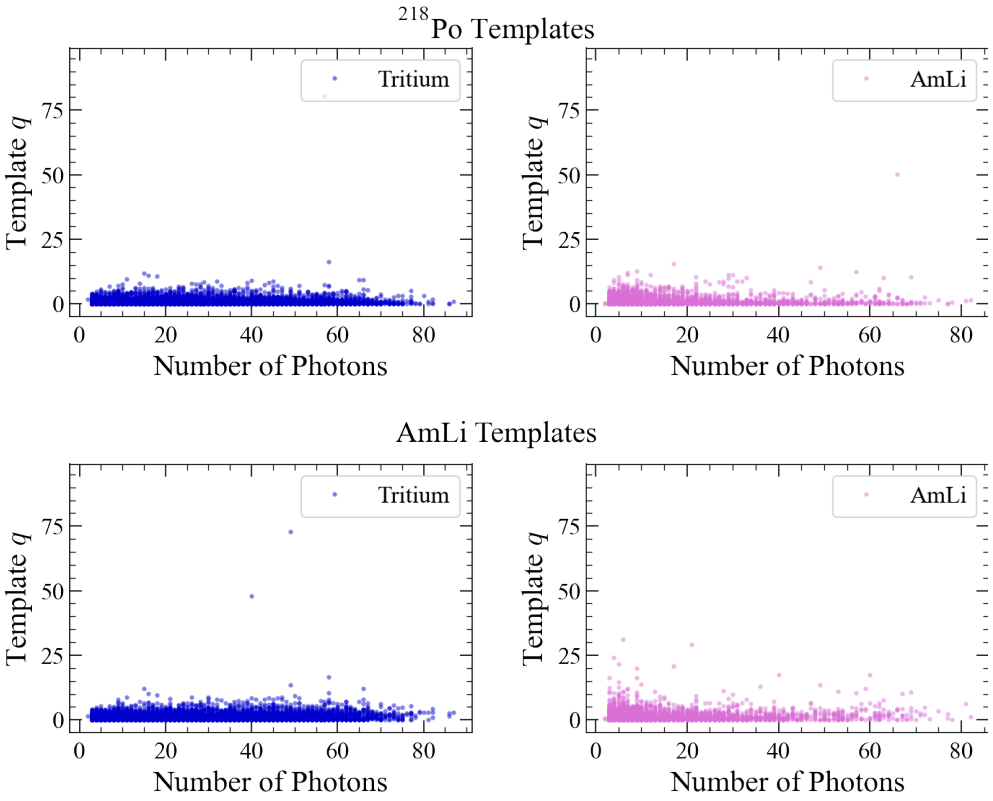


Figure 6.15: Relationship between the template comparison likelihood ratio and the number of S1 photons for a tritium source (left) and an AmLi source (right) using ^{218}Po templates (top) and AmLi templates (bottom).

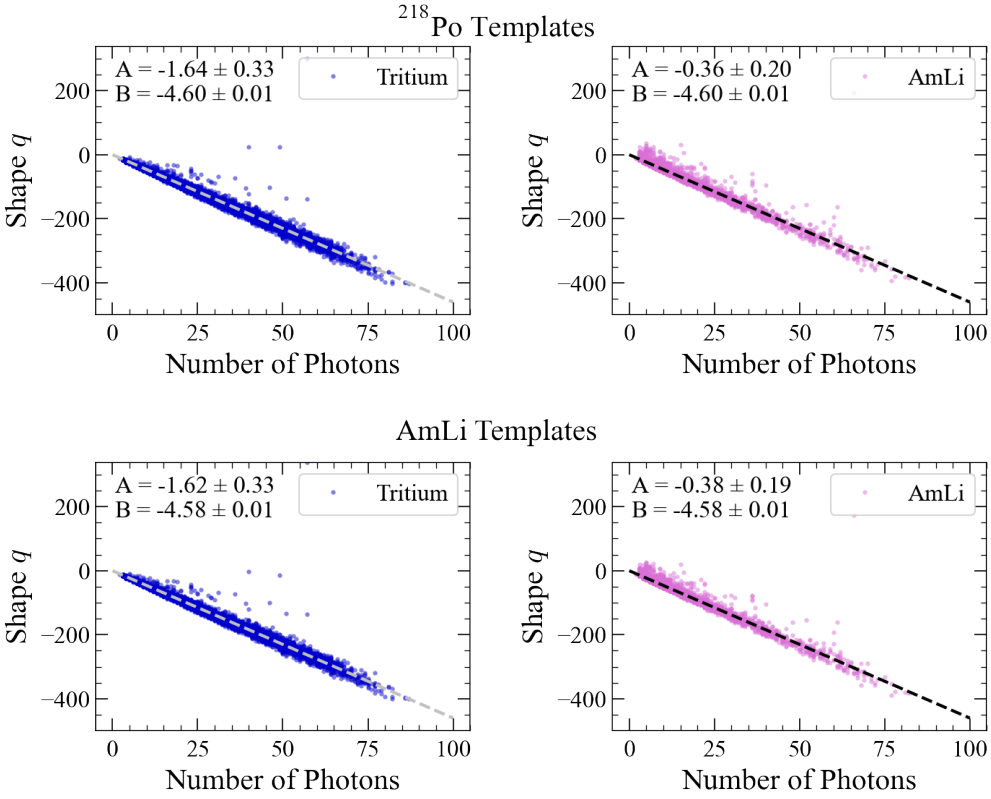


Figure 6.16: Relationship between the shape-based likelihood ratio and the number of S1 photons for a tritium source (left) and an AmLi source (right) using ^{218}Po templates (top) and AmLi templates (bottom). The fit parameters for the function $f(x) = A + Bx$ are shown.

normalised to provide a one-dimensional distribution of test statistics. The cut values were reselected to again achieve a signal acceptance of 95%. The results of this process are shown in Figs. 6.17–6.19, with background rejection efficiencies of 25.4%, 25.7% and 28.1% for the GOF, template comparison and shape-based cuts. Following further tuning of other cuts, the individual power of the cut was tested. The S1 likelihood cuts were tuned on a tritium dataset with all other cuts applied, and the resulting distributions are shown in Figure 6.20, with accidental rejection efficiencies of 18.2%, 21.2% and 26.4% for the GOF, template comparison and shape-based cuts.

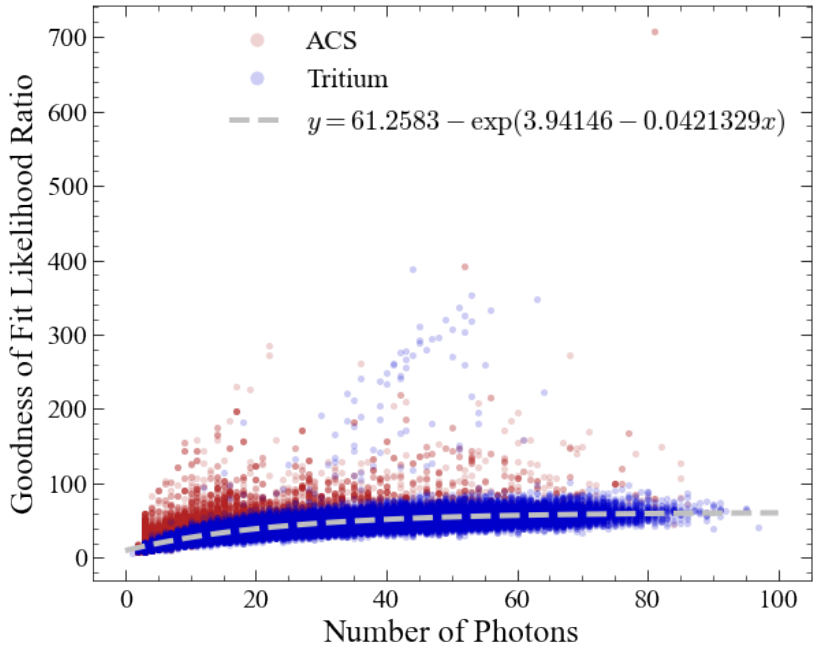
6.5.2 Cut Efficiency

Acceptance curves were generated for each cut, for tritium and ACS datasets. For the samples with only baseline cuts applied (Figure 6.21), the combined cut acceptance for the tritium signal was found to be 87.4%, which resulted in a background rejection efficiency of 47.1%. As expected from the results when using low-energy AmLi as a signal source, there is a dip in signal acceptance at low S1c, predominantly resulting from the shape-based cut. When all other data quality cuts had been applied to the samples, including S1-based cuts which target accidentals, the combined cut acceptance of 87.6% had an associated background rejection efficiency of 42.4% (Figure 6.22). The signal acceptance and background rejection efficiency of each likelihood cut is presented in Table 7.2 for each of the samples described. As ACS events were removed across the entire range of S1c tested, an investigation was carried out, using the Offline Event Viewer, into which events were targeted by each likelihood cut that were not removed by any other cut.

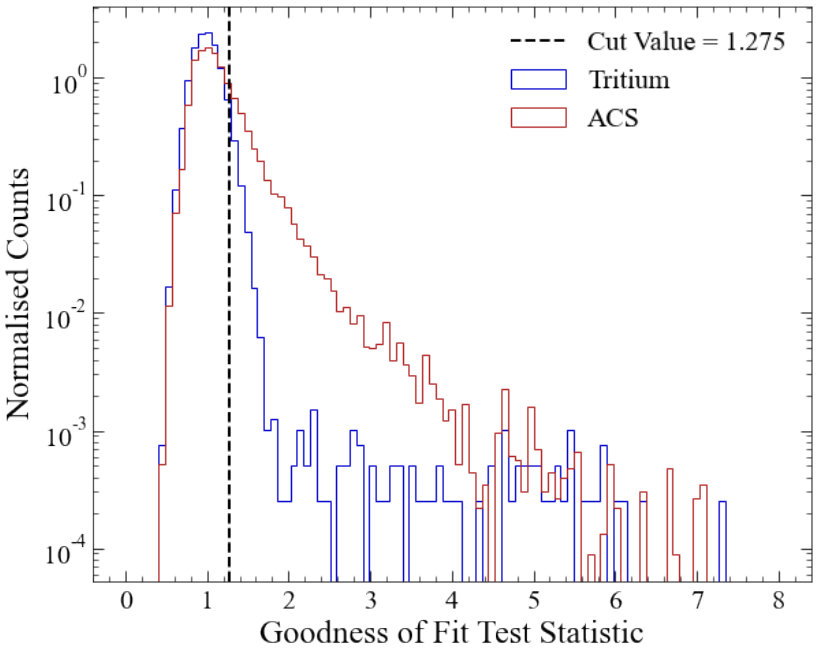
6.5.3 Pathologies Removed

The template comparison cut removes events with an unexpected top-bottom area split or timing offset for the assigned drift time. The example in Figure 6.23 shows an event with a short drift time of 88 µs, where the bottom array detects photons earlier, and in a greater quantity, than expected for an event close to the top of the detector.

The shape-based cut was found to predominantly target small S1s, such as the example pulse shown in Figure 6.24, which has an area of 2.4 phd. The fact that these pulses are often surrounded by a high rate of SPEs supports the idea that they are likely due to SPE pile-up. Larger pulses are also removed by this cut, where the unusual shape can be

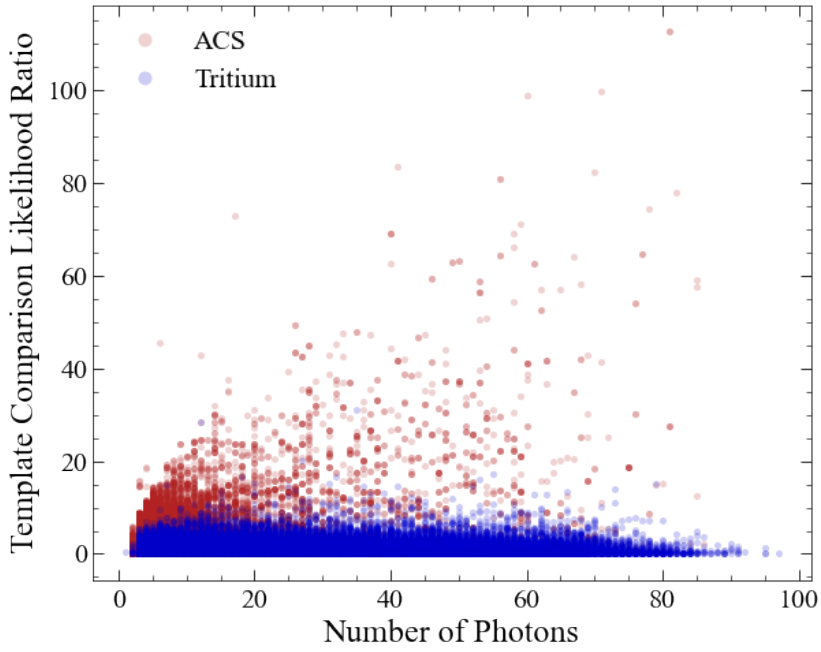


(a)

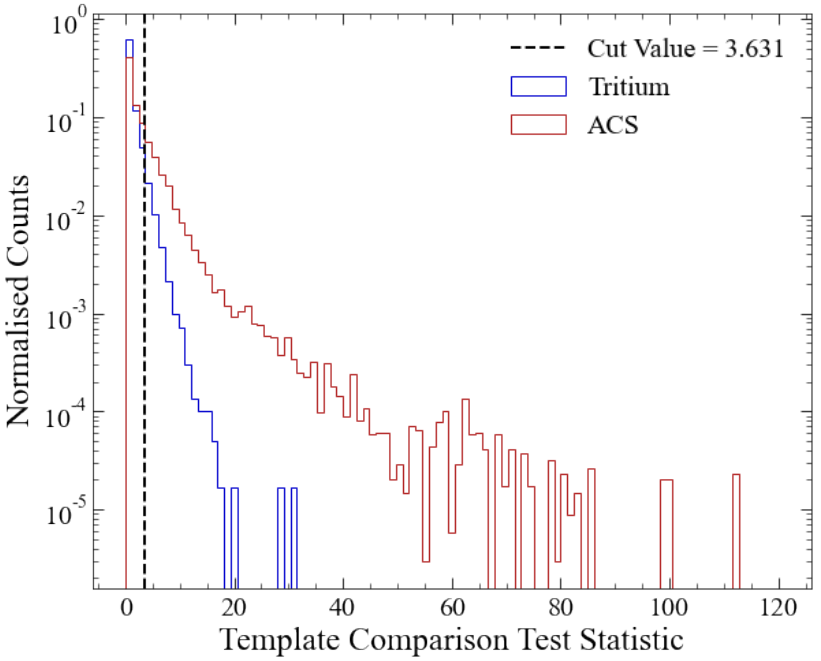


(b)

Figure 6.17: Distribution of test statistics for the GOF likelihood ratio, for tritium and ACS SR3 data with only baseline cuts applied, against number of S1 photons in (a), and following normalisation by the exponential fit to tritium data in (b). The cut value corresponding to 95% signal acceptance is shown.

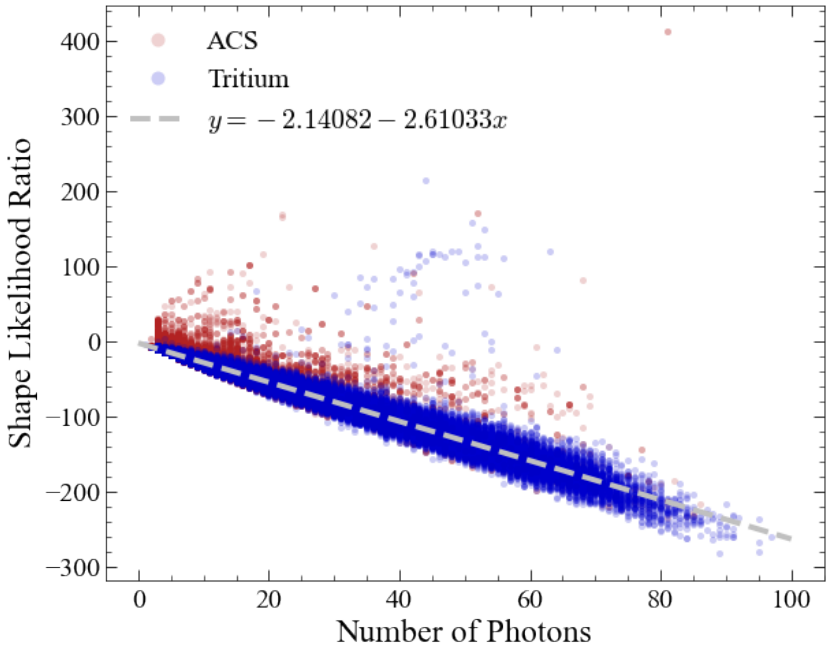


(a)

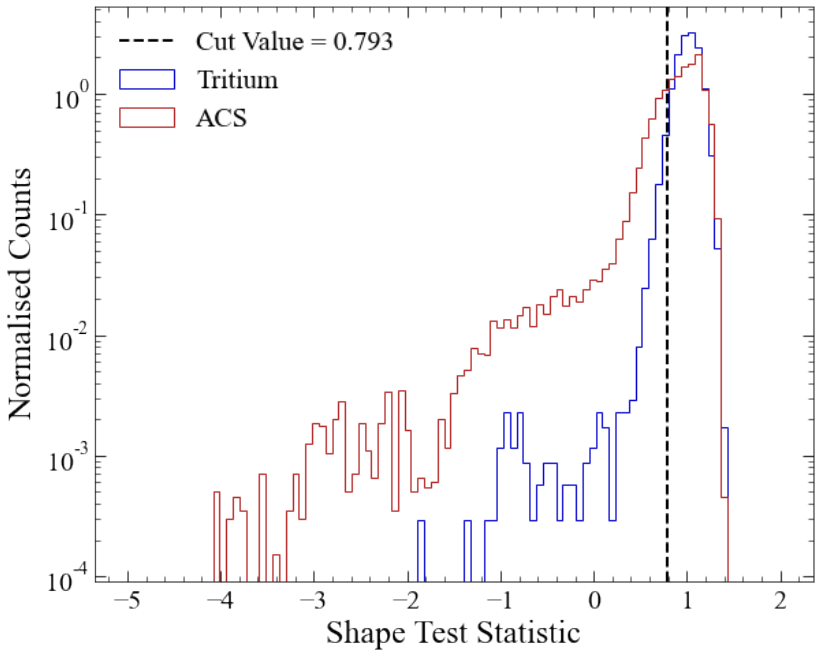


(b)

Figure 6.18: Distribution of test statistics for the template comparison likelihood ratio, for tritium and ACS SR3 data with only baseline cuts applied, against number of S1 photons in (a), and as a one-dimensional distribution, not requiring normalisation by a fit, in (b). The cut value corresponding to 95% signal acceptance is shown.



(a)



(b)

Figure 6.19: Distribution of test statistics for the shape-based likelihood ratio, for tritium and ACS SR3 data with only baseline cuts applied, against number of S1 photons in (a), and following normalisation by the linear fit to tritium data in (b). The cut value corresponding to 95% signal acceptance is shown.

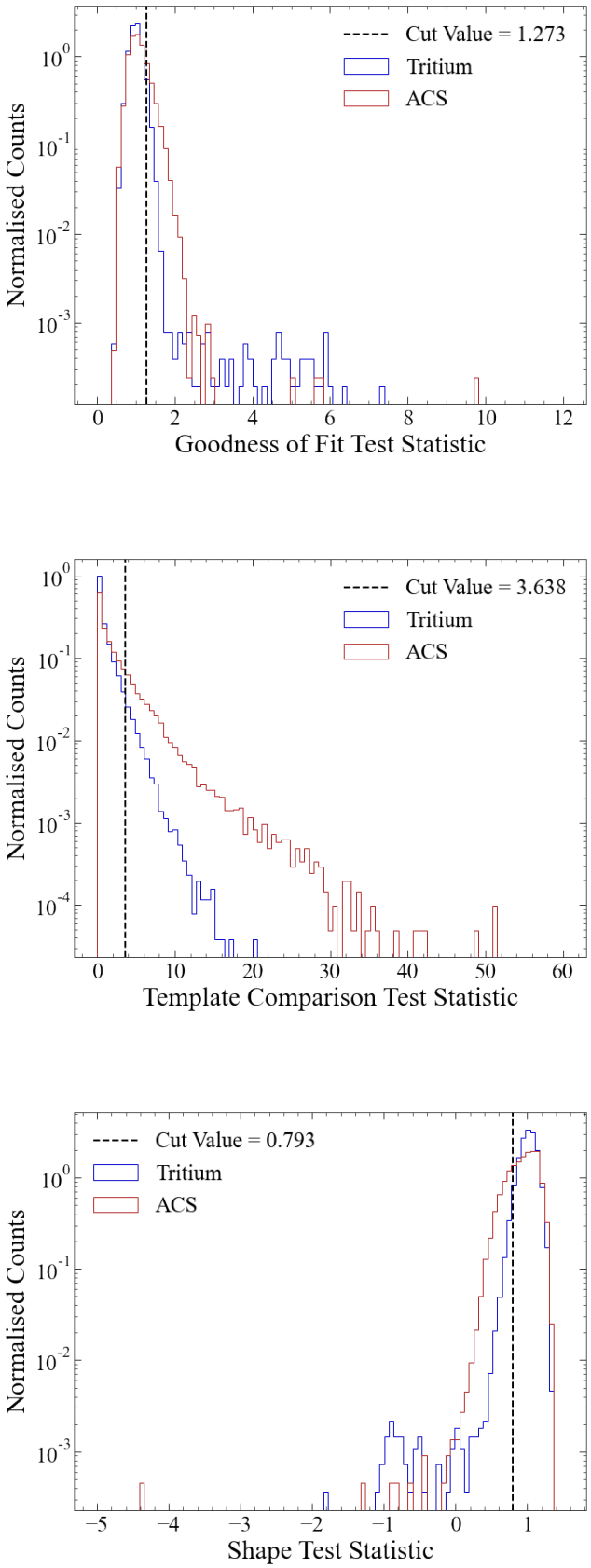


Figure 6.20: One-dimensional distributions of GOF, template comparison and shape-based test statistics, for SR3 tritium and ACS events with all other data quality cuts applied, following normalisation by tritium fits.

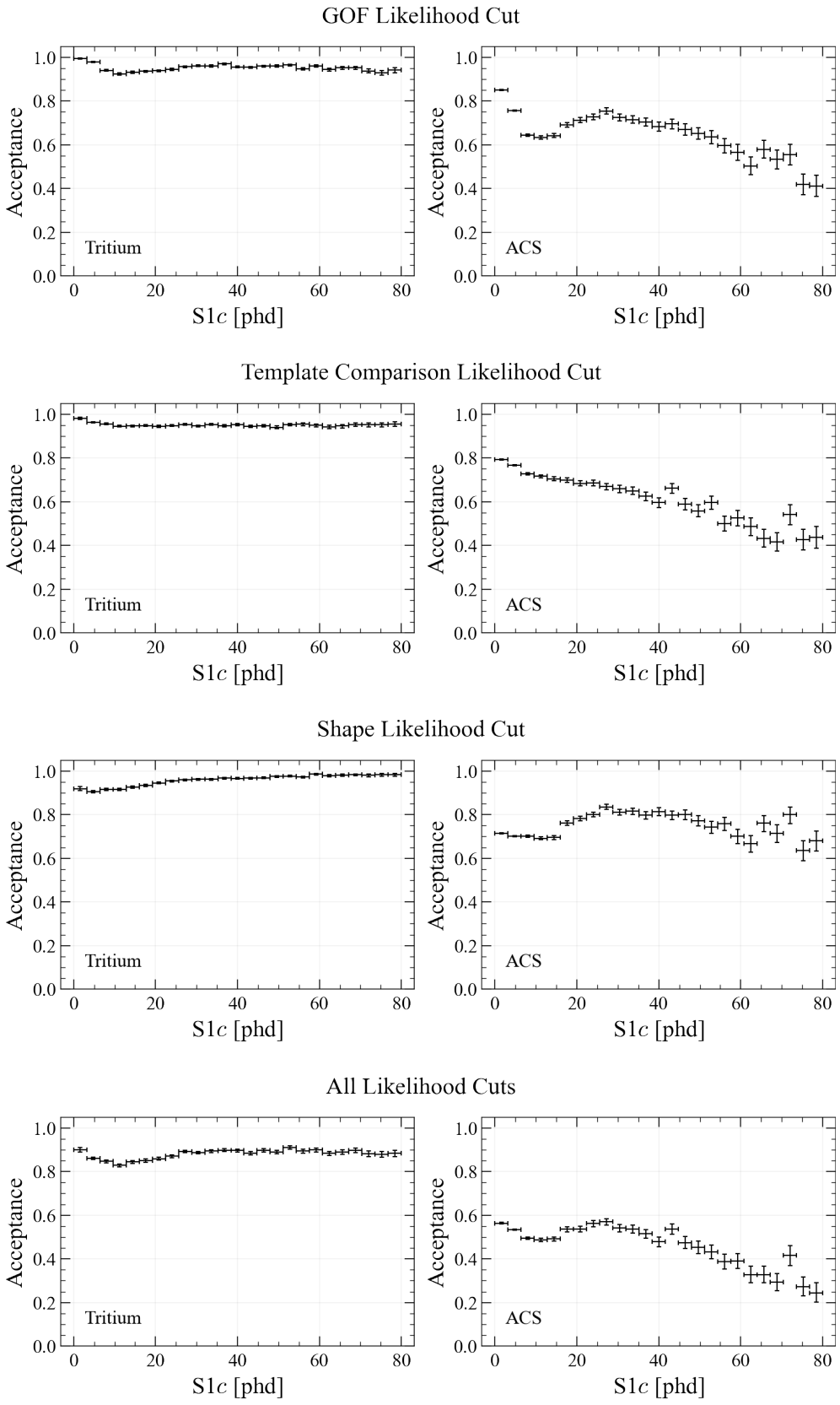


Figure 6.21: Acceptance for SR3 tritium (left) and ACS (right) as a function of S1c for each likelihood cut individually and all combined, following application of baseline cuts only.

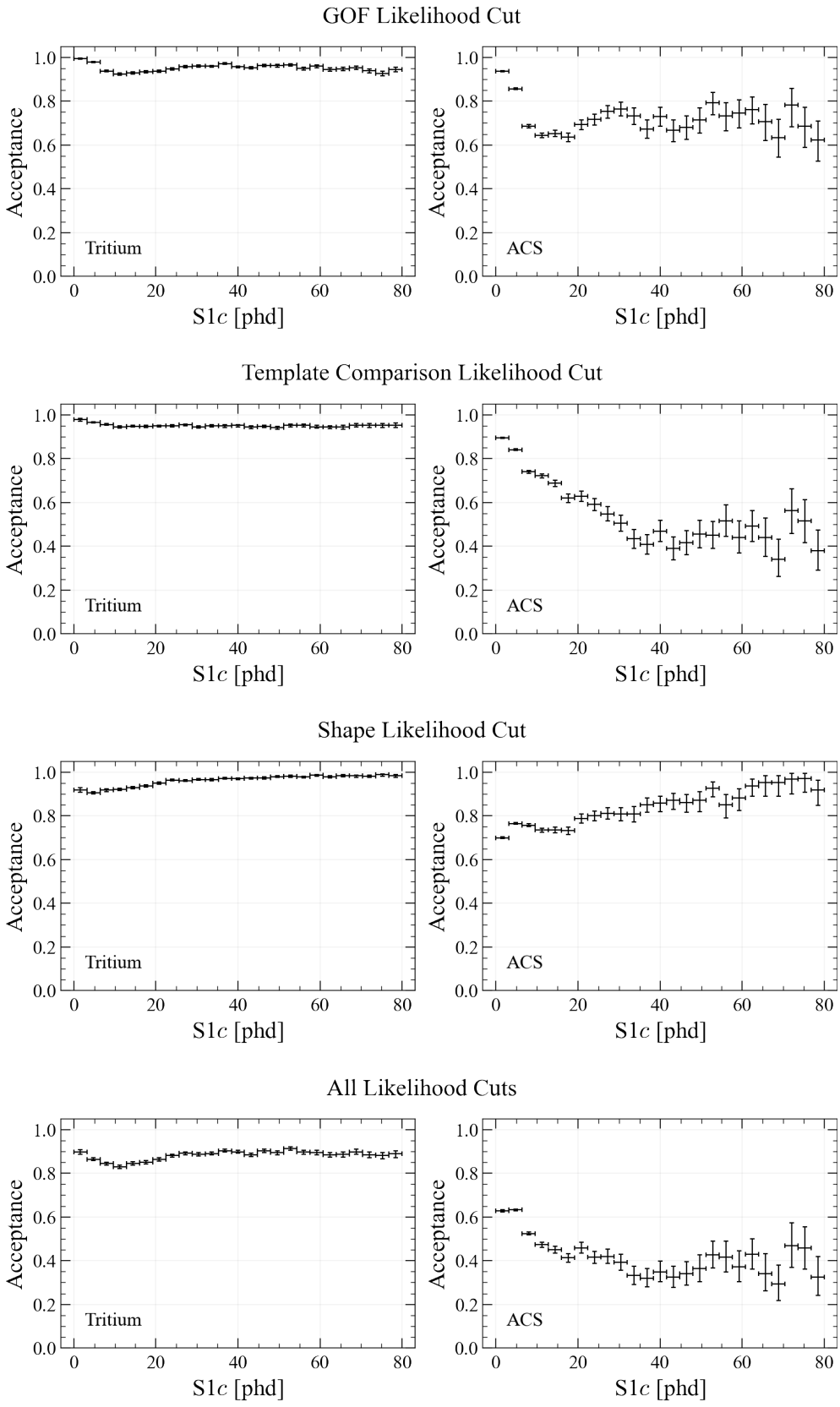
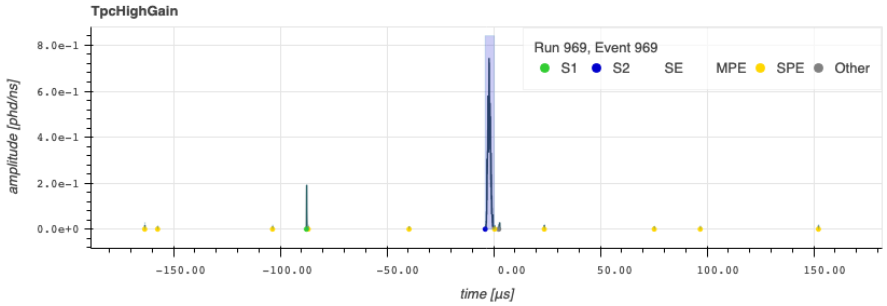
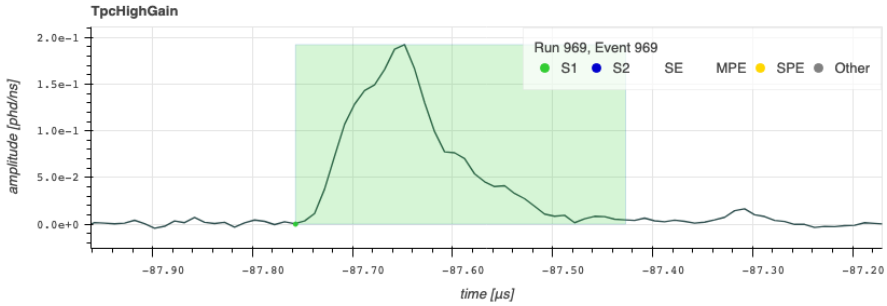


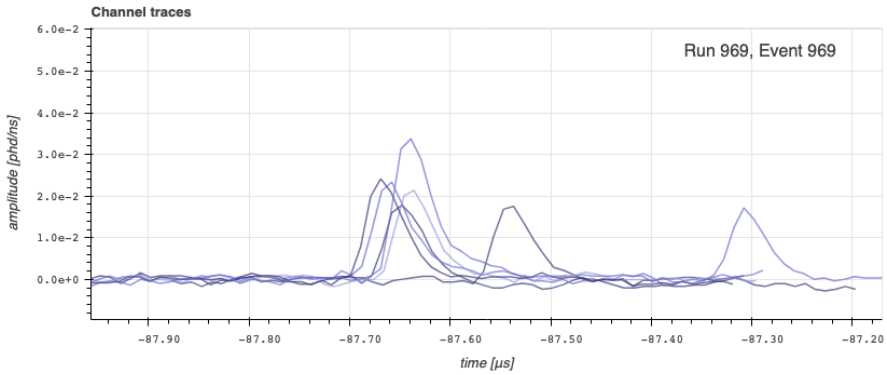
Figure 6.22: Acceptance for SR3 tritium (left) and ACS (right) as a function of S1c for each likelihood cut individually and all combined, following application of all other data quality cuts.



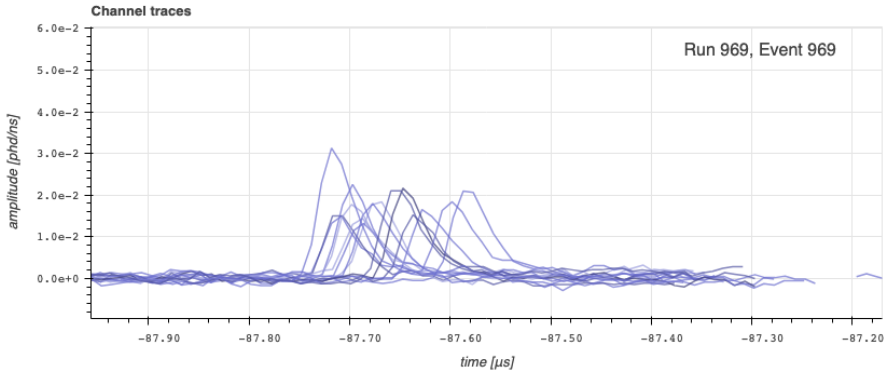
(a)



(b)



(c)



(d)

Figure 6.23: Event Viewer screenshot of an ACS event removed by the template comparison likelihood cut. The short time separation between S1 and S2 pulses in green and blue, respectively, can be seen in (a). The S1 waveform is magnified in (b), and the S1 channel waveforms are shown for the top (c) and bottom (d) PMT arrays.

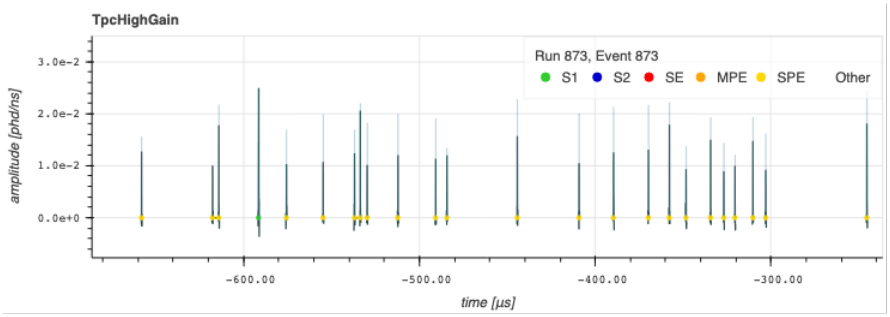
Table 6.1: The tritium signal acceptance and ACS background rejection efficiency for each likelihood cut and for all likelihood cuts combined, for the case where the individual cuts have been tuned to produce a signal acceptance of 95%. The samples on which the likelihood cuts were tuned and tested are specified in the first row. In the first scenario, only baseline cuts were applied to the samples, and in the second scenario all other SR3 data quality cuts were applied.

	Baseline Selection Cuts		All Data Quality Cuts	
Likelihood Cut	Acceptance	Efficiency	Acceptance	Efficiency
GOF	95.0%	25.4%	95.0%	18.2%
Template Comparison	95.0%	25.7%	95.0%	21.2%
Shape-based	95.0%	28.1%	95.0%	26.4%
All Likelihood Cuts	87.4%	47.1%	87.6%	42.4%

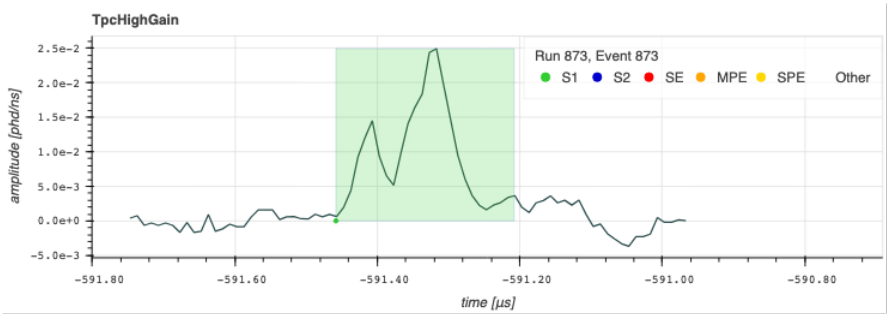
attributed to a pile-up of multiple S1 pulses that has been classified as a single S1, such as the example in Figure 6.25, which has an S1 area of 31.2 phd. Some pulses removed were likely misclassified as S1s, such as the example in Figure 6.26, where the S1 of size 11.1 phd looks to be part of a single electron pulse. The GOF cut is able to remove events with both unusual shapes and timing offsets, due to its agnostic nature.

6.5.4 Cut Correlation

As each cut has been tuned separately to provide a signal acceptance of 95%, the overall cut acceptance of 87.4% may be considered to be too low for the WIMP search. This could be improved by varying the acceptance of each cut individually, or alternatively the potential of combining multiple cuts simultaneously can be considered. To evaluate this potential for improvement, the correlation between each of the three cuts has been studied in Figure 6.27, using SR3 data with only baseline cuts applied. An excess of ACS events in a region which cannot be targeted by horizontal and vertical linear cuts alone indicates that a combined approach would result in an improved cut efficiency. This is particularly the case for the template comparison cut in combination with both the GOF cut and the shape-based cut. These observations resulted in the decision to investigate methods of combining the three cuts using machine learning, as discussed in chapter 7.



(a)



(b)

Figure 6.24: Event Viewer screenshot of an ACS event removed by the shape-based likelihood cut, likely due to SPE pile-up. The S1 (green marker) can be seen amongst SPE pulses (yellow markers) in (a) and magnified in (b).

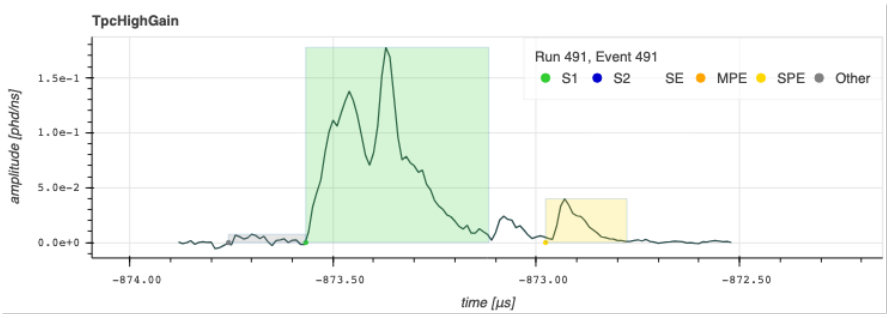


Figure 6.25: Event Viewer screenshot of an ACS event removed by the shape-based likelihood cut, likely due to S1 pile-up.

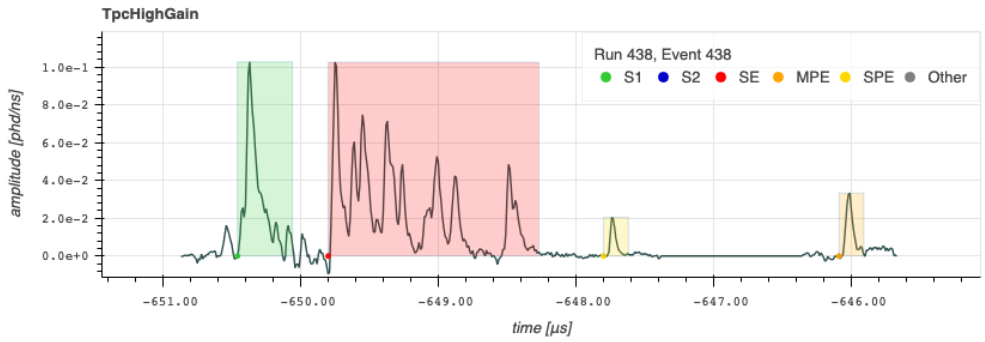


Figure 6.26: Event Viewer screenshot of an ACS event removed by the shape-based likelihood cut, likely due to pulse misclassification.

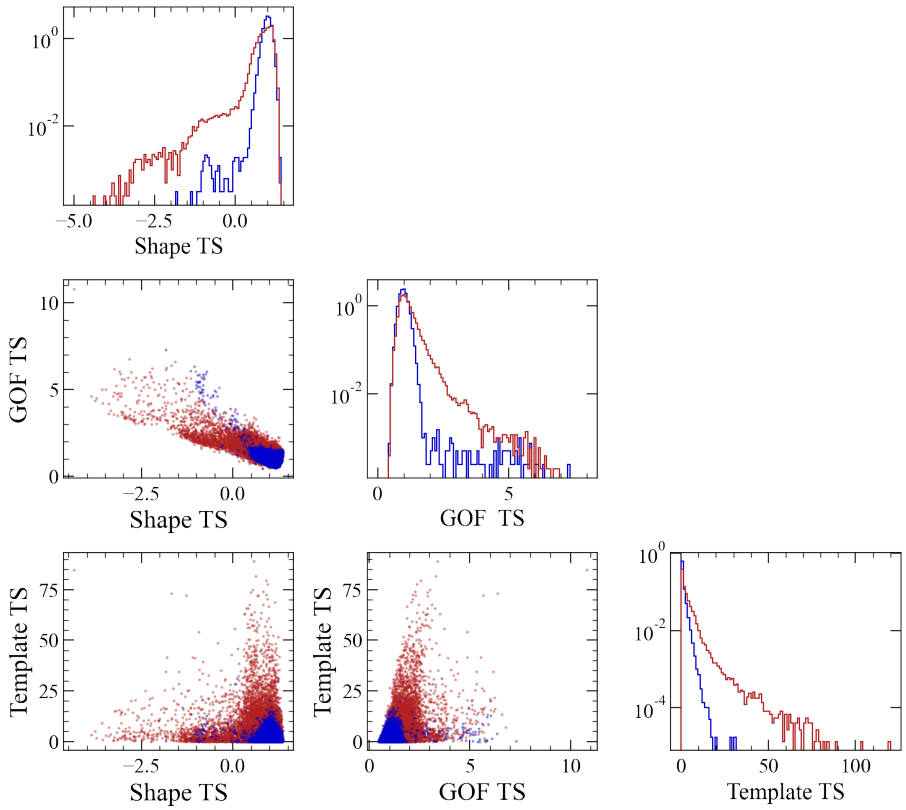


Figure 6.27: Test statistics for each likelihood cut plotted against each other for tritium (blue) and ACS (red) data to investigate cut correlation.

6.6 Summary

The photon timing offset observed in both LZ data and simulations has been exploited to construct three test statistics to target accidental-coincidence backgrounds. This resulted in the development of data quality cuts using SR1 data, including investigations of the dependence of pulse shape on area and interaction type. When applying all three linear cuts on SR3 data, a combined tritium signal acceptance of 87.4% and a combined background rejection efficiency of 47.1% was achieved, or 87.6% and 42.4% following application of all other data quality cuts. Accidental events across the entire S1c range were removed, including pathologies such as the pile-up of SPEs or S1s, as well as misclassified pulses. A study of the correlations between the three cuts indicates that applying the cuts simultaneously rather than individually would allow for an improved background rejection efficiency for the same signal acceptance. The focus of chapter 7 is a boosted decision tree, which was developed for the purpose of combining these cuts.

Chapter 7

Accidental-Coincidence Background Removal with Machine Learning

Boosted decision trees (BDTs) are a powerful tool for event classification in particle physics. It has been shown in subsection 6.5.4 that the test statistics which formed the basis of linear cuts in chapter 6 would be more effective for background rejection if utilised simultaneously. To achieve this, the test statistics can be used as input parameters for a BDT. This allows for all three metrics to be considered in the classification decision, and is thus expected to improve the rejection efficiency of accidental-coincidence backgrounds. The development of the analysis from a single preliminary BDT to a tuned array of 14 models, and its impact on the sensitivity of the LZ experiment to WIMPs, will be the focus of this chapter.

7.1 Boosted Decision Trees

For more than 30 years, machine-learning techniques have been employed in the field of high energy physics [161] [162], where the objective is typically to identify a rare physics signal amongst background processes occurring at a much higher rate. Traditional cut-and-count methods, which optimise the ratio of signal to background events passing thresholds on various parameters, are not sensitive in regions of substantial overlap between the two populations. Often, an event will not meet one of the selection criteria, but if other parameters indicate that it is signal-like, it would be undesirable to reject it [163].

A decision tree is a method of event classification, in which a model is trained on (often simulated) signal and background events, labelled 1

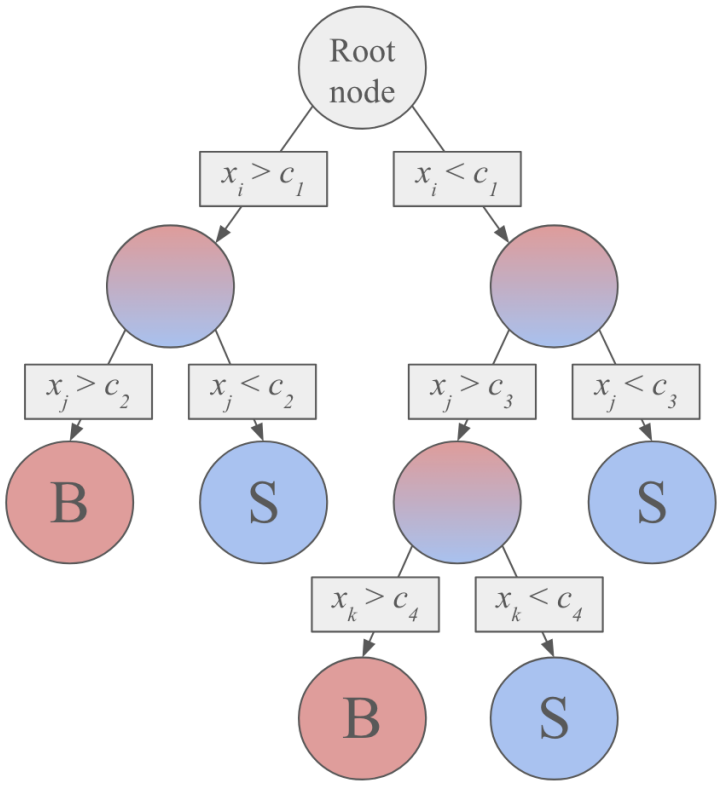


Figure 7.1: Schematic of a single decision tree. At each node, splits are made on features $x_{i,j,k}$ with cut values $c_{1,2,3,4}$, to classify events into signal, "S", and background, "B".

and 0, respectively. At each node, a binary cut is made on a feature, as shown in Figure 7.1, such that multiple criteria are considered in the classification decision. A single decision tree is susceptible to overtraining, whereby the model is trained on statistical fluctuations in the training data. BDTs employ the process of boosting, which overcomes this issue by using an ensemble of trees. The final prediction for the i^{th} event is defined as

$$\hat{y}_i = \sum_{k=1}^{K} f_k\left(x_i\right), \tag{7.1.1}$$

where $\{f_1, f_2, ..., f_k\}$ is a set of trees, which are functions that map attributes to weights, w_i. For the Extreme Gradient Boosting (XGBoost) algorithm [164] used in this analysis, the following objective function is minimised:

$$\begin{aligned} \mathcal{L}(\phi) &= \sum_i l\left(\hat{y}_i, y_i\right) + \sum_k \Omega\left(f_k\right), \\ \text{where } \Omega(f_t) &= \gamma T + \alpha \sum_{j=1}^{T} |w_j| + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \end{aligned} \tag{7.1.2}$$

where l is the loss function that measures the difference between the

prediction, $\hat{y}_i$, and target, y_i. Ω is a regularisation term, which is a feature of the XGBoost algorithm that penalises complexity to avoid overfitting. Here, γ, α and λ are regularisation parameters, with the former penalising trees with a large number of leaves (terminal nodes where the splitting stops), resulting in shallower trees. α and λ are L1 and L2 regularisation terms, which reduce the sensitivity of the model to individual observations. T is the number of leaves in the tree and w is the sum of the leaf weights. Equation 7.1.2 cannot be optimised by traditional methods in Euclidean space, so the model is trained additively, using a second-order Taylor approximation of the objective function for fast optimisation. This is the process of boosting, which can be thought of as "greedily" adding the tree f_t that minimises

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t\left(\mathbf{x}_i\right)\right) + \Omega\left(f_t\right), \tag{7.1.3}$$

where $\hat{y}_i^t$ is the prediction of the i-th instance at the t-th iteration. In this way, each new tree gradually corrects for errors in the previous trees.

The dissimilarity between the target and prediction is commonly measured using the binary cross entropy or log loss, in which the loss function is defined as [165]:

$$\sum_i l\left(\hat{y}_i, y_i\right) = -y_i \ln \hat{y}_i - (1 - y_i) \ln(1 - \hat{y}_i). \tag{7.1.4}$$

This function is minimised when the prediction $\hat{y}_i$ is close to zero for background events, which have a target y_i of zero, or close to one for signal events, where y_i is one.

The use of BDTs is prevalent in the particle physics community, due to their speed and scalability, as well as their robustness against missing values and imbalanced datasets [163]. For example, a BDT was used in producing the first evidence for $t\bar{t}t\bar{t}$ production in the ATLAS experiment [166]. The LZ collaboration utilised a BDT in searches for WIMP-nucleon effective field theory couplings in SR1 data over an extended energy range. The BDT was used for the rejection of γ-X backgrounds, in which multiply-scattering γ rays can be misclassified as Single Scatters, due to at least one interaction occurring in a region of the detector from which charge cannot be collected [167].

In the context of background rejection in particle physics, the BDT assigns a score in the range [0,1], where 0 and 1 are the target scores for background and signal events, respectively. By examining the label of events in a validation dataset, it is then possible to place a cut on BDT

score at the desired signal acceptance. Tuning of a single cut on BDT score is more feasible than adjusting multiple linear cuts.

In the case of the S1 likelihood cut, the three test statistics can be used as the input features for a BDT. SR3 tritium and AccidentalChopStitch events were again chosen for the signal and background datasets. Initially, the BDT was trained on datasets with only baseline data quality cuts applied. To remove the dependence on the number of S1 photons, the likelihood ratios were again normalised by fits to the tritium data prior to input into the BDT, as in the linear cut method.

7.2 Hyperparameters

In order to optimise the XGBoost model, several hyperparameters can be tuned. The tunable parameters considered in this study, as informed by Ref. [168], are as follows:

- **Number of trees:** This parameter controls the number of boosting iterations, in which a new tree is applied. A sufficiently high value is required to avoid underfitting, whereby the model is unable to learn the training data effectively, resulting in poor performance on both training and validation data. Conversely, a very large number of trees results in a more complex model prone to overfitting.

- **Learning rate, η:** After each boosting stage, η reduces the feature weights. A low value of η therefore slows down learning and leads to a more conservative boosting process, avoiding overfitting. For a given learning rate, as the number of trees is increased, the performance of the model will generally improve, before plateauing. A lower η typically requires a larger number of trees to reach the optimal performance. This smaller step size results in slow computation, but makes it easier to reach the optimum performance for a model.

- **Maximum depth:** This hyperparameter governs the maximum number of levels in a tree. A high value results in a complex model, which may be prone to overfitting.

- **Subsampling fraction:** The subsampling fraction controls the proportion of training data sampled for use in each boosting iteration. Selection of a low value produces a more conservative model, preventing overfitting, but the chosen value should not be so low as to result in underfitting.

Table 7.1: Values of each hyperparameter selected for initial BDT investigations. Further tuning was later carried out for the "number of trees" and "maximum depth" parameters.

Parameter	Value
Number of Trees	500
Learning Rate η	0.01
Maximum Depth	6
Subsampling Fraction	0.8
L1 Regularisation, α	0.4
L2 Regularisation, λ	0.4

- **L1 and L2 regularisation:** α and λ are L1 and L2 regularisation terms applied to the feature weights as in Equation 7.1.2, and are analogous to Lasso and Ridge regression, respectively. L1 regularisation penalises the absolute value of the weights, shrinking some to zero. It is robust to outliers in the data. L2 regularisation acts on the squares of the weights, forcing them to be small but never zero. It is preferable for data with highly correlated features, as it distributes their impact more evenly [169]. Regularisation reduces the sensitivity of the model to small changes in the training data, and hence reduces overfitting.

In preliminary investigations, these parameters were set heuristically, guided by conventional values used in similar studies (for example, Ref [170]). Their chosen values are shown in Table 7.1. The number of trees and maximum tree depth were later tuned for each model using a grid search, as described in further detail in subsection 7.4.2.

7.3 Initial BDT Investigations

7.3.1 BDT Performance

Initially, a dataset containing an equal number of tritium and ACS events was divided into training and validation samples, with an 80-20 split. A single XGBoost model was trained on events across the entire range of S1c values simultaneously. The trained model was then used to assign a score to events in the validation dataset. The distributions of BDT scores for events labelled as signal and background are shown in Figure 7.2. The performance of the BDT was measured using a Receiver Operating Characteristic (ROC) curve, as shown in Figure 7.3, which shows the true

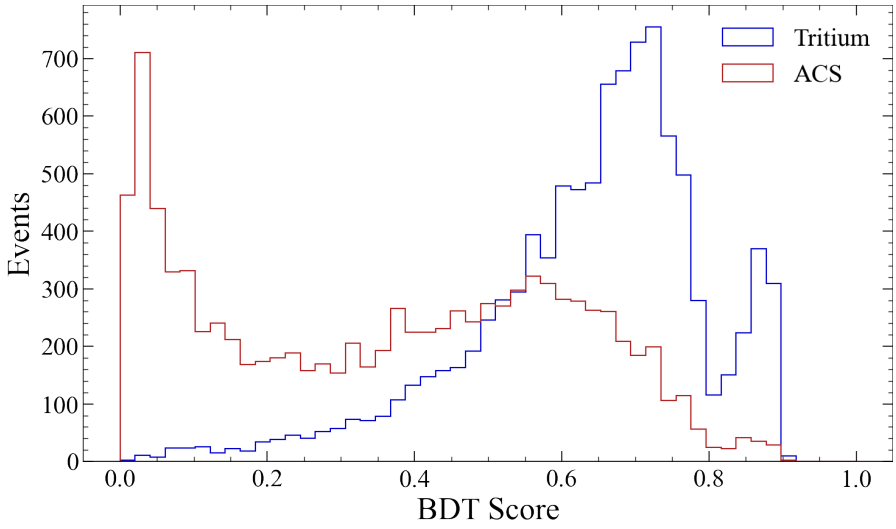


Figure 7.2: Distribution of BDT scores for tritium signal and ACS background events in validation data using the preliminary BDT model.

positive rate as a function of false positive rate. This was determined by finding the tritium acceptance and accidental acceptance for every BDT cut value tested. The optimal scenario of 100% signal acceptance for 0% background leakage corresponds to the top-left corner of the plot. The BDT performed well on the validation data, with a 44.7% accidental rejection efficiency for a 95% signal acceptance.

In order to identify which features were most useful in deciding the classification of events, the importance of each input variable was quantified in three ways. The weight refers to the frequency with which a feature is used to make splits; the coverage measures the number of events for which a given feature affects the splits; and the gain is the reduction in the loss function (and hence the improvement of the accuracy of predictions) that is achieved by including a feature. The relative importance described by these three metrics is shown for each test statistic in Figure 7.4. The shape-based test statistic provides the highest weight, and has values for the cover and gain that are only slightly below those for the template comparison test statistic.

To ensure the validity of the results from this model, three BDTs were trained, each with one of the test statistics as the single input parameter. In this case, the BDT is expected to produce a similar accidental rejection efficiency to the equivalent linear cut for the same tritium signal acceptance. For both the template comparison and goodness of fit tests, this was indeed the case. However, a cut on BDT score at 95% tritium signal acceptance for the model which used the shape-based test statistic returned an efficiency of 35.3%, which was a significant improvement on the linear cut efficiency of 28.1%. In order to investigate this, events fail-

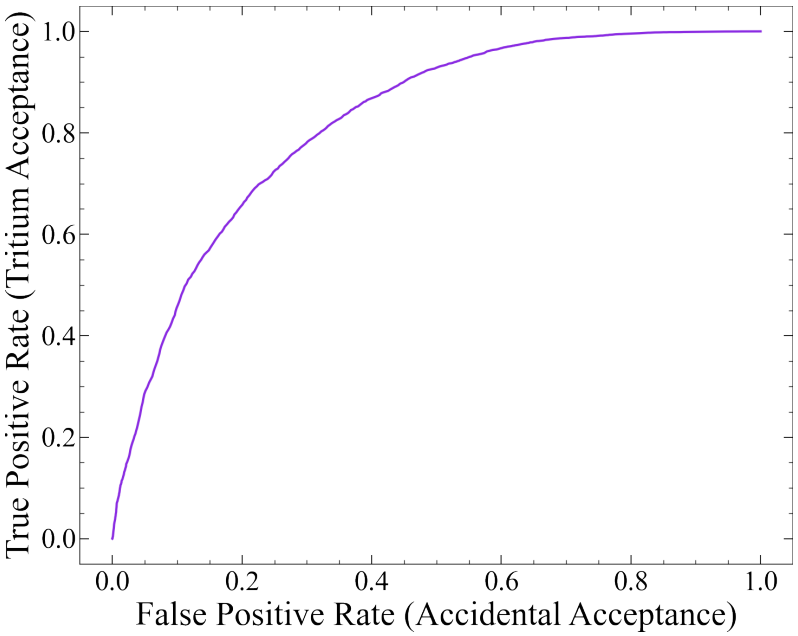


Figure 7.3: ROC curve for the preliminary BDT model showing tritium acceptance as a function of accidental acceptance for many possible BDT score cut values.

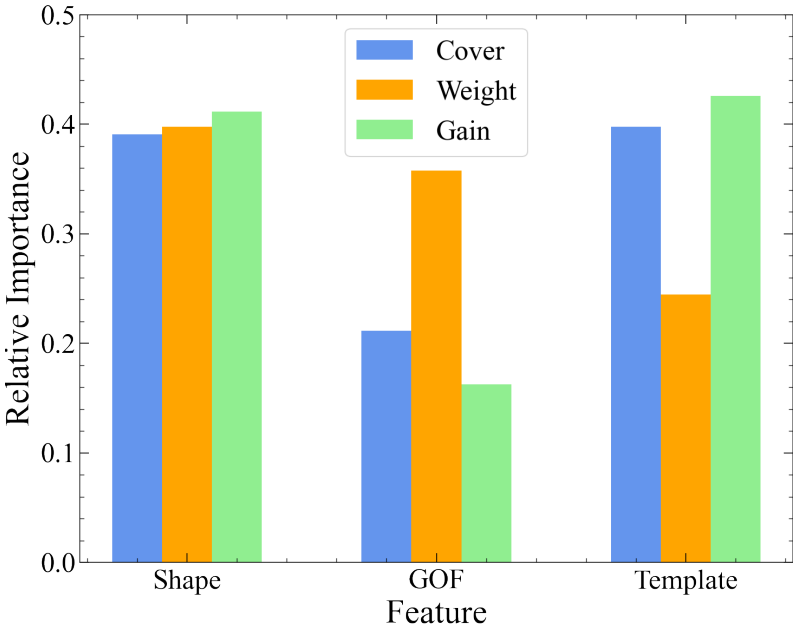


Figure 7.4: Relative importance of each feature in the preliminary BDT model, as defined by the cover, weight and gain metrics.

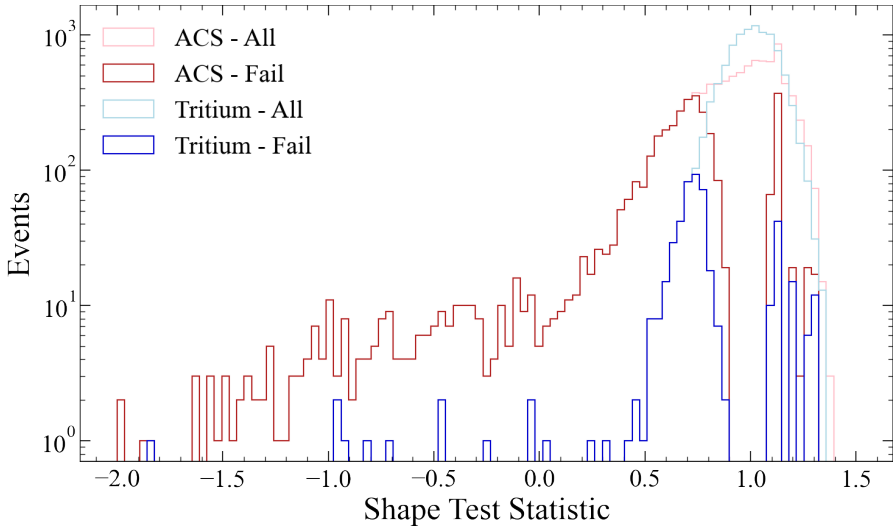


Figure 7.5: Distribution of shape-based test statistic values for tritium and ACS validation data. The dark blue and red histograms indicate events failing a cut on BDT score with 95% tritium signal acceptance, for a model with the shape-based test statistic as the only input feature. The light blue and pink histograms show all tritium and ACS validation data, respectively.

ing the BDT score cut were identified in the distribution of shape-based test statistic values, as shown in Figure 7.5.

7.3.2 Test Statistic Quantisation

It would be expected that primarily events with a low shape-based test statistic would be assigned a low BDT score, and thus fail the cut. Close to the peak of the distribution, the rate of signal events is higher than that of background events, so a higher BDT score would be expected. However, it can be seen that some signal and background events in this region (specifically, in the ranges ∼[0.82,0.90] and ∼[1.06,1.12]) are in fact removed by the BDT cut.

Reproducing this histogram with a finer resolution for a small range of values close to the peak of the distribution, as in Figure 7.6, revealed that the rate of accidental events did in fact repeatedly spike above the rate of signal events. Statistical fluctuations were unlikely to be the origin of these spikes, as the variation of event counts is large compared to the tritium data. The good performance of the BDT on the validation data implied that the spikes occurred at the same test statistic values in both the training and testing datasets.

An even finer resolution, as in Figure 7.7, demonstrated that there was quantisation present, as exactly the same test statistic values were produced for multiple events. This is expected, as the same likelihood value can occur if the distribution of S1 photons between the 10 ns wave-

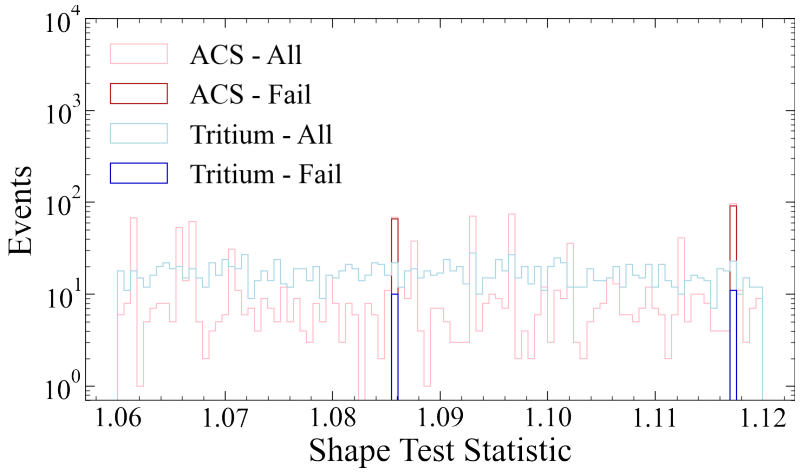


Figure 7.6: As in Figure 7.5, but with finer binning, to demonstrate the spikes in ACS rate close to the peak of the shape-based test statistic distribution.

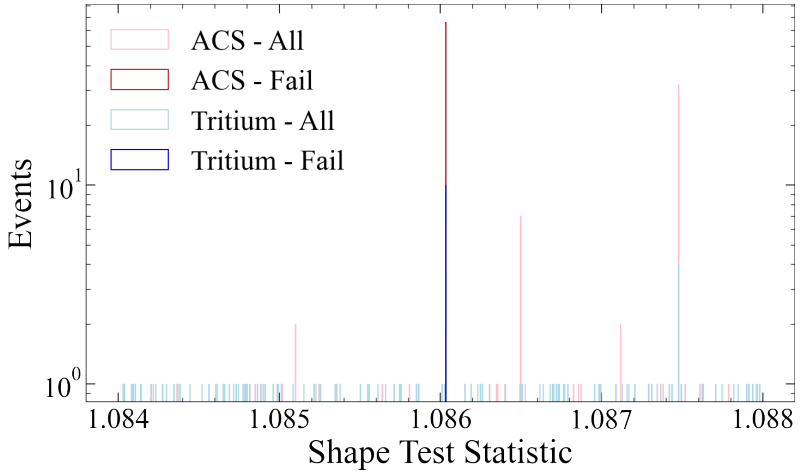


Figure 7.7: As in Figure 7.6, but with even finer binning, to demonstrate the quantisation of test statistic values.

form sampling time bins is the same for two events. This is more likely to transpire for events with a small number of S1 photons, for which there are fewer possible combinations. This quantisation was also seen in the template-comparison and GOF test statistics, but to a lesser extent, as the drift time must also be in the same range in order to produce identical values for the likelihood ratio.

This introduced a dependence of event classification on S1c, as the ACS dataset contains a larger number of small pulses, which are more likely to produce duplicate test statistic values. The BDT would learn to treat events with these repeated values as accidental-like, so events, both signal and background, with a small number of photons, were more likely to be rejected by the BDT cut. This undesirable behaviour is demonstrated most clearly by the tritium acceptance curve, shown in Figure 7.8, which declines significantly for low S1c values.

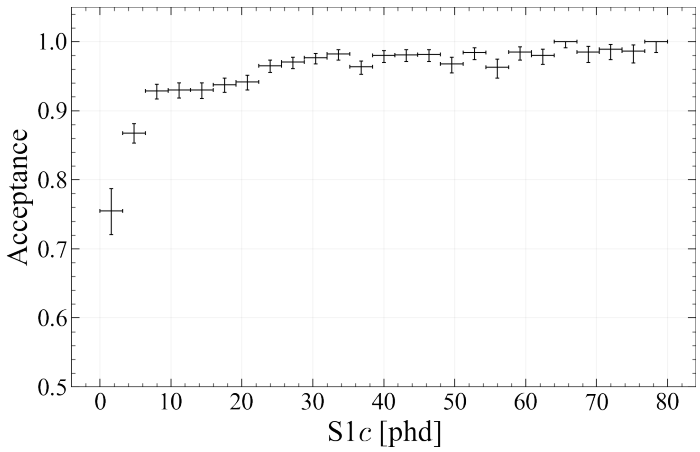


Figure 7.8: Tritium acceptance as a function of S1c for the BDT cut with an overall tritium signal acceptance of 95%, for a model with the shape-based test statistic as the only input feature.

A temporary solution was introduced, in which the BDT input variables were binned to smear out the quantisation. When the shape-based single-variable BDT was retrained on this data, the model did not learn the spikes in the data, so, as expected, events close to the peak of the test statistic distribution were no longer removed by the BDT cut, as shown in Figure 7.9. This approach likely incurs a cost in terms of loss of information, so was not used to produce the final results.

7.3.3 Cherenkov Light

A population of accidental events with the highest test statistic values was observed at a higher rate than the tritium signal. On investigation, it was discovered that many of these events had a higher corrected S1 area than expected for the number of S1 photons detected, as shown in Figure 7.10. Here, 4000 signal and background events are shown, with test statistics less then and greater than one. For accidental events with a high test statistic, a new population of events is observed, away from the correlated band. It was taken into consideration that ACS events have a randomly-assigned drift time, so the correction factor applied to the pulse area is likely to be incorrect for the true position of the S1, which may result in a larger area. However, the effect observed here is too large to be attributed solely to this behaviour. By applying a cut to select and inspect these events, it was found that they generally had properties consistent with the Cherenkov light that can be produced in PMT windows.

The majority of light from a Cherenkov S1 pulse is detected in a single PMT. Inspection of the channel waveforms from the selected population of events, such as those shown in Figure 7.11, demonstrated that this

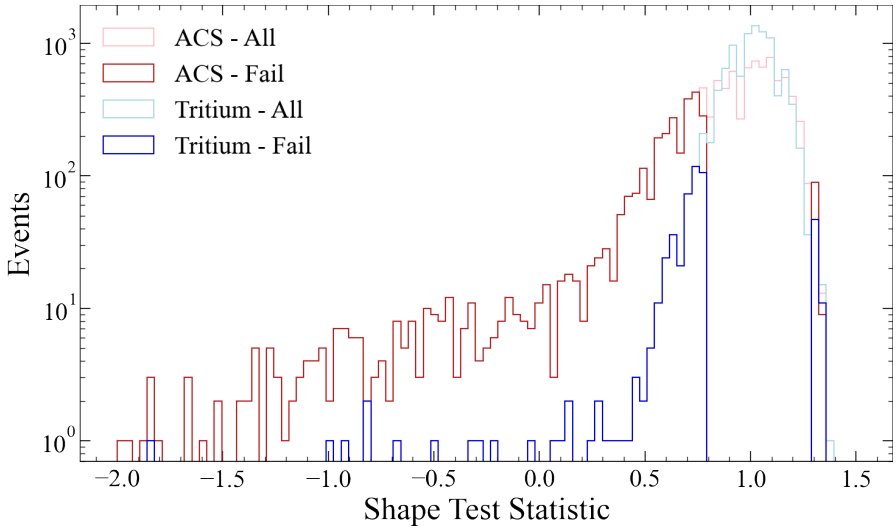


Figure 7.9: Distribution of shape-based test statistic values for tritium and ACS validation data. The dark blue and red histograms indicate events failing a cut on BDT score with 95% tritium signal acceptance. The shape-based test statistic is the only input feature for this model, and the input values were binned to smear out quantisation. The light blue and pink histograms show all tritium and ACS validation data, respectively.

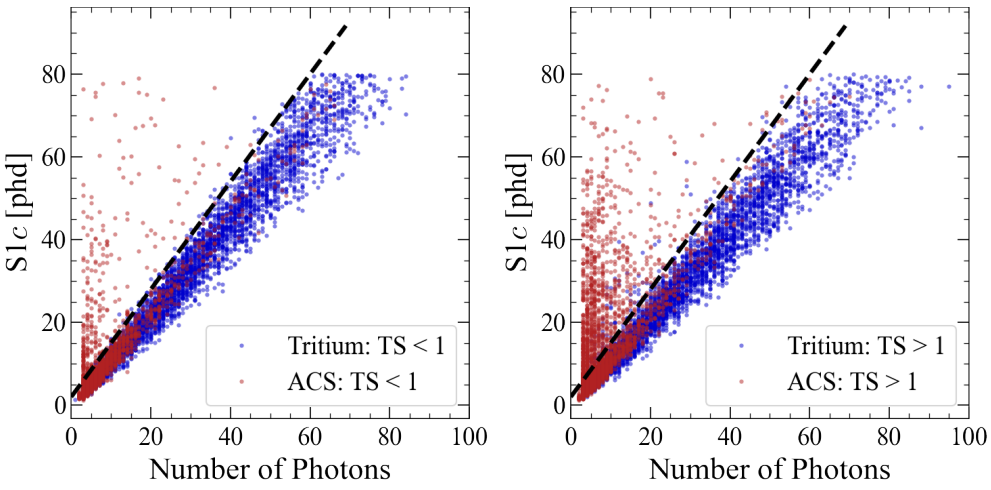


Figure 7.10: Distribution of tritium and ACS events in S1c-$n_{photons}$ space, for events with a shape-based test statistic less than one (left) and greater than one (right). The black dashed line indicates a cut used to select events with a high S1c value for the given number of S1 photons.

was indeed the case. A small pulse width is also expected for Cherenkov S1s, and Figure 7.12 shows that this population of S1 pulses is narrower than other accidental events.

An explanation as to why these events have a high test statistic value is that, during calculation of the likelihood ratio, the waveform alignment can be chosen such that narrow pulses are contained entirely within the peak of the S1 template, whereas wider pulses incur a penalty from regions of the template where the probability is low. This introduces another small dependence on S1c. It can be seen in Figure 7.13 that pulse length is generally independent of S1c, except for at small pulse areas, where the pulse tends to be narrower. Small tritium pulses are therefore more likely to have a similar test statistic value to Cherenkov S1s, and thus be rejected by the BDT cut. This remaining small dependence on S1c can be seen in the tritium acceptance curve in Figure 7.14.

It may be concluded that it is acceptable for the BDT to treat these events with a high shape-based test statistic as accidental-like, however there already exists a data quality cut specifically designed to target Cherenkov events: the high single channel cut. It was therefore decided that the BDT will instead be trained only on the accidental events surviving all other data quality cuts. This would allow for the determination of the power of the BDT cut on top of the existing analysis.

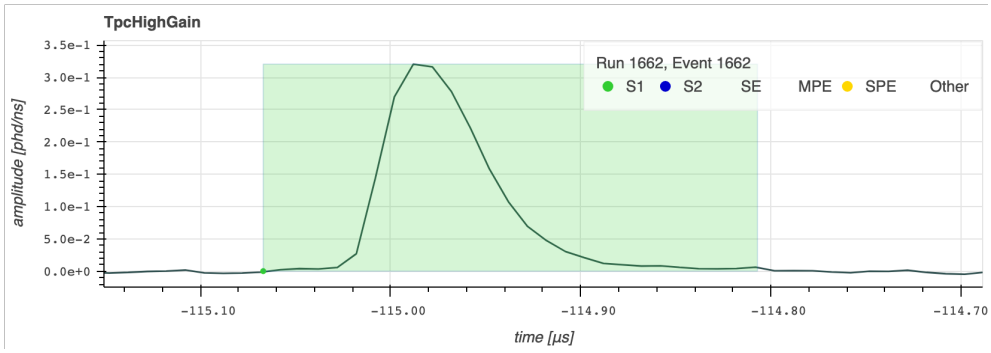
7.4 Binned BDT Analysis

7.4.1 Approach

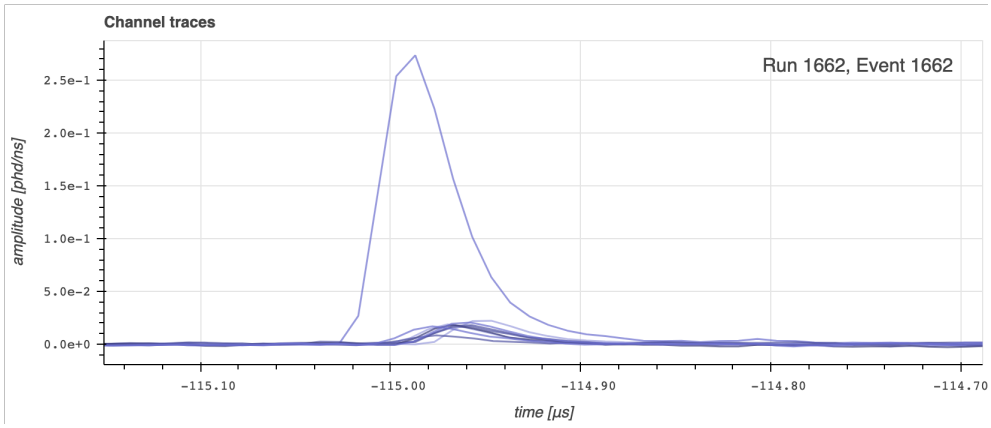
In order to preserve all available information, and minimise dependence on S1c, it was decided that, rather than binning the BDT input variables, multiple BDTs would be trained separately. Each input dataset would contain an equal number of signal and background events, with each possible number of S1 photons up to 15. The number of available ACS events with more than 15 photons was low, so, to maintain high statistics, a single BDT model was trained for events with $n_{photons} > 15$. This was deemed to be acceptable, as no test statistic quantisation was observed in this region.

7.4.2 k-fold Cross Validation

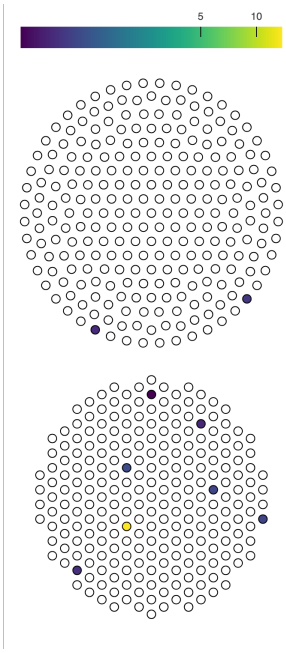
The "maximum depth" and "number of trees" hyperparameters were tuned separately for each BDT using a grid search, where k-fold cross validation was employed. This is the process in which the total dataset



(a)



(b)



(c)

Figure 7.11: Screenshots from the LZ Event Viewer, showing an example of an ACS S1 pulse with a high-shape based test statistic, likely to be Cherenkov light. The relatively narrow S1 pulse is shown in (a), and the channel waveforms in (b), as well as the TPC PMT heatmap in (c), demonstrate that the majority of light is detected by one PMT.

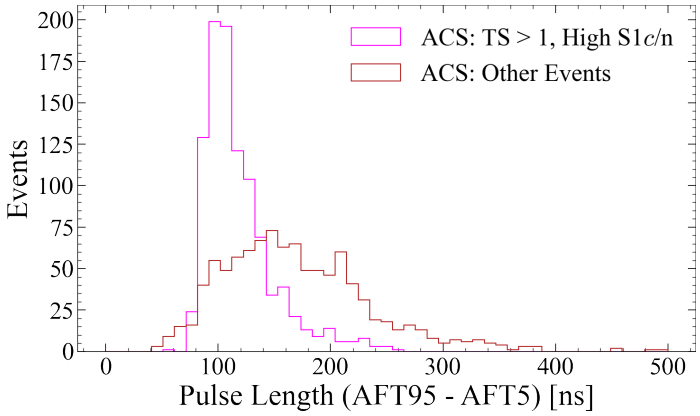


Figure 7.12: S1 pulse length, defined as AFT95 - AFT5, for ACS events with a high shape-based test statistic and high S1c for the given number of photons (pink), compared to other ACS events (dark red).

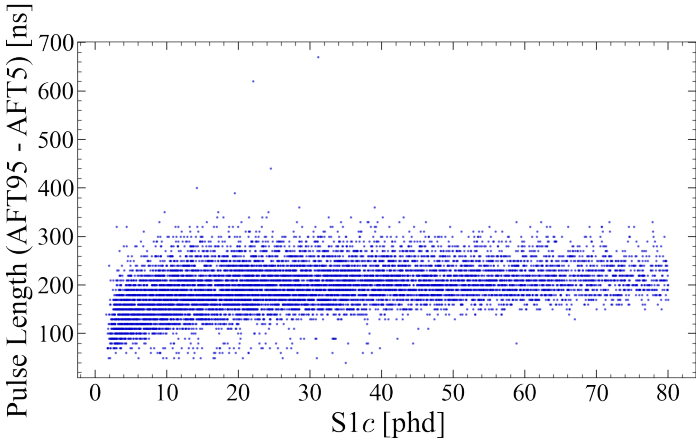


Figure 7.13: S1 pulse length, defined as AFT95 - AFT5, against S1c for tritium events.

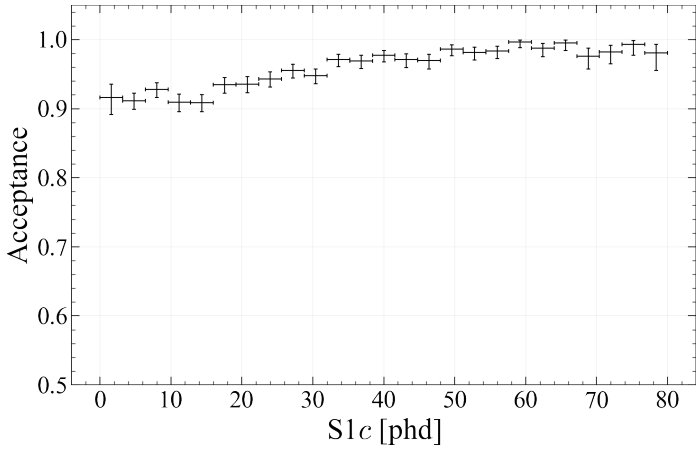


Figure 7.14: Tritium acceptance as a function of S1c for the BDT cut with an overall signal acceptance of 95%, for a model with the shape-based test statistic as the only input feature, in which the input values were binned to smear out quantisation.

is divided into k subsets of approximately equal size. One of k models is trained on all subsets except one, which is treated as the validation dataset for that model and scored. This means that all events are used for both training and validation, as opposed to the previous approach, in which only a small subset of the data was available for scoring. Five and 10 are typical values for k [171], and $k = 5$ was selected for this study. The GridSearchCV module from the sklearn Python package was utilised to operate a cross-validated parameter grid search. The combination of "number of trees" and "maximum depth" that maximised the "negative log loss" metric, and thus minimised the loss function, was selected for each model.

7.4.3 BDT Performance

The trained BDT models were then used to produce scores for every event. The distribution of BDT scores for each model is shown in Figure 7.15. It can be seen that separation of the background and signal populations generally improves as number of photons is increased, as expected. An ROC curve was then generated for each model, as shown in Figure 7.16. To maintain a constant signal acceptance across all models, a different BDT cut value would be selected for each value of $n_{photons}$. Example acceptance curves for a BDT cut with a signal acceptance of 95% are shown in Figure 7.17. By setting the acceptance requirement for each BDT separately, the dip in acceptance at low S1c suffered by the combined linear cuts is not encountered here. The background rejection efficiency could then be calculated separately for each model, as well as over the entire ACS dataset. When choosing an 87.6% signal acceptance for each model, equivalent to the combined linear cut acceptance in chapter 6, the total accidental rejection efficiency is 48.1%, compared to 42.4% for the linear cuts. The feature importance for each model is shown in Figure 7.18, where it can be observed that no single variable is consistently outperforming the other two.

7.5 Cut Optimisation

It is desirable to determine the cut value that would provide the optimal sensitivity to WIMPs. This process was carried out for both the linear and BDT versions of the S1 likelihood cut, to allow for an accurate comparison. Feldman and Cousins introduced the ratio

$$R = P(n \mid \mu)/P\left(n \mid \mu_{\text{best}}\right) \tag{7.5.1}$$

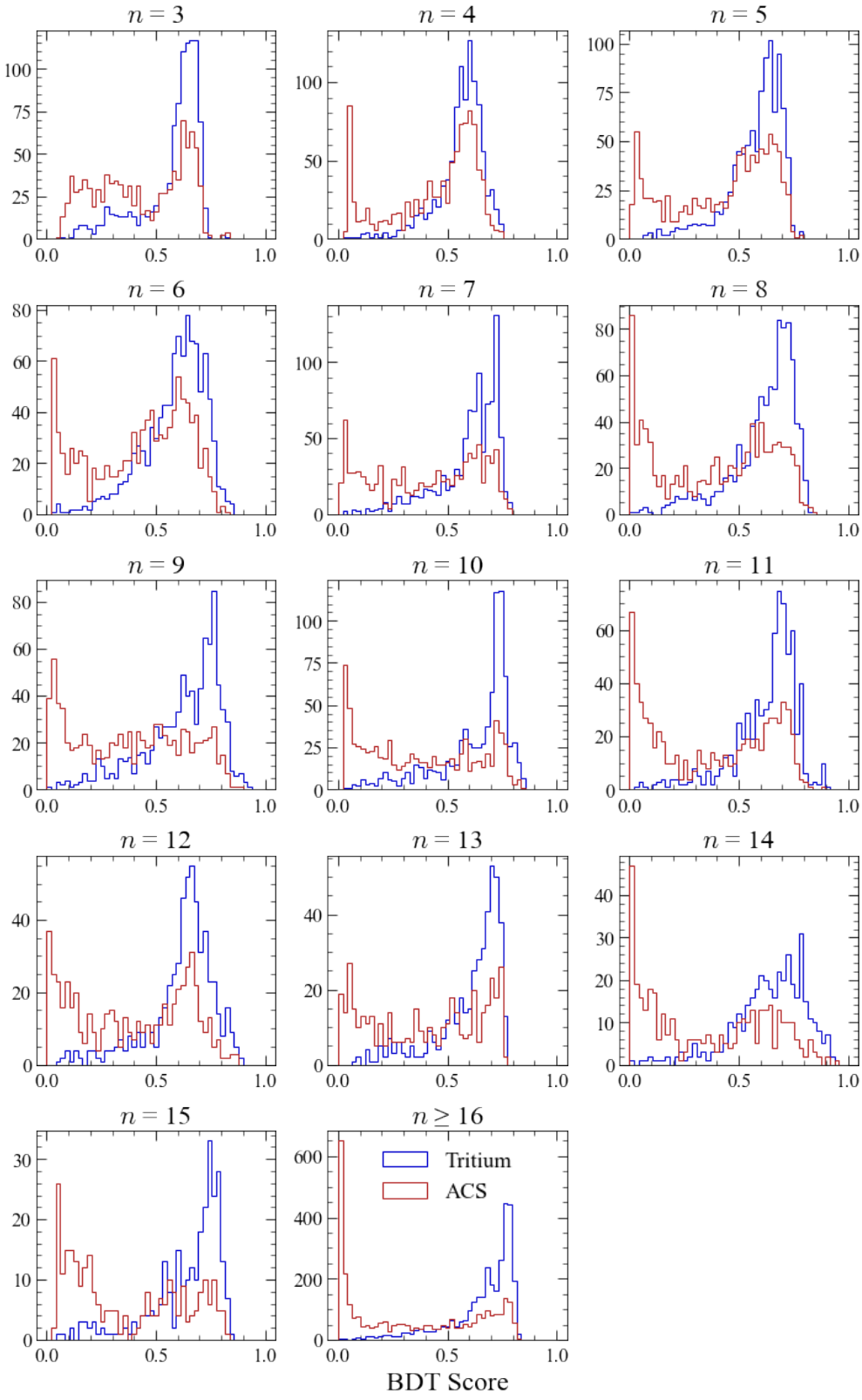


Figure 7.15: Distribution of BDT scores for each model, for all tritium signal and ACS background events, where n is the number of S1 photons.

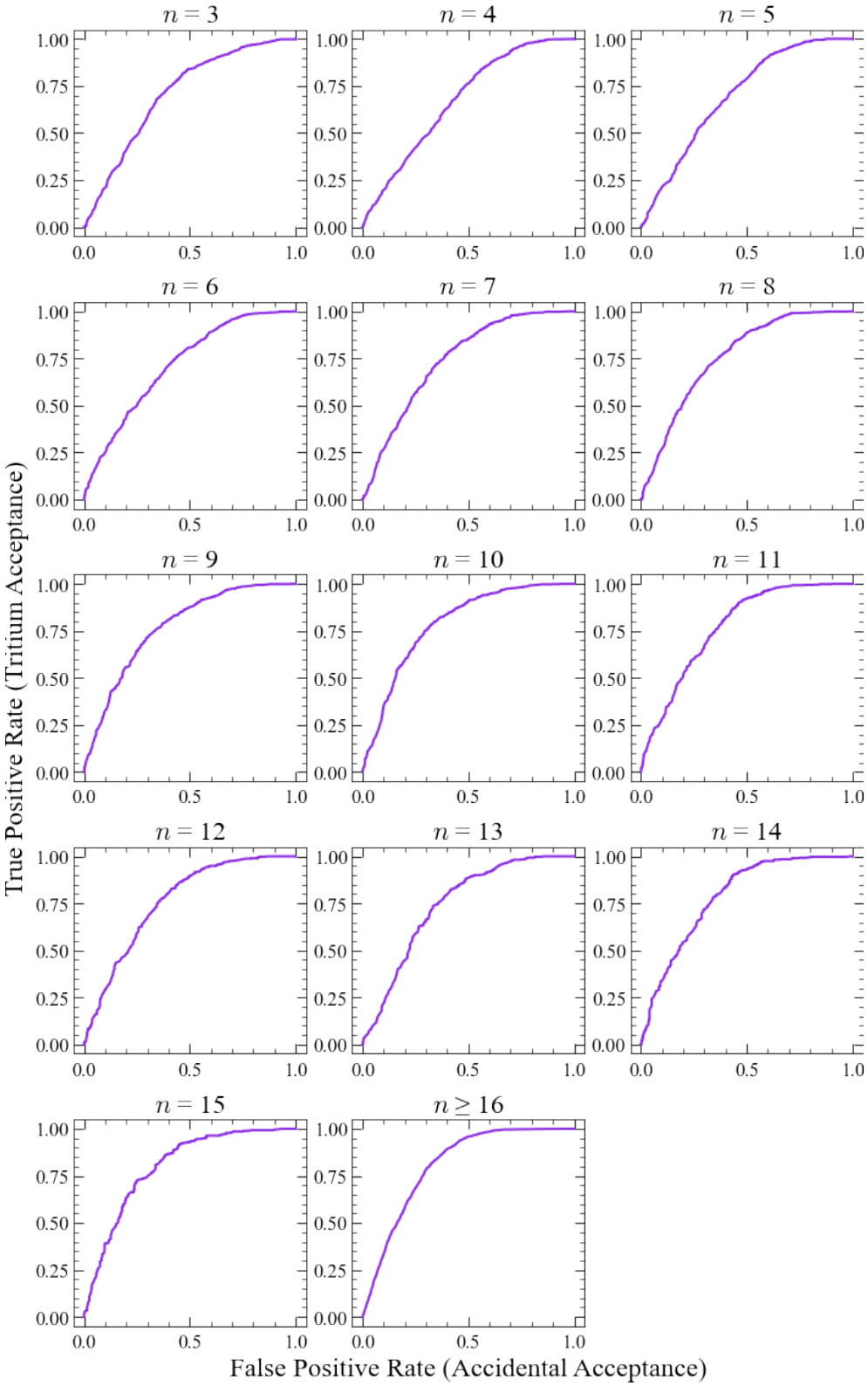


Figure 7.16: ROC curves showing tritium acceptance as a function of ACS acceptance for each BDT model, where n is the number of S1 photons.

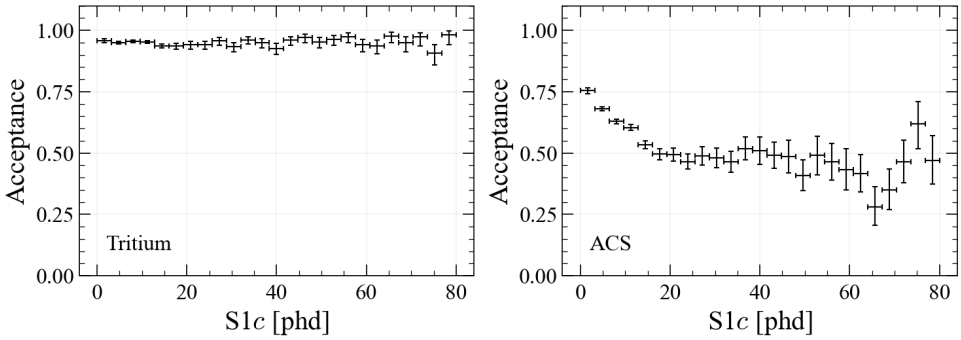


Figure 7.17: Tritium and ACS acceptance as a function of S1c, for a BDT cut where the tritium signal acceptance has been set to 95% for each model.

as the quantity on which to base an ordering principle, where n is the observed number of events [172]. R is the ratio of the probability of obtaining n for a true signal mean, μ, and the probability of obtaining n given the best-fit value of μ. For a known mean background, b, and a given value of μ, values of n are added to an acceptance region in order of decreasing R, until the sum of $P(n \mid \mu)$ meets the required confidence level, where here the standard value of 90% is used.

The optimal cut value depends on the exposure of the experiment, and can be estimated using a background-only Feldman-Cousins upper limit calculation. WIMP models for which the expected number of signal events exceeds this upper limit would be rejected at the 90% confidence level if no signal was observed.

One required input for this calculation is the signal acceptance. As this was designed to be approximately constant with S1c, it was applied as an overall percentage. This entered the calculation only as a penalty on the upper limit, leading to a weaker limit for a lower signal acceptance. The required quantity is the total signal acceptance for all data quality cuts on a pure (i.e. following livetime and physics cuts) signal dataset. The existing acceptance curve corresponded to the acceptance of the BDT cut alone, on a dataset with all other cuts applied. This was therefore required to be normalised by the acceptance of the other S1-based and S2-based cuts:

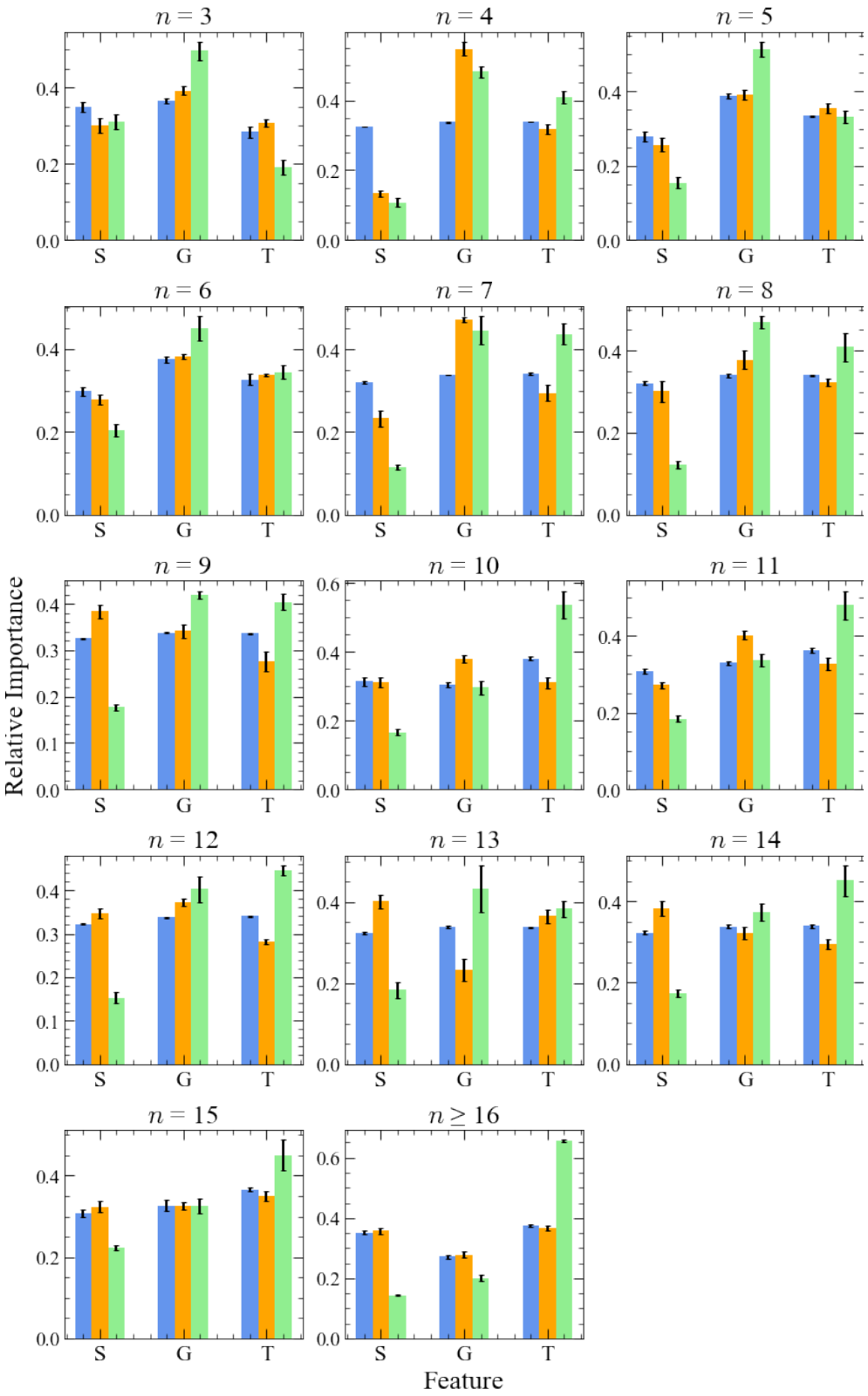


Figure 7.18: Relative importance, as defined by the cover (blue), weight (orange) and gain (green) metrics, of the shape-based ("S"), GOF ("G") and template-comparison ("T") input features in each of the 14 BDT models, where n is the number of S1 photons. The error bars show the standard deviation in the importance scores across the five models employed in k-fold cross validation.

$$\begin{aligned}&\text{Signal Acceptance} = \\ &\frac{\text{S1 Likelihood} + \text{Physics} + \text{Livetime} + \text{S1-based} + \text{S2-based}}{\text{Physics} + \text{Livetime}} \\ &= \frac{\text{S1 Likelihood} + \text{Physics} + \text{Livetime} + \text{S1-based} + \text{S2-based}}{\text{Physics} + \text{Livetime} + \text{S1-based} + \text{S2-based}} \\ &\times \frac{\text{Physics} + \text{Livetime} + \text{S1-based} + \text{S2-based}}{\text{Physics} + \text{Livetime}}\end{aligned} \tag{7.5.2}$$

The acceptance curve corresponding to the second term was already available as part of the standard SR3 analysis, and was again found to be approximately constant across S1c.

The calculation also requires the expected number of background events and the background rejection efficiency, both used to determine b. The expected number of background events was chosen to be the number of accidental events within the 1σ contour of a 40 GeV-mass WIMP during the total run time, following all other data quality cuts, as shown in Figure 7.19. For this study, the 40 GeV WIMP was selected as a representative example, as this is the region in which LZ has peak sensitivity. However, it would also be possible to tune the likelihood cut thresholds to target sensitivity improvements at different WIMP masses. For smaller masses, the distribution of signal events would be concentrated in the region dominated by accidentals, so the run would become background limited more quickly. A more aggressive cut would therefore be required, where livetime is sacrificed for improved background rejection. For higher masses, the WIMP contour extends further along the NR band, leading to less overlap with accidentals. A less aggressive cut, for which more livetime is preserved, would therefore be preferable.

As background rejection efficiency varies with S1c, the ACS acceptance curve was applied to the accidentals distribution, in order to calculate the overall percentage efficiency in the selected region.

For every value of signal acceptance tested, 10,000 toys were generated, with a Poisson distributed number of observed background events, n_0. The upper limit on μ was determined using the TFeldmanCousins class in ROOT, which queries tables of confidence intervals for combinations of b and n_0. The mean upper limit was plotted against the cut acceptance, and a polynomial fit was used to determine the minimum point, corresponding to the acceptance at which the weakest signals can be rejected.

The calculation was undertaken for two possible run times: 200 live-days, which is the approximate expected livetime for SR3, and 1000 live-

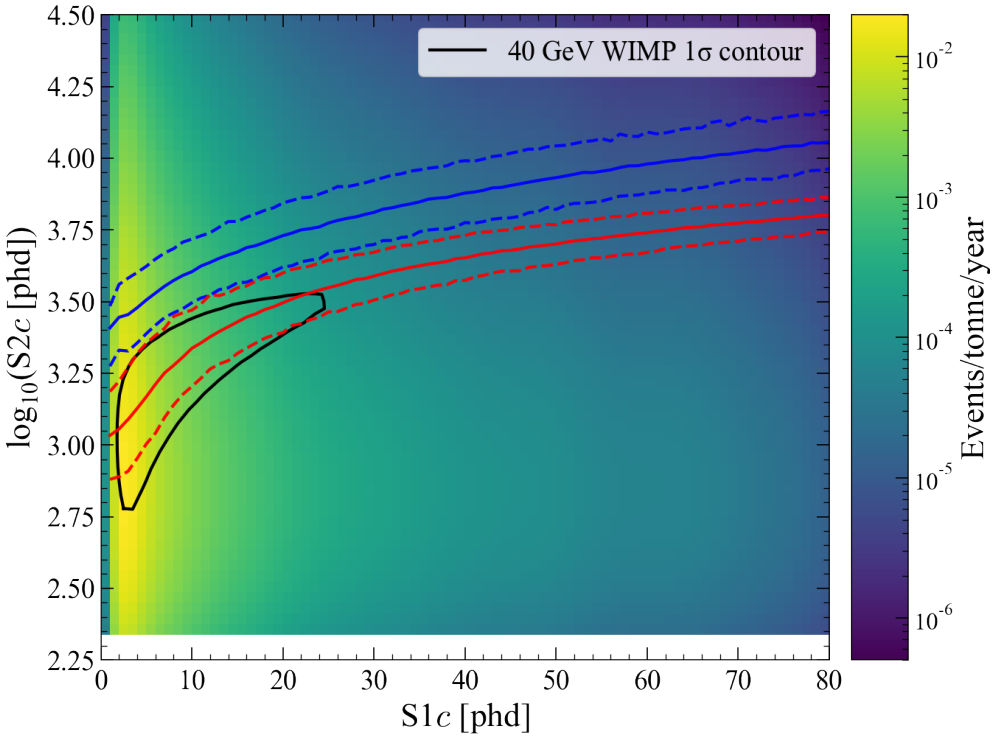


Figure 7.19: SR3 distribution of AccidentalChopStitch events in $\log_{10}$ S2c-S1c space following all data quality cuts and a smoothing procedure. The 1σ contour for a 40 GeV WIMP is shown in black.

days, which is the estimated total livetime of the LZ experiment. The latter will be in a regime in which accidental-coincidence backgrounds are more dominant, so a more aggressive cut is expected to be effective. The results for both the combined linear cuts and the BDT cut are shown in Figure 7.20, which also includes the total number of accidental events within the 40 GeV WIMP 1σ contour that survived each cut. For the 1000-liveday search, it can be seen that acceptances across the full range tested outperform the 100% signal acceptance scenario, which is equivalent to no cut. The optimal acceptance and the resulting mean upper limit are shown for each scenario in Table 7.2, together with the number of surviving accidentals. A small reduction in the upper limit and accidental counts can be seen for the BDT cut compared to the linear cuts.

7.6 Impact on WIMP Sensitivity

Sensitivity projections were then produced for each cut scenario, in order to demonstrate the potential impact on the WIMP search. It was decided that likelihood cut acceptances of 92% and 95% would be used, as the optimal values returned by the Feldman-Cousins study for 1000 and 200 livedays, respectively. Acceptances of 80% and 99% were also tested as

Table 7.2: The optimal likelihood cut acceptance which provides the lowest mean upper limit for each of the four scenarios described in the text. The number of accidental events within a 40 GeV WIMP 1σ contour that survive a cut with this acceptance is also shown.

Cut	Livetime [d]	Acceptance	Accidentals	Upper Limit
Linear	200	95.1%	1.51	3.81
BDT	200	95.2%	1.45	3.73
Linear	1000	91.7%	6.73	6.25
BDT	1000	91.5%	6.46	6.17

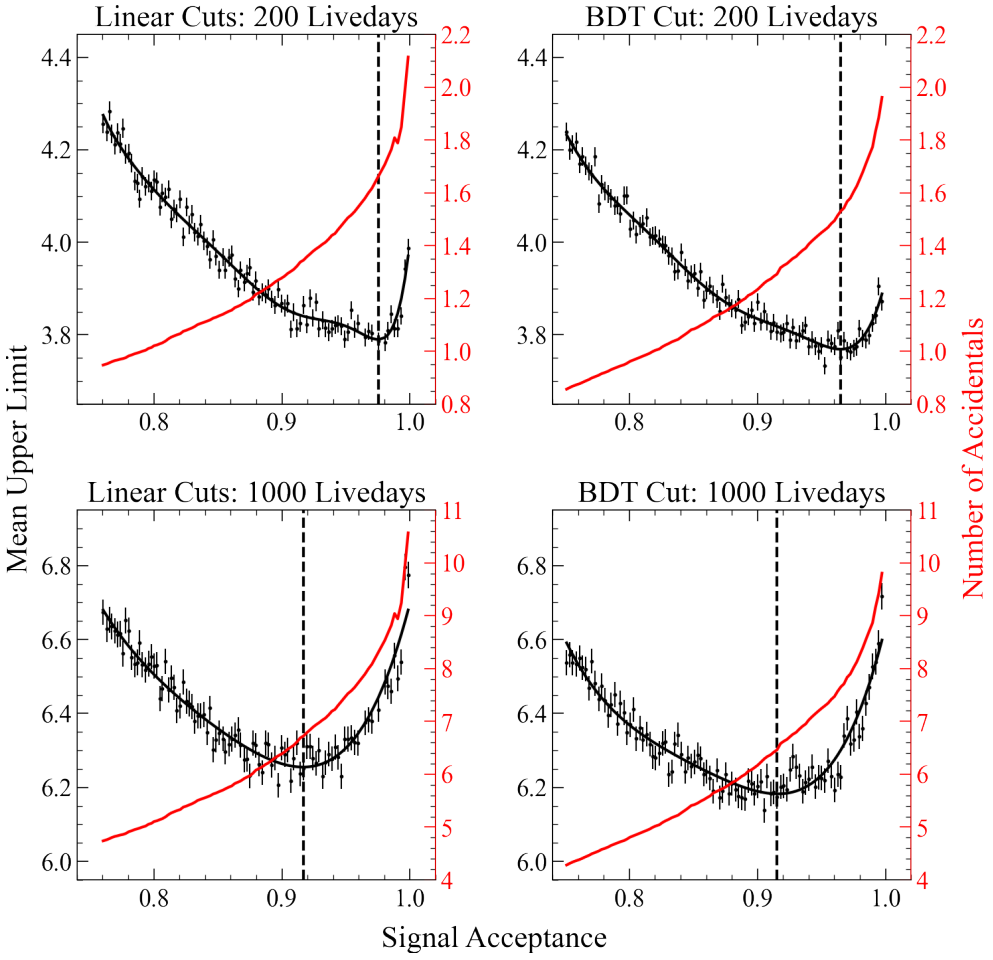


Figure 7.20: Mean upper limits determined using a Feldman-Cousins calculation for each value of total signal acceptance for the S1 likelihood cut. Combined linear cuts and a single cut on BDT score are shown for 200- and 1000-liveday searches. The error bars show the standard error across the different toys and the solid black line is a polynomial fit to the data, for which the minimum point is indicated by the black dashed line. The number of surviving accidental events within the chosen contour is shown in red.

extreme scenarios.

In the process of limit setting, a profile likelihood ratio is used to define a two-sided test statistic, $\tilde{t}_\mu$ for a signal strength, μ:

$$\lambda\left(\mu\right) = \frac{\mathcal{L}\left(\mu, \hat{\hat{\boldsymbol{\theta}}}\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\boldsymbol{\theta}}\right)} \tag{7.6.1}$$

$$\tilde{t}_\mu \equiv -2\ln\lambda(\mu), \tag{7.6.2}$$

in which the Feldman-Cousins method has been modified to include a set of nuisance parameters, $\boldsymbol{\theta}$. Here, a single hat corresponds to parameters that maximise the global likelihood, whereas a double hat indicates parameters that maximise the likelihood for a specific value of μ.

As prescribed in Ref. [147], the test statistic recommended for limit-setting in dark matter direct detection experiments is

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu) = \begin{cases} -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0, \\ -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0. \end{cases} \tag{7.6.3}$$

In the case of dark matter searches, the signal can only increase the event count, so μ is required to be positive. If $\hat{\mu} < 0$ for a certain dataset, then the best agreement between the data and a physical value of μ occurs at $\mu = 0$. The test statistic definition therefore ensures that if $\hat{\mu} < 0$, it will always be fixed to zero in the denominator of the likelihood ratio.

For limit setting, a signal-plus-background model forms the null hypothesis, H_0, and a background-only model forms the alternative hypothesis, H_1. For every WIMP mass tested, for a range of possible values of μ, Monte Carlo toys are used to generate distributions of test statistics for each hypothesis. When testing the level of agreement between H_0 and data, the p-value is defined as the integral of the H_0 distribution above the value of the test statistic observed in data, $\tilde{t}_{\mu,\mathrm{obs}}$:

$$p(\mu) = P\left(\tilde{t}_\mu \geq \tilde{t}_{\mu,\mathrm{obs}} \mid \mu\right) = \int_{\tilde{t}_{\mu,\mathrm{obs}}}^{\infty} f\left(\tilde{t}_\mu \mid \mu\right) d\tilde{t}_\mu. \tag{7.6.4}$$

The p-value is therefore the probability of obtaining a value of $\tilde{t}_\mu$ greater than $\tilde{t}_{\mu,\mathrm{obs}}$.

When evaluating a projected sensitivity, the median of the distribution for H_1 is determined, and replaces $\tilde{t}_{\mu,\mathrm{obs}}$ in the calculation of the p-value. For each WIMP mass, the value of μ for which the p-value is 0.1 is scaled to produce a limit on the WIMP-nucleon cross-section. The

p-values obtained when the median of the H_1 distribution is shifted by $\pm 1\sigma$ or 2σ are indicated by green and yellow bands, respectively, in limit plots, in order to quantify statistical fluctuations over repeated experiments. To evaluate the impact of the S1 likelihood cut on the WIMP search, this process was carried out using the Flamedisx package in the non-asymptotic regime [173], [174].

Templates for WIMPs, backgrounds and accidentals were produced for use as inputs for the statistical inference. It should be noted that at the time of this study, the official data quality cuts and background model were still undergoing development, so results will not be fully consistent with the final SR3 analysis. In particular, the S2 threshold for the WIMP search has since been raised from 4.5 electrons to 14.5 electrons in order to continue acquiring data for the search for a ^{8}B signal, while keeping the ^{8}B region blinded separately from the WIMP search. As this study considers only projected sensitivities and not the observed data, it does not interfere with the bias mitigation in place for SR3. A consequence of the lower S2 threshold is an additional population of accidentals, which are correlated in nature, meaning that a small spurious S1 and S2 arise from the same source, such as an e-/ph-train, and are misclassified as a Single Scatter. This population was not modelled in this investigation, but it is expected that the S1 likelihood cut would also target these events, leading to further improvement. The sources of tritium and ^{14}C, which contaminated early SR3 data, were also not included in the background model for this sensitivity study, but will be accounted for in the official analysis. The S1 likelihood cut acceptance curves for tritium were applied to the PDFs for signal and background sources, and the accidental acceptance curves were similarly applied to the ACS PDF, to produce templates for all cut scenarios.

The sensitivity projections for 200 and 1000 livedays are shown in figs. 7.21 and 7.22, compared to the case with no likelihood cut applied. As expected, for the 200-liveday search, the impact of the S1 likelihood cut was limited, with small visible improvements for the 95% and 99% cut acceptances, which are close to the minimum of the upper limit curves resulting from the Feldman-Cousins calculation. The most significant relative improvement on the most stringent limit from the baseline scenario was 3.9% for the BDT cut (with a cut acceptance of 99%) and 2.3% for the linear cuts (with a cut acceptance of 95%). As a cut with an acceptance of 80% produced a weaker upper limit than no cut in the Feldman-Cousins study (see Figure 7.20), it is unsurprising that in this scenario the median sensitivity worsens compared to the baseline case.

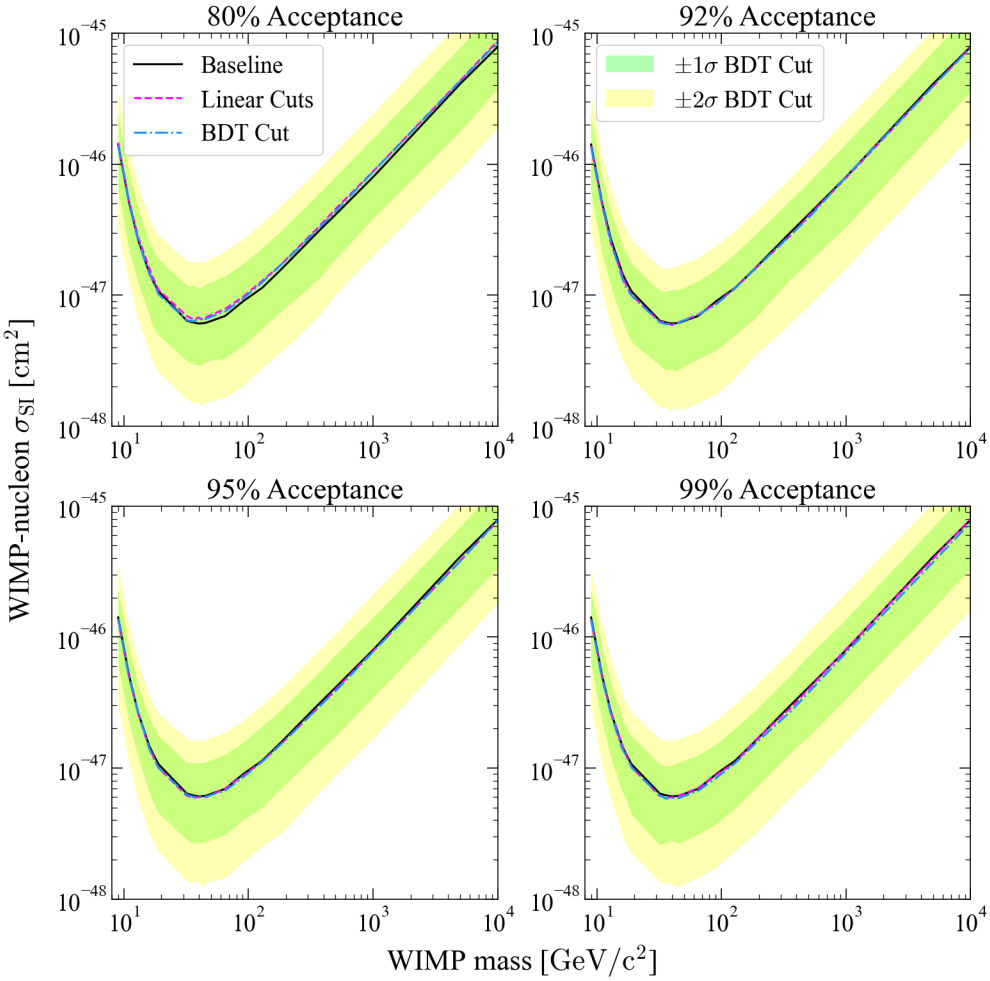


Figure 7.21: Sensitivity projections for a 200-liveday WIMP search, computed using Flamedisx. In the baseline scenario (black), no further cuts are applied in addition to the standard SR3 cuts. The pink line shows the impact of the S1 likelihood combined linear cuts and the blue line shows the BDT cut. 1σ and 2σ bands are also shown for the BDT cut.

The BDT cut outperforms the linear cuts here, as is generally the case. For 1000 livedays, the 92% cut acceptance has the biggest impact, resulting in an improvement of the most stringent limit for a 40 GeV WIMP from $2.12 \times 10^{-48}\,\mathrm{cm}^2$ for the baseline scenario to $1.97 \times 10^{-48}\,\mathrm{cm}^2$ for the BDT cut, an 8% decrease. The limit for an ideal scenario of 100% acceptance for the likelihood cut and 100% removal of accidentals is also shown. The most stringent limit in this case corresponds to a cross-section of $1.04 \times 10^{-48}\,\mathrm{cm}^2$, a 51.2% decrease from the baseline scenario, demonstrating the potential for significant further improvement by targeting this important background.

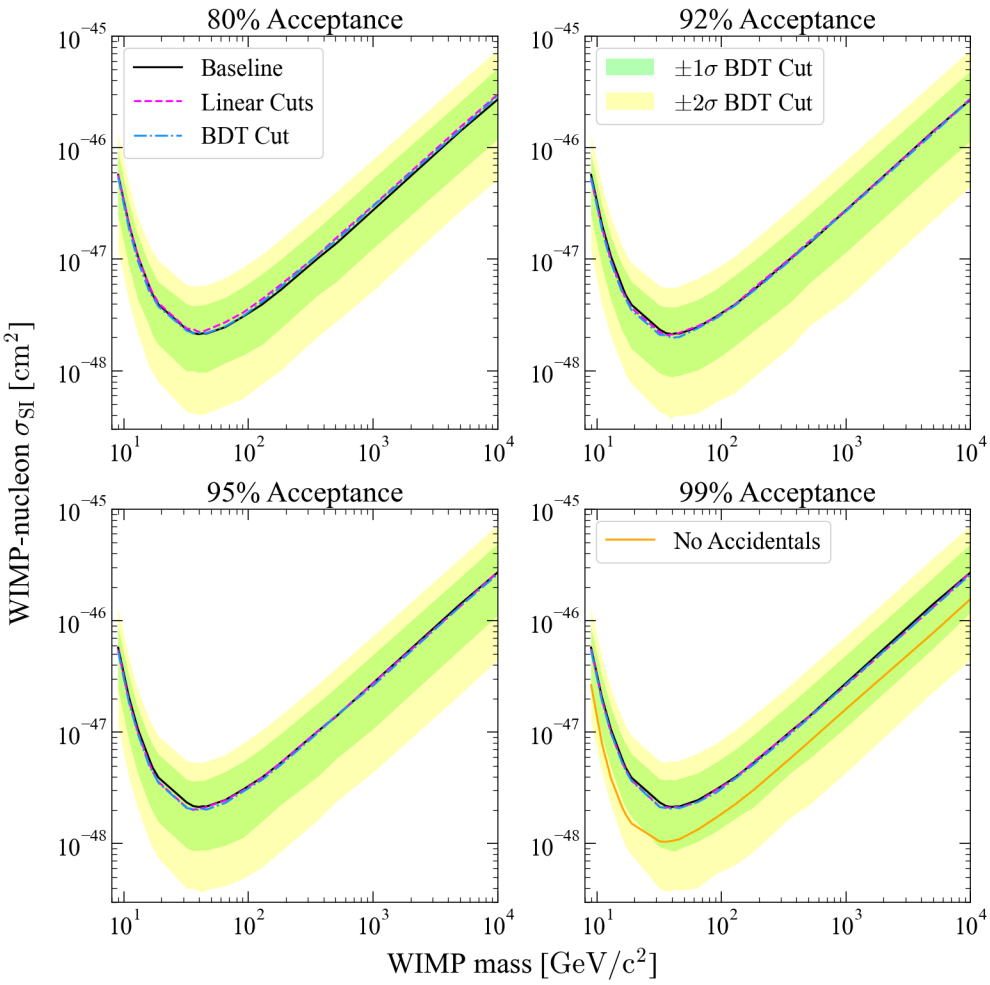


Figure 7.22: Sensitivity projections, as in Figure 7.21, but for a 1000-liveday WIMP search. The limit for an ideal scenario of 100% cut acceptance and 100% removal of accidentals is also shown.

7.7 Impact on WIMP Discovery Potential

In the case that a new signal is detected, the background only model becomes the null hypothesis, and a special case of the test statistic is used to test the significance at which it can be rejected in favour of H_1, the signal-plus-background model:

$$\tilde{t}_0 := q_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} \geq 0 \\ 0, & \hat{\mu} < 0 \end{cases} . \tag{7.7.1}$$

The p-value is then determined using:

$$p_0 = P\left(\tilde{t}_0 \geq \tilde{t}_{0,\,\mathrm{obs}} \mid 0\right) = \int_{\tilde{t}_{0,\,\mathrm{obs}}}^{\infty} f\left(\tilde{t}_0 \mid 0\right) d\tilde{t}_0 \tag{7.7.2}$$

The p-value can be expressed in terms of the number of standard deviations above the mean of a Gaussian distribution that would result in a probability of p_0 in the upper tail. 3σ and 5σ are the significances usually required to claim "evidence" or a "discovery", respectively, in particle physics. For a projected discovery calculation, a signal is injected into simulated background-only datasets, where the number of events is Poisson distributed about a mean value of the chosen μ. The median of the H_1 distribution is again used as the point above which the integral is calculated to determine the p-value.

For this investigation, masses of 9 GeV, 11 GeV and 40 GeV, corresponding to WIMP interactions at threshold, just above threshold, and at the optimal sensitivity, respectively, were chosen to demonstrate the impact of the S1 likelihood cut on the WIMP discovery potential. A cut acceptance of 92% for a 1000-liveday search was chosen to show the most significant impact. For each WIMP mass, a single value of $\mu = 15$ was chosen to obtain a discovery significance approaching 3σ[1]. This corresponds to WIMP-nucleon cross-sections of $8.23 \times 10^{-47}\,\mathrm{cm}^2$, $3.28 \times 10^{-47}\,\mathrm{cm}^2$ and $3.99 \times 10^{-48}\,\mathrm{cm}^2$ for the 9 GeV, 11 GeV and 40 GeV WIMPs, respectively. These values are realistic as they are below the current LZ limit, but within the design sensitivity goal of LZ of $3\times10^{-48}\,\mathrm{cm}^2$ in the WIMP parameter space. For each cut scenario, μ was scaled appropriately to maintain the same WIMP-nucleon cross-section, so that the relative improvement could be observed. Although both the combined linear and BDT cuts were chosen to have the same total acceptance across S1c, the differing shapes of the acceptance curves result in slightly different acceptances for the specific WIMP masses considered,

[1] This approach was adopted to avoid computationally-intensive scans of toy Monte Carlo trials for many values of μ.

resulting in different values of μ.

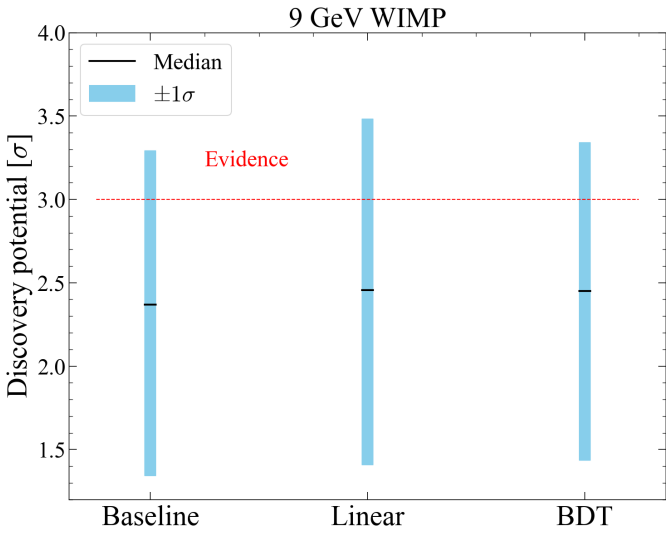
The distribution of values for the resulting discovery significance is shown in Figure 7.23. The most significant improvement is seen for a 40 GeV WIMP, where the median discovery potential increases from 2.58σ for the baseline scenario, to 2.70σ with the linear cuts applied, to 2.85σ for the BDT cut. This constitutes an 11% increase from the baseline to BDT scenarios. The chosen μ value, cross-section and resulting discovery potential for each scenario are presented in Table 7.3. This demonstrates how the cuts increase the potential for LZ to claim evidence for WIMPs, in the case of a detection. The effect is larger for the discovery potential than the sensitivity, which is expected when considering a population such as accidentals, which can directly overlap the signal. The impact of the background level on the upper limit becomes more significant for the smaller p-values associated with discovery tests compared to sensitivity studies. It should be noted that in a full study of discovery potential, the desired p-value would be chosen, and all μ values would be tested to determine the cross-section that results in this value.

In the event that a direct detection experiment observes an excess approaching 3σ, it is recommended that the "look elsewhere effect" (LEE) should be taken into account. This is the phenomenon which can occur when testing a null hypothesis against several different alternative hypotheses, such as different WIMP masses. A statistical fluctuation could be observed for any of these signal hypotheses, leading to an overestimation of the discovery significance. The impact of the LEE is dependent on both the number of hypotheses and the ease of distinguishing between them, and is expected to be small, particularly for WIMP masses greater than 40 GeV [147].

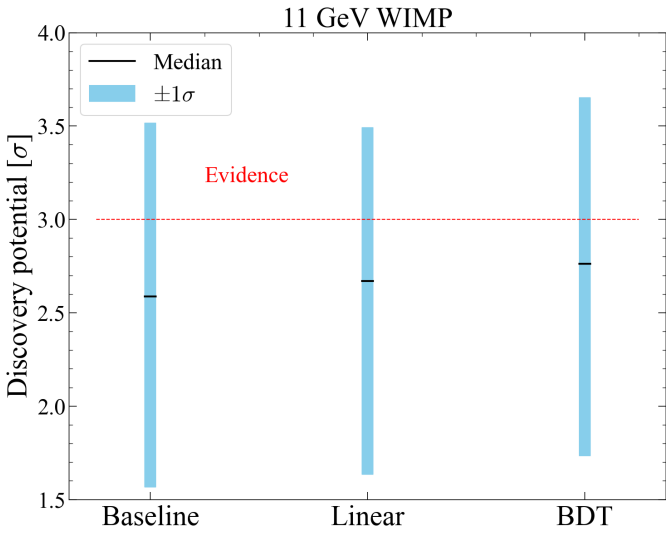
7.8 Future Improvements

The binned BDT implementation described enhances the sensitivity of the LZ experiment to WIMPs, but its performance could be further improved through more extensive tuning of the available hyperparameters, as well as by using a larger input dataset. The creation of separate BDT models for each value of $n_{photons}$ resulted in small training datasets, which could be boosted with more calibration data, as well as a larger production of ACS events.

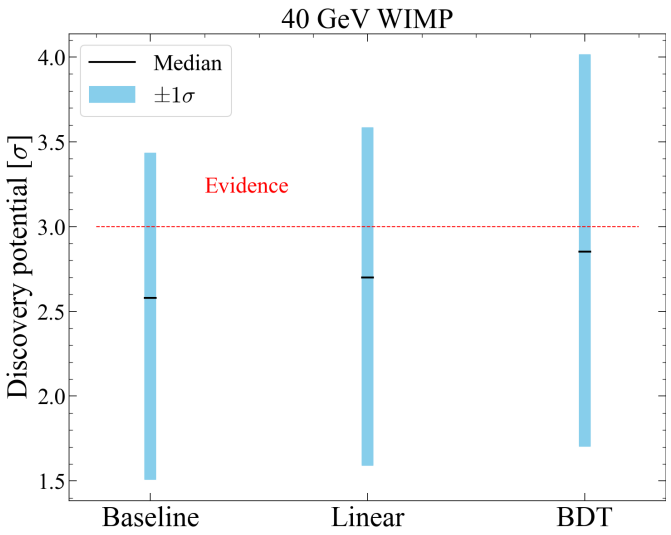
This BDT used only three input parameters, but there are many other pulse features that can be used to distinguish between signal events and accidentals. The same process could be followed with all pulse-based



(a)



(b)



(c)

Figure 7.23: Projected discovery potential of LZ for a 1000-liveday search, for scenarios with no additional cuts, combined linear cuts and a BDT cut. For each WIMP mass, a value of $\mu = 15$ was used for the baseline scenario, and was scaled to maintain the same cross-section for the other cases.

Table 7.3: The potential discovery significance for WIMP masses of 9 GeV, 11 GeV and 40 GeV, with no likelihood cut (baseline), combined linear cuts, and a single BDT cut. The number of observed signal events μ was chosen to be 15 for the baseline scenario, and then adjusted to maintain the same cross-section when each of the cuts was applied.

Mass [GeV]	Cut	μ	Cross-section [cm^2]	Discovery Significance
9	Baseline	15	8.23×10^{-47}	2.37σ
9	Linear	13.5	8.23×10^{-47}	2.46σ
9	BDT	13.9	8.23×10^{-47}	2.45σ
11	Baseline	15	3.28×10^{-47}	2.59σ
11	Linear	13.5	3.28×10^{-47}	2.67σ
11	BDT	13.9	3.28×10^{-47}	2.76σ
40	Baseline	15	3.99×10^{-48}	2.58σ
40	Linear	13.5	3.99×10^{-48}	2.70σ
40	BDT	13.8	3.99×10^{-48}	2.85σ

event selection implemented within the BDT model, rather than as individual cuts that are applied in advance. "Feature engineering" would be carried out to select physically-motivated input variables, which generally have significant overlap with those used in WIMP search data quality cuts. An example of this approach can be seen in Ref. [170], in which ten input parameters, primarily describing the shape and size of S1 and S2 pulses, were utilised to increase removal of accidental-coincidence events in SR1 data. The addition of the three S1 likelihood input variables to a model such as this would further improve rejection of accidentals, and thus increase sensitivity to WIMPs.

7.9 Summary

In chapter 6, three test statistics were introduced with the purpose of targeting accidentals based on arrival times of S1 photons. Rejection of events based on linear cuts on each quantity resulted in a loss of sensitivity due to overlap between the distributions for signal and background events. A boosted decision tree can be employed to consider all three metrics when classifying events, and facilitates optimisation of cut acceptance and accidental rejection efficiency through the tuning of a single cut on BDT score.

Initial investigations revealed interesting behaviour of the XGBoost model. In particular, quantisation of test statistic values, a general fea-

ture of the binned likelihood calculation, resulted in a reduced acceptance for events with a low number of S1 photons. As well as this, a population of accidental events were found to produce the highest values for the shape-based test statistic. Upon further investigation, evidence indicated that these were narrow S1s resulting from Cherenkov light in PMT windows.

It was therefore decided that several BDT models would be trained separately for each value of $n_{photons}$ below 15, in order to remove the effects of test statistic quantisation. All other data quality cuts would be applied on the input data to target pathologies such as Cherenkov light, and demonstrate the additional impact of the BDT cut beyond the existing WIMP search analysis. The resulting signal acceptance showed a reduced variation with S1c in comparison to the linear cuts.

In order to determine the impact of the BDT cut on the WIMP search, a Feldman-Cousins calculation was used to calculate the optimal cut acceptance for 200- and 1000-liveday searches. For the 1000-liveday case, a BDT cut with a signal acceptance of 92% resulted in an improvement of the projected discovery potential of up to 11% and of the most stringent exclusion limit by up to 8%, compared to the case with no additional cuts applied. An ideal scenario, in which all accidental-coincidence events were removed, was observed to have a significant impact on the WIMP sensitivity, so it is essential to continue to target this background as effectively as possible. Further improvement can be expected through larger training datasets, more extensive tuning of hyperparameters, and, most importantly, inclusion of additional input features.

Chapter 8

Conclusions

The LZ direct detection experiment set world-leading limits on spin-independent WIMP-nucleon interactions during its first science run. This was achieved through the detection of primary and secondary scintillation signals in a dual-phase xenon TPC, as described in chapter 3. The Physics Readiness Monitor was used to check for problems in the data, and new tools were developed to facilitate this. During validation of SR2 data, Algorithm Groups and two-dimensional plot comparisons were used in the identification of both long-term trends and brief irregularities. The TPC_Health module was employed to observe features such as xenon activation, temporal variation of pulse sizes, residual ^{83m}Kr, hotspots, and light emission detected in the Skin. This enabled the compilation of a list of acquisitions suitable for the WIMP search analysis. Whilst SR2 was terminated due to detector instabilities, the monitoring tools and validation procedures developed in this thesis will be key to evaluating the data quality for ongoing and future physics runs.

The tuning of optical simulations to reproduce the position-dependent light collection efficiency in SR1 data was described in chapter 5. Three optical parameters, the LXe-PTFE reflectivity, LXe absorption length and LXe Rayleigh scattering length, were found to yield the most significant impact on the LCE. These were optimised in a grid search, which used a χ^2 comparison to data. After accounting for issues with the models for ^{83m}Kr photon yields and secondary PMT effects, an improved match to data was achieved throughout the detector for the total LCE. However, it was found that separate consideration of photon counts in each PMT array was required to produce accurate top and bottom array LCE curves. The optimal Rayleigh scattering length of 0.43 m was selected and, as the reflectivity and absorption length were shown to be degenerate, a sensible pair of values of 97.3% and 121 m were chosen for these parameters, based on an external measurement of the reflectivity.

The tuned parameters provided good results in every region of the detector, apart from close to the bottom. Several potential solutions were suggested, including adjustments to the refractive index of the quartz PMT windows or the gaseous xenon. Pulses were observed to be slightly narrower in simulations than data, so future tuning should take this into consideration. These tuned simulations will be advantageous for analysis of complex pathologies and development of waveform-based cuts in the future.

A new waveform-based data quality cut was designed to target accidental coincidences of lone S1s and S2s, which form one of the most significant backgrounds to the WIMP search. This was the focus of chapter 6 and chapter 7. An offset of up to ~ 45 ns was observed between photon arrival times in the top and bottom PMT arrays, which was strongly correlated to the drift time of an event. This could be used to identify accidentals, for which the drift time is assigned randomly and is therefore not meaningful. As well as this, the shapes of pulses classified as S1s resulting from the pile-up of dark counts or SPEs are often irregular. Data-driven S1 templates were generated for slices of drift time throughout the detector, and three test statistics were designed to exploit both of these identifying features. The test statistics consisted of an S1 goodness of fit test, a comparison of S1 templates at various drift times, and a ratio of likelihoods for S1 and flat rate hypotheses. Test statistic values were calculated for a tritium signal and a background dataset consisting of AccidentalChopStitch (ACS); artificially-paired S1s and S2s used to model accidental-coincidence events.

Following normalisation to account for a dependence on the number of S1 photons, linear cuts were applied to reject ACS events. Investigations conducted using SR1 data showed that the pulse area and interaction type would not have a significant impact on the template shape. Each cut was then tuned on SR3 data to produce a signal acceptance of 95%. In the case where only baseline selection cuts were applied to the signal and background datasets, the acceptance of the combined likelihood cuts on the signal sample was 87.4%, with a background rejection efficiency of 47.1%. When all other SR3 data quality cuts were applied to the signal and background samples prior to tuning of the cut parameters, a combined likelihood cut acceptance of 87.6% and an accidental rejection efficiency of 42.4% were achieved. The cuts removed accidental-coincidence events across the full range of S1c within the WIMP search region of interest. This included pathologies such as pile-up of SPEs or S1s, misclassified pulses, and events with an incorrect assigned drift time.

It has therefore been demonstrated that the S1 likelihood cut is able to effectively remove accidentals. Correlation was observed between the three individual cuts, indicating that simultaneous consideration of the three test statistics would lead to improved background rejection efficiency for the same signal loss.

This was achieved through the use of a boosted decision tree, for which the three test statistics were used as input features. The BDT assigned a score to events, with 0 and 1 as target scores for background and signal, respectively. A single cut was then placed on the BDT score to achieve the desired signal acceptance. A preliminary BDT performed well, and models with a single test statistic as the only input feature were used to validate this. The BDT using the shape-based test statistic achieved a background rejection efficiency that was significantly better than the equivalent linear cut, prompting further investigation. Quantisation of test statistic values for events with a low S1 photon count resulted in a reduced signal acceptance for this population. To avoid this, several BDT models were trained separately for events with a different number of S1 photons. A population of accidental-coincidence events with a high test statistic value was also observed, which, upon further investigation, was deemed likely to be Cherenkov light. Although it was logical that these events should be rejected by the BDT cut, it was decided that all other SR3 data quality cuts would be applied to datasets before training the BDT, to exploit existing cuts that target specific pathologies, such as Cherenkov light. Hyperparameters were tuned separately for each model, employing k-fold cross validation. The BDT models achieved a total accidental rejection efficiency of 48.1% for an 87.6% signal acceptance, compared to 42.4% for the combined linear cuts. The BDT score cut values for each model were selected to ensure that acceptance was approximately constant across S1c. This was superior to the combined linear cuts, for which a dip in acceptance is present at low S1c.

In order to evaluate the impact on the WIMP search of both the linear and BDT approaches to the S1 likelihood cut, a Feldman-Cousins calculation was used to determine the optimal cut acceptance. The background signal was chosen to be the surviving accidental-coincidence events within the 1σ contour of a 40 GeV WIMP, following all other data quality cuts. For a 200-liveday search, the optimal cut acceptance was found to be $\sim$95%, corresponding to a reduction of accidentals in the chosen region from 2.15 to 1.51 for the linear cuts and 1.45 for the BDT cut. For a 1000-liveday search, a more aggressive cut of $\sim$92% was found to be optimal, which reduced the number of accidentals in the chosen region from

10.74 to 6.73 for the linear cuts and 6.46 for the BDT cut.

Sensitivity projections were then produced, demonstrating that application of the BDT cut could improve the sensitivity of LZ to WIMPs by up to 8% for a 1000-liveday search, relative to the baseline scenario. For a 40 GeV WIMP with an SI WIMP-nucleon cross-section of $3.99 \times 10^{-48}\,\mathrm{cm}^2$, the BDT cut increased the projected discovery potential by 11% for a 1000-liveday search. The significant improvement to WIMP sensitivity when all accidentals were removed (a 51.2% decrease in cross-section for the most stringent limit) showed that it is vital to continue to target this background. A future comprehensive BDT analysis with additional input features would increase the potential of the LZ experiment to either exclude further parameter space, or claim a WIMP discovery. Other avenues for future developments for both the linear and BDT versions of the S1 likelihood cut include the incorporation of uncertainties in the S1 templates and the separation of S1 templates by interaction type, depending on the physics search. The XLZD next-generation experiment will also require efficient removal of accidental-coincidence backgrounds, as it explores the remainder of the WIMP parameter space before reaching the neutrino fog. The framework and techniques developed here represent an important first step in addressing this challenge. If no signal is observed, this remaining parameter space can be excluded, and the focus of direct detection searches will turn to directional detectors which can overcome the neutrino fog, and searches for other dark matter candidates beyond WIMPs.

Bibliography

[1] B. Allanach, A. Bevan, M. D'Onofrio, *et al.*, Particle Physics Advisory Panel, "PPAP 2021 Roadmap," 2021.

[2] L. Hu and L. Li, "Using Tree-Based Machine Learning for Health Studies: Literature Review and Case Series," *Int J Environ Res Public Health*, vol. 19, 2022.

[3] L. Grunwald, A.-K. Schneider, B. Schröder, and S. Weber, "Predicting urban cold-air paths using boosted regression trees," *Landscape and Urban Planning*, vol. 201, p. 103 843, 2020.

[4] J. Cooley *et al.*, "Report of the Topical Group on Particle Dark Matter for Snowmass 2021," *arXiv 2209.07426*, Sep. 2022.

[5] N. Aghanim, Y. Akrami, M. Ashdown, *et al.*, Planck Collaboration, "Planck 2018 results: VI. Cosmological parameters," *Astronomy and Astrophysics*, vol. 641, A6, Sep. 2020.

[6] F. Zwicky, "Republication of: The redshift of extragalactic nebulae.," *Gen Relativ Gravit*, vol. 41, pp. 207–224, Jan. 2009.

[7] R. Gavazzi, T. Treu, J. D. Rhodes, *et al.*, "The Sloan Lens ACS Survey. IV. The Mass Density Profile of Early-Type Galaxies out to 100 Effective Radii," *The Astrophysical Journal*, vol. 667, no. 1, pp. 176–190, Sep. 2007.

[8] P. Natarajan, U. Chadayammuri, M. Jauzac, *et al.*, "Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations," *Monthly notices of the Royal Astronomical Society.*, vol. 468, no. 2, pp. 1962–1980, Jun. 2017.

[9] D. Coe, N. Benítez, T. Broadhurst, and L. A. Moustakas, "A HIGH-RESOLUTION MASS MAP OF GALAXY CLUSTER SUBSTRUCTURE: LensPerfect ANALYSIS OF A1689," *The Astrophysical Journal*, vol. 723, no. 2, pp. 1678–1702, Oct. 2010.

[10] D. Clowe, M. Bradač, A. H. Gonzalez, *et al.*, "A Direct Empirical Proof of the Existence of Dark Matter," *The Astrophysical Journal*, vol. 648, no. 2, p. L109L113, Aug. 2006.

[11] V. C. Rubin and W. K. Ford Jr, "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions," *The Astrophysical Journal*, vol. 159, p. 379, Feb. 1970.

[12] K. G. Begeman, A. H. Broeils, and R. H. Sanders, "Extended rotation curves of spiral galaxies: dark haloes and modified dynamics," *Monthly Notices of the Royal Astronomical Society*, vol. 249, no. 3, pp. 523–537, Apr. 1991.

[13] A. Bosma, "21-cm line studies of spiral galaxies. II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types.," *Astronomical Journal*, vol. 86, pp. 1825–1846, Dec. 1981.

[14] V. C. Rubin, W. K. Ford Jr., and N. Thonnard, "Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc," *Astrophys. J. Lett.*, vol. 225, pp. L107–L111, 1978.

[15] K. Freese, "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars," *EAS Publications Series*, vol. 36, pp. 113–126, 2009.

[16] A. Liddle, *An Introduction to Modern Cosmology*, 2nd ed. Chichester: Wiley, 2003, ch. 10.

[17] G. F. Smoot, C. L. Bennett, A. Kogut, *et al.*, "Structure in the COBE Differential Microwave Radiometer First-Year Maps," *The Astrophysical Journal*, vol. 396, p. L1, Sep. 1992.

[18] S. Dodelson and F. Schmidt, *Modern Cosmology*, 2nd ed. Academic Press, 2021, ch. 9.

[19] Y. Akrami, M. Ashdown, and J. Aumont, Planck Collaboration, "Planck 2018 results: IV. Diffuse component separation," *Astronomy and Astrophysics*, vol. 641, A4, 2020.

[20] W. Hu, N. Sugiyama, and J. Silk, "The physics of microwave background anisotropies," *Nature*, vol. 386, no. 6620, pp. 37–43, 1997.

[21] M. Taoso, G. Bertone, and A. Masiero, "Dark Matter Candidates: A Ten-Point Test," *JCAP*, vol. 03, p. 022, 2008.

[22] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees, "Formation of galaxies and large-scale structure with cold dark matter," *Nature*, vol. 311, no. 5986, pp. 517–525, 1984.

[23] P. Bode, J. P. Ostriker, and N. Turok, "Halo Formation in Warm Dark Matter Models," *The Astrophysical Journal*, vol. 556, no. 1, pp. 93–107, Jul. 2001.

[24] A. Coc and E. Vangioni, "Primordial nucleosynthesis," *International Journal of Modern Physics E*, vol. 26, no. 08, p. 1 741 002, 2017.

[25] S. Burles, K. M. Nollett, and M. S. Turner, "Big Bang Nucleosynthesis Predictions for Precision Cosmology," *The Astrophysical Journal*, vol. 552, no. 1, pp. L1–L5, May 2001.

[26] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradač, "Constraints on the Self-Interaction Cross Section of Dark Matter from Numerical Simulations of the Merging Galaxy Cluster 1E 0657–56," *The Astrophysical Journal*, vol. 679, no. 2, p. 1173, Jun. 2008.

[27] J. Lesgourgues and S. Pastor, "Neutrino Mass from Cosmology," *Advances in High Energy Physics*, vol. 2012, pp. 1–34, 2012.

[28] M. Aker, A. Beglarian, J. Behrens, *et al.*, KATRIN Collaboration, "Direct neutrino-mass measurement with sub-electronvolt sensitivity," *Nature Physics*, vol. 18, no. 2, pp. 160–166, 2022.

[29] S. Dodelson and L. M. Widrow, "Sterile neutrinos as dark matter," *Phys. Rev. Lett.*, vol. 72, no. 1, pp. 17–20, Jan. 1994.

[30] M. Shaposhnikov, "Sterile neutrinos," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 228–248.

[31] 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.

[32] G. G. Raffelt, "Astrophysical axion bounds," in *Axions: Theory, Cosmology, and Experimental Searches*, M. Kuster, G. Raffelt, and B. Beltrán, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 51–71.

[33] P. Sikivie, "Axions," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 228–248.

[34] R. Khatiwada, D. Bowring, A. S. Chou, *et al.*, "Axion Dark Matter Experiment: Detailed design and operations," *Review of Scientific Instruments*, vol. 92, no. 12, Dec. 2021.

[35] E. Aprile, K. Abe, F. Agostini, *et al.*, XENON Collaboration, "Search for New Physics in Electronic Recoil Data from XENONnT," *Phys. Rev. Lett.*, vol. 129, p. 161 805, 16 Oct. 2022.

[36] J. Aalbers, D. S. Akerib, A. K. Al Musalhi, *et al.*, LZ Collaboration, "Search for new physics in low-energy electron recoils from the first LZ exposure," *Phys. Rev. D*, vol. 108, p. 072 006, 7 Oct. 2023.

[37] K. Griest, "WIMPs and MACHOs," in *Encyclopedia of Astronomy and Astrophysics*, P. Murdin, Ed., Boca Raton, 2002, E2634.

[38] C. Alcock, R. A. Allsman, D. R. Alves, *et al.*, "The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations," *The Astrophysical Journal*, vol. 542, no. 1, pp. 281–307, Oct. 2000.

[39] P. Villanueva-Domingo, O. Mena, and S. Palomares-Ruiz, "A Brief Review on Primordial Black Holes as Dark Matter," *Frontiers in Astronomy and Space Sciences*, vol. 8, 2021.

[40] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, LIGO Scientific Collaboration and Virgo Collaboration, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.*, vol. 116, p. 061 102, 6 Feb. 2016.

[41] S. Bird, I. Cholis, J. B. Muñoz, *et al.*, "Did LIGO Detect Dark Matter?" *Phys. Rev. Lett.*, vol. 116, p. 201 301, 20 May 2016.

[42] A. Yacine, E. D. Kovetz, and M. Kamionkowski, "Merger rate of primordial black-hole binaries," *Phys. Rev. D*, vol. 96, p. 123 523, 12 Dec. 2017.

[43] B. J. Kavanagh, D. Gaggero, and G. Bertone, "Merger rate of a subdominant population of primordial black holes," *Phys. Rev. D*, vol. 98, p. 023 536, 2 Jul. 2018.

[44] B. Famaey and S. S. McGaugh, "Modified Newtonian Dynamics (MOND): Observational Phenomenology and Relativistic Extensions," *Living Reviews in Relativity*, vol. 15, no. 1, p. 10, 2012.

[45] M. Milgrom, "Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis," *The Astrophysical Journal*, vol. 270, Jul. 1983.

[46] K.-H. Chae, "Breakdown of the Newton–Einstein Standard Gravity at Low Acceleration in Internal Dynamics of Wide Binary Stars," *The Astrophysical Journal*, vol. 952, no. 2, p. 128, Jul. 2023.

[47] X. Hernandez, "Internal kinematics of Gaia DR3 wide binaries: anomalous behaviour in the low acceleration regime," *Monthly Notices of the Royal Astronomical Society*, vol. 525, no. 1, pp. 1401–1415, Aug. 2023.

[48] I. Banik, C. Pittordis, W. Sutherland, *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.

[49] G. Gelmini and P. Gondolo, "DM production mechanisms," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 121–141.

[50] D. G. Cerdeño, "DARK MATTER 101: From production to detection," 2016.

[51] J. Ellis and K. A. Olive, "Supersymmetric dark matter candidates," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 142–163.

[52] O. Buchmueller, C. Doglioni, and L.-T. Wang, "Search for dark matter at colliders," *Nature Physics*, vol. 13, no. 3, pp. 217–223, Mar. 2017.

[53] J. Abdallah, H. Araujo, A. Arbey, *et al.*, "Simplified models for dark matter searches at the LHC," *Physics of the Dark Universe*, vol. 9–10, pp. 8–23, Sep. 2015.

[54] A. Boveia and C. Doglioni, "Dark matter searches at colliders," *Annual Review of Nuclear and Particle Science*, vol. 68, no. 1, pp. 429–459, Oct. 2018.

[55] G. Aad, B. Abbott, D. C. Abbott, *et al.*, ATLAS Collaboration, "Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Rev. D*, vol. 103, p. 112 006, 11 Jun. 2021.

[56] G. Aad, B. Abbott, D. Abbott, *et al.*, ATLAS Collaboration, "Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett. B*, vol. 829, p. 137 066, Jun. 2022.

[57] A. M. Sirunyan *et al.*, CMS Collaboration, "Search for dark matter produced in association with a leptonically decaying Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 81, no. 1, p. 13, 2021.

[58] M. Aaboud, G. Aad, B. Abbott, and J. Abdallah, ATLAS Collaboration, "Search for dark matter in association with a Higgs boson decaying to b-quarks in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett. B*, vol. 765, pp. 11–31, Feb. 2017.

[59] L. Bergström and G. Bertone, "Gamma-rays," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 491–506.

[60] L. Goodenough and D. Hooper, "Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope," *arXiv 0910.2998*, Oct. 2009.

[61] H. Abdalla, F. Aharonian, F. A. Benkhali, *et al.*, H.E.S.S. Collaboration, "Search for Dark Matter Annihilation Signals in the H.E.S.S. Inner Galaxy Survey," *Phys. Rev. Lett.*, vol. 129, p. 111 101, 11 Sep. 2022.

[62] O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, *et al.*, "An anomalous positron abundance in cosmic rays with energies 1.5-100GeV," *Nature*, vol. 458, no. 7238, pp. 607–609, Apr. 2009.

[63] M. Aguilar *et al.*, AMS Collaboration, "Antiproton Flux, Antiproton-to-Proton Flux Ratio, and Properties of Elementary Particle Fluxes in Primary Cosmic Rays Measured with the Alpha Magnetic Spectrometer on the International Space Station," *Phys. Rev. Lett.*, vol. 117, p. 091 103, 9 Aug. 2016.

[64] P. J. Fox and E. Poppitz, "Leptophilic dark matter," *Phys. Rev. D*, vol. 79, no. 8, Apr. 2009.

[65] I. Cholis, T. Linden, and D. Hooper, "A robust excess in the cosmic-ray antiproton spectrum: Implications for annihilating dark matter," *Phys. Rev. D*, vol. 99, no. 10, May 2019.

[66] R. K. Leane, "Indirect Detection of Dark Matter in the Galaxy," in *3rd World Summit on Exploring the Dark Side of the Universe*, 2020, pp. 203–228.

[67] M. G. Aartsen, R. Abbasi, Y. Abdou, *et al.*, IceCube Collaboration, "Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector," *Science*, vol. 342, no. 6161, p. 1 242 856, 2013.

[68] B. Feldstein, A. Kusenko, S. Matsumoto, and T. T. Yanagida, "Neutrinos at IceCube from heavy decaying dark matter," *Phys. Rev. D*, vol. 88, no. 1, Jul. 2013.

[69] T. Cohen, K. Murase, N. L. Rodd, B. R. Safdi, and Y. Soreq, "γ-ray Constraints on Decaying Dark Matter and Implications for IceCube," *Phys. Rev. Lett.*, vol. 119, no. 2, Jul. 2017.

[70] S. Ando and K. Ishiwata, "Constraints on decaying dark matter from the extragalactic gamma-ray background," *JCAP*, vol. 2015, no. 05, p. 024, May 2015.

[71] D. Cerdeño and A. Green, "Direct Detection of WIMPS," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 347–369.

[72] D. Cerdeño, C. Cuesta, M. Fornasa, *et al.*, "Complementarity of dark matter direct detection: the role of bolometric targets," *JCAP*, vol. 2013, no. 07, pp. 028–028, Jul. 2013.

[73] A. Al Musalhi, "Searches for rare nuclear decays in the LUX-ZEPLIN experiment," Ph.D. dissertation, University of Oxford, 2023.

[74] J. Aalbers *et al.* "JelleAalbers/wimprates." version 0.5.0. (2023), [Online]. Available: `https://zenodo.org/records/7636982`.

[75] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, and Y. Xu, "Model Independent Direct Detection Analyses," *arXiv 1211.2818*, 2012.

[76] M. Schumann, "Direct detection of WIMP dark matter: concepts and status," *Journal of Physics G: Nuclear and Particle Physics*, vol. 46, no. 10, p. 103 003, Aug. 2019.

[77] S. Ahlen, F. Avignone, R. Brodzinski, A. Drukier, G. Gelmini, and D. Spergel, "Limits on cold dark matter candidates from an ultralow background germanium spectrometer," *Phys. Lett. B*, vol. 195, pp. 603–608, Sep. 1987.

[78] W. H. Dai, L. P. Jia, H. Ma, *et al.*, CDEX Collaboration, "Exotic Dark Matter Search with the CDEX-10 Experiment at China's Jinping Underground Laboratory," *Phys. Rev. Lett.*, vol. 129, no. 22, Nov. 2022.

[79] A. Aguilar-Arevalo, D. Amidei, D. Baxter, *et al.*, DAMIC Collaboration, "Results on Low-Mass Weakly Interacting Massive Particles from an 11 kg d Target Exposure of DAMIC at SNOLAB," *Phys. Rev. Lett.*, vol. 125, no. 24, Dec. 2020.

[80] L. Barak, I. M. Bloch, M. Cababie, *et al.*, SENSEI Collaboration, "SENSEI: Direct-Detection Results on sub-GeV Dark Matter from a New Skipper CCD," *Phys. Rev. Lett.*, vol. 125, no. 17, Oct. 2020.

[81] K. Freese, M. Lisanti, and C. Savage, "Colloquium: Annual modulation of dark matter," *Reviews of Modern Physics*, vol. 85, no. 4, pp. 1561–1581, Nov. 2013.

[82] R. Bernabei, P. Belli, A. Bussolotti, *et al.*, "Further results from DAMA/LIBRA-phase2 and perspectives," *Nuclear Physics and Atomic Energy*, vol. 22, no. 4, pp. 329–342, Dec. 2021.

[83] G. Adhikari, E. B. de Souza, N. Carlin, *et al.*, "Strong constraints from COSINE-100 on the DAMA dark matter results using the same sodium iodide target," *Science Advances*, vol. 7, no. 46, Nov. 2021.

[84] J. Amaré, S. Cebrián, D. Cintas, *et al.*, "Annual modulation results from three-year exposure of ANAIS-112," *Phys. Rev. D*, vol. 103, no. 10, May 2021.

[85] H. S. Lee, H. Bhang, J. H. Choi, *et al.*, KIMS Collaboration, "Search for low-mass dark matter with CsI(Tl) crystal detectors," *Phys. Rev. D*, vol. 90, no. 5, Sep. 2014.

[86] D. W. Amaral, T. Aralis, T. Aramaki, *et al.*, SuperCDMS Collaboration, "Constraints on low-mass, relic dark matter candidates from a surface-operated SuperCDMS single-charge sensitive detector," *Phys. Rev. D*, vol. 102, no. 9, Nov. 2020.

[87] Q. Arnaud, E. Armengaud, C. Augier, *et al.*, EDELWEISS Collaboration, "First Germanium-Based Constraints on Sub-MeV Dark Matter with the EDELWEISS Experiment," *Phys. Rev. Lett.*, vol. 125, no. 14, Oct. 2020.

[88] A. H. Abdelhameed, G. Angloher, P. Bauer, *et al.*, CRESST Collaboration, "First results from the CRESST-III low-mass dark matter program," *Phys. Rev. D*, vol. 100, p. 102 002, 10 Nov. 2019.

[89] R. Agnese, T. Aralis, T. Aramaki, *et al.*, SuperCDMS Collaboration, "Search for low-mass dark matter with CDMSlite using a profile likelihood fit," *Phys. Rev. D*, vol. 99, no. 6, Mar. 2019.

[90] R. Ajaj, P.-A. Amaudruz, G. R. Araujo, *et al.*, DEAP Collaboration, "Search for dark matter with a 231-day exposure of liquid argon using DEAP-3600 at SNOLAB," *Phys. Rev. D*, vol. 100, no. 2, Jul. 2019.

[91] K. Abe, K. Hiraide, N. Kato, *et al.*, XMASS Collaboration, "Direct dark matter searches with the full data set of XMASS-I," *Phys. Rev. D*, vol. 108, no. 8, Oct. 2023.

[92] P. Agnes, I. F. M. Albuquerque, T. Alexander, *et al.*, DarkSide Collaboration, "Low-Mass Dark Matter Search with the DarkSide-50 Experiment," *Phys. Rev. Lett.*, vol. 121, p. 081 307, 8 Aug. 2018.

[93] J. Aalbers, D. S. Akerib, C. W. Akerlof, *et al.*, LZ Collaboration, "First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment," *Phys. Rev. Lett.*, vol. 131, no. 4, Jul. 2023.

[94] E. Aprile, K. Abe, F. Agostini, *et al.*, XENON Collaboration, "First Dark Matter Search with Nuclear Recoils from the XENONnT Experiment," *Phys. Rev. Lett.*, vol. 131, p. 041 003, 4 Jul. 2023.

[95] Y. Meng, Z. Wang, Y. Tao, *et al.*, PandaX-4T Collaboration, "Dark Matter Search Results from the PandaX-4T Commissioning Run," *Phys. Rev. Lett.*, vol. 127, no. 26, Dec. 2021.

[96] C. Amole, M. Ardid, I. J. Arnquist, *et al.*, PICO Collaboration, "Dark matter search results from the complete exposure of the PICO-60 C_3F_8 bubble chamber," *Phys. Rev. D*, vol. 100, p. 022 001, 2 Jul. 2019.

[97] E. Daw, J. Fox, J.-L. Gauvreau, *et al.*, "Spin-dependent limits from the DRIFT-IId directional dark matter detector," *Astroparticle Physics*, vol. 35, no. 7, pp. 397–401, 2012.

[98] M. Schumann, L. Baudis, L. Bütikofer, A. Kish, and M. Selvi, "Dark matter sensitivity of multi-ton liquid xenon detectors," *JCAP*, vol. 2015, no. 10, pp. 016–016, Oct. 2015.

[99] S. Kang, S. Scopel, G. Tomar, and J.-H. Yoon, "Present and projected sensitivities of dark matter direct detection experiments to effective WIMP-nucleus couplings," *Astroparticle Physics*, vol. 109, pp. 50–68, May 2019.

[100] J. Billard, M. Boulay, S. Cebrián, *et al.*, "Direct detection of dark matter - APPEC committee report," *Reports on Progress in Physics*, vol. 85, no. 5, p. 056 201, Apr. 2022.

[101] V. Chepel and H. Araújo, "Liquid noble gas detectors for low energy particle physics," *Journal of Instrumentation*, vol. 8, Jul. 2012.

[102] T. Doke, A. Hitachi, J. Kikuchi, K. Masuda, H. Okada, and E. Shibamura, "Absolute scintillation yields in liquid argon and xenon for various particles," *Japanese Journal of Applied Physics*, vol. 41, no. 3R, p. 1538, Mar. 2002.

[103] P. Sorensen and C. E. Dahl, "Nuclear recoil energy scale in liquid xenon with application to the direct detection of dark matter," *Phys. Rev. D*, vol. 83, p. 063 501, 6 Mar. 2011.

[104] D. S. Akerib, S. Alsum, H. M. Araújo, *et al.*, LUX Collaboration, "Liquid xenon scintillation measurements and pulse shape discrimination in the LUX dark matter detector," *Phys. Rev. D*, vol. 97, no. 11, Jun. 2018.

[105] E. Aprile and L. Baudis, "Liquid noble gases," in *Particle Dark Matter: Observations, Models and Searches*, G. Bertone, Ed. Cambridge: Cambridge Univ. Press, 2010, pp. 413–436.

[106] B. J. Mount, S. Hans, R. Rosero, *et al.*, LZ Collaboration, "LUX-ZEPLIN (LZ) Technical Design Report," *arXiv 1703.09144*, 2017.

[107] M. Szydagis, J. Balajthy, G. A. Block, *et al.*, "A Review of NEST Models, and Their Application to Improvement of Particle Identification in Liquid Xenon Experiments," *arXiv 2211.10726*, 2023.

[108] 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.

[109] J. Lindhard, V. Nielsen, M. Scharff, and P. V. Thomsen, "Integral equations governing radiation effects. (notes on atomic collisions, iii)," *Kgl. Danske Videnskab., Selskab. Mat. Fys. Medd.*, vol. 33, Jan. 1963.

[110] C. Faham, V. Gehman, A. Currie, A. Dobi, P. Sorensen, and R. Gaitskell, "Measurements of wavelength-dependent double photoelectron emission from single photons in VUV-sensitive photomultiplier tubes," *Journal of Instrumentation*, vol. 10, no. 09, P09010, Sep. 2015.

[111] 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.

[112] G. Adhikari, N. Carlin, J. Choi, *et al.*, "An induced annual modulation signature in COSINE-100 data by DAMA/LIBRA's analysis method," *Scientific Reports*, vol. 13, no. 1, p. 4676, 2023.

[113] G. D'Imperio, SABRE collaboration, "Dark matter search with the SABRE experiment," *Journal of Physics: Conference Series*, vol. 1342, no. 1, p. 012 060, Jan. 2020.

[114] J. Xu, D. Adams, B. G. Lenardo, *et al.*, "Search for the Migdal effect in liquid xenon with keV-level nuclear recoils," *Phys. Rev. D*, vol. 109, p. L051101, 5 Mar. 2024.

[115] P. Agnes, I. F. M. Albuquerque, T. Alexander, *et al.*, DarkSide Collaboration, "Search for Dark-Matter–Nucleon Interactions via Migdal Effect with DarkSide-50," *Phys. Rev. Lett.*, vol. 130, p. 101 001, 10 Mar. 2023.

[116] A. M. Suliga, "Diffuse Supernova Neutrino Background," in *Handbook of Nuclear Physics*. Springer Nature Singapore, 2023, pp. 3789–3806.

[117] J. Aalbers, S. S. AbdusSalam, K. Abe, *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.

[118] B. López Paredes, H. Araújo, F. Froborg, *et al.*, "Response of photomultiplier tubes to xenon scintillation light," *Astroparticle Physics*, vol. 102, pp. 56–66, 2018.

[119] R. Linehan, R. Mannino, A. Fan, *et al.*, "Design and production of the high voltage electrode grids and electron extraction region for the LZ dual-phase xenon time projection chamber," *NIM-A*, vol. 1031, p. 165 955, 2022.

[120] D. Akerib, C. Akerlof, D. Akimov, *et al.*, LZ Collaboration, "The LUX-ZEPLIN (LZ) experiment," *NIM-A*, vol. 953, p. 163 047, 2020.

[121] S. Haselschwardt, S. Shaw, H. Nelson, *et al.*, "A liquid scintillation detector for radioassay of gadolinium-loaded liquid scintillator for the LZ Outer Detector," *NIM-A*, vol. 937, pp. 148–163, 2019.

[122] V. Solovov, V. Belov, D. Akimov, *et al.*, "Position reconstruction in a dual phase xenon scintillation detector," in *2011 IEEE Nuclear Science Symposium Conference Record*, 2011, pp. 1226–1233.

[123] D. S. Akerib, C. W. Akerlof, D. Y. Akimov, *et al.*, LZ Collaboration, "The LUX-ZEPLIN (LZ) radioactivity and cleanliness control programs," *Eur. Phys. J. C*, vol. 80, no. 11, p. 1044, 2020.

[124] D. S. Akerib, C. W. Akerlof, S. K. Alsum, *et al.*, LZ Collaboration, "Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment," *Phys. Rev. D*, vol. 101, no. 5, Mar. 2020.

[125] J. Aalbers, D. S. Akerib, A. K. A. Musalhi, *et al.*, LZ Collaboration, "Background determination for the LUX-ZEPLIN dark matter experiment," *Phys. Rev. D*, vol. 108, no. 1, Jul. 2023.

[126] J. Aalbers, D. S. Akerib, A. K. Al Musalhi, *et al.*, LZ Collaboration, "Cosmogenic production of ^{37}Ar in the context of the LUX-ZEPLIN experiment," *Phys. Rev. D*, vol. 105, p. 082 004, 8 Apr. 2022.

[127] F. Gray, C. Ruybal, J. Totushek, D.-M. Mei, K. Thomas, and C. Zhang, "Cosmic ray muon flux at the Sanford Underground Laboratory at Homestake," *NIM-A*, vol. 638, no. 1, pp. 63–66, 2011.

[128] D. Akerib, C. Akerlof, S. Alsum, *et al.*, LZ Collaboration, "Measurement of the gamma ray background in the Davis cavern at the Sanford Underground Research Facility," *Astroparticle Physics*, vol. 116, p. 102 391, 2020.

[129] D. Akerib, C. Akerlof, A. Alqahtani, *et al.*, LZ Collaboration, "Simulations of events for the LUX-ZEPLIN (LZ) dark matter experiment," *Astroparticle Physics*, vol. 125, p. 102 480, 2021.

[130] E. Aprile, J. Aalbers, F. Agostini, *et al.*, XENON Collaboration, "Observation of two-neutrino double electron capture in ^{124}Xe with XENON1T," *Nature*, vol. 568, no. 7753, pp. 532–535, Apr. 2019.

[131] E. Aprile, K. Abe, F. Agostini, *et al.*, XENON Collaboration, "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.

[132] J. B. Albert, M. Auger, D. J. Auty, *et al.*, EXO Collaboration, "Improved measurement of the $2\nu\beta\beta$ half-life of ^{136}Xe with the EXO-200 detector," *Phys. Rev. C*, vol. 89, p. 015 502, 1 Jan. 2014.

[133] A. Singh, E. Bernard, A. Biekert, *et al.*, "Analysis of ^{83m}Kr prompt scintillation signals in the PIXeY detector," *Journal of Instrumentation*, vol. 15, no. 01, P01023–P01023, Jan. 2020.

[134] D. S. Akerib, H. M. Araújo, X. Bai, *et al.*, LUX Collaboration, "Tritium calibration of the LUX dark matter experiment," *Phys. Rev. D*, vol. 93, no. 7, Apr. 2016.

[135] D. S. Akerib, S. Alsum, H. M. Araújo, *et al.*, LUX Collaboration, "Low-energy (0.7-74 keV) nuclear recoil calibration of the LUX dark matter experiment using D-D neutron scattering kinematics," *arXiv 1608.05381*, 2016.

[136] A. Sazzad, J. Busenitz, A. Piepke, S. Poudel, H. Trewin, and A. LeViness, "Design and characterization of AmLi neutron sources for the LZ experiment," *Journal of Instrumentation*, vol. 18, no. 05, P05006, May 2023.

[137] Hamamatsu Photonics K. K., *Photomultiplier Tubes: Basics and Applications.* 4th ed. 2017, p. 180.

[138] D. S. Akerib, S. Alsum, H. M. Araújo, *et al.*, LUX Collaboration, "Investigation of background electron emission in the LUX detector," *Phys. Rev. D*, vol. 102, p. 092 004, 9 Nov. 2020.

[139] T. Takahashi, J.-z. Ruan(Gen), S. Kubota, and F. Shiraishi, "Time delay of recombination luminescence: Electron thermalization in xenon gas and xenon-nitrogen gas mixtures excited by ^{252}Cf fission fragments," *Phys. Rev. A*, vol. 25, pp. 600–603, 1 Jan. 1982.

[140] F. James, *Statistical Methods in Experimental Physics*, 2nd ed. Singapore: World Scientific Publishing Co. Pte. Ltd., 2006, ch. 11.

[141] A. Best, "Measurement of alpha capture reactions on O-17 and O-18 for the s process," Ph.D. dissertation, University of Notre Dame, Dec. 2011.

[142] L. Orawo, "Confidence intervals for the binomial proportion: A comparison of four methods," *Open Journal of Statistics*, no. 11, pp. 806–816, 2021.

[143] M. Agostini, K. Altenmüller, S. Appel, *et al.*, Borexino Collaboration, "Simultaneous precision spectroscopy of *pp*, ^{7}Be, and *pep* solar neutrinos with Borexino Phase-II," *Phys. Rev. D*, vol. 100, p. 082 004, 8 Oct. 2019.

[144] N. Vinyoles, A. M. Serenelli, F. L. Villante, *et al.*, "A New Generation of Standard Solar Models," *The Astrophysical Journal*, vol. 835, no. 2, p. 202, Jan. 2017.

[145] B. Aharmim, S. N. Ahmed, A. E. Anthony, *et al.*, SNO Collaboration, "Combined analysis of all three phases of solar neutrino data from the Sudbury Neutrino Observatory," *Phys. Rev. C*, vol. 88, p. 025 501, 2 Aug. 2013.

[146] M. Berglund and M. E. Wieser, "Isotopic compositions of the elements 2009 (IUPAC Technical Report)," *Pure and Applied Chemistry*, vol. 83, no. 2, pp. 397–410, 2011.

[147] D. Baxter, I. M. Bloch, E. Bodnia, *et al.*, "Recommended conventions for reporting results from direct dark matter searches," *Eur. Phys. J. C*, vol. 81, no. 10, Oct. 2021.

[148] J. Allison, K. Amako, J. Apostolakis, *et al.*, "Recent developments in Geant4," *NIM-A*, vol. 835, pp. 186–225, 2016.

[149] L. Althueser, B. Antunović, E. Aprile, *et al.*, DARWIN Collaboration, "GPU-based optical simulation of the DARWIN detector," *Journal of Instrumentation*, vol. 17, no. 07, P07018, Jul. 2022.

[150] "GATE Documentation: Generating and tracking optical photons." (2024), [Online]. Available: `https://opengate.readthedocs.io/en/latest/generating_and_tracking_optical_photons.html#unified-model` (visited on 07/08/2024).

[151] S. Kravitz, R. J. Smith, L. Hagaman, *et al.*, "Measurements of angle-resolved reflectivity of PTFE in liquid xenon with IBEX," *Eur. Phys. J. C*, vol. 80, no. 3, p. 262, 2020.

[152] A. Baldini, C. Bemporad, F. Cei, *et al.*, "Absorption of scintillation light in a 100l liquid xenon gamma-ray detector and expected detector performance," *NIM-A*, vol. 545, no. 3, pp. 753–764, 2005.

[153] D. S. Akerib, C. W. Akerlof, D. Y. Akimov, *et al.*, LZ Collaboration, "LUX-ZEPLIN (LZ) Conceptual Design Report," *arXiv 1509.02910*, 2015.

[154] F. Neves, A. Lindote, A. Morozov, *et al.*, "Measurement of the absolute reflectance of polytetrafluoroethylene (PTFE) immersed in liquid xenon," *Journal of Instrumentation*, vol. 12, no. 01, P01017–P01017, Jan. 2017.

[155] E. Aprile, R. Budnik, B. Choi, *et al.*, "Measurement of the scintillation yield of low-energy electrons in liquid xenon," *Phys. Rev. D*, vol. 86, no. 11, Dec. 2012.

[156] A. Azzalini, *The Skew-Normal and Related Families*. Cambridge: Cambridge Univ. Press, 2013, ch. 2.

[157] T. Fruth, "PMT Studies and Loop Antenna Development for the LZ Dark Matter Search," Ph.D. dissertation, University of Oxford, 2019.

[158] A. G. Wright, *The Photomultiplier Handbook*. Oxford: Oxford Univ. Press, 2017, ch. 5.

[159] D. S. Akerib, S. Alsum, H. M. Araújo, *et al.*, LUX Collaboration, "Ultralow energy calibration of LUX detector using ^{127}Xe electron capture," *Phys. Rev. D*, vol. 96, no. 11, Dec. 2017.

[160] S. Baker and R. D. Cousins, "Clarification of the use of CHI-square and likelihood functions in fits to histograms," *NIM*, vol. 221, no. 2, pp. 437–442, 1984.

[161] B. Denby, "Neural networks and cellular automata in experimental high energy physics," *Computer Physics Communications*, vol. 49, no. 3, pp. 429–448, 1988.

[162] L. Lönnblad, C. Peterson, and T. Rögnvaldsson, "Finding gluon jets with a neural trigger," *Phys. Rev. Lett.*, vol. 65, pp. 1321–1324, 11 Sep. 1990.

[163] A. S. Cornell, W. Doorsamy, B. Fuks, G. Harmsen, and L. Mason, "Boosted decision trees in the era of new physics: A smuon analysis case study," *JHEP*, vol. 2022, no. 4, Apr. 2022.

[164] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, ACM, Aug. 2016.

[165] J. Mushava and M. Murray, "Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring," *Expert Systems with Applications*, vol. 238, p. 121 876, 2024.

[166] G. Aad, B. Abbott, D. Abbott, *et al.*, ATLAS Collaboration, "Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 80, p. 1085, Nov. 2020.

[167] J. Aalbers, D. S. Akerib, A. K. Al Musalhi, *et al.*, LZ Collaboration, "First constraints on WIMP-nucleon effective field theory couplings in an extended energy region from LUX-ZEPLIN," *Phys. Rev. D*, vol. 109, no. 9, May 2024.

[168] T. Chen and C. Guestrin. "XGBoost Documentation." (2022), [Online]. Available: `https://xgboost.readthedocs.io/en/stable/` (visited on 03/27/2024).

[169] A. Ng, "Feature selection, L 1 vs. L 2 regularization, and rotational invariance," in *Proceedings of the Twenty-First International Conference on Machine Learning*, Sep. 2004.

[170] I. Khurana, "Mitigating Accidental Coincidence Backgrounds in the LZ experiment," Ph.D. dissertation, University College London, 2023.

[171] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction To Statistical Learning: With Applications In R*. New York: Springer Science + Business Media, 2013, ch. 5.

[172] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals," *Phys. Rev. D*, vol. 57, no. 7, pp. 3873–3889, Apr. 1998.

[173] 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, no. 7, Oct. 2020.

[174] 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.