
Machine Learning based Signal Enhancement and the Search for Bosonic Dark Matter with the XENONnT Experiment

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik
des Karlsruher Instituts für Technologie (KIT)

angenommene

Dissertation

von

Sebastian Helmut Vetter

geboren am 01. Dezember 1996 in Pforzheim

Referentin: Prof. Dr. Kathrin Valerius
Institut für Astroteilchenphysik
Karlsruher Institut für Technologie

Korreferent: Prof. Torben Ferber
Institut für Experimentelle Teilchenphysik
Karlsruher Institut für Technologie

Tag der mündlichen Prüfung: 24.07.2026

Erklärung der selbstständigen Anfertigung der Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertation mit dem Titel

Machine Learning based Signal Enhancement and the Search for Bosonic Dark Matter with the XENONnT Experiment

selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht habe, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde. Ich versichere außerdem, dass ich die Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Karlsruhe, den 25. August 2026

(Sebastian Vetter)

”We step out of our Solar System into the Universe, seeking only peace and friendship, to teach if we are called upon, to be taught if we are fortunate.” - Kurt Waldheim, Golden Record of the Voyager Spacecraft

Abstract

Various astrophysical and cosmological observations point to the existence of some additional mass in the universe that is approximately five times as abundant as baryonic matter. This additional matter is commonly referred to as *Dark Matter* (DM). Despite numerous efforts at its detection, the particle-nature of DM persists to be unknown. One of the world leading experiments in the search for DM via direct detection is the XENONnT experiment, a liquid-xenon dual-phase time projection chamber (TPC) operating at the Laboratori Nazionali del Gran Sasso (LNGS). This thesis presents three projects completed in the context of XENONnT.

The first project is the development and implementation of a Conditional Normalizing Flow (CNF) position reconstruction model. In the first two science runs (SRs), SR0 and SR1, XENONnT relied on classical feed-forward neural networks for the position reconstruction, which provide a point estimate in the horizontal (x, y) plane. The CNF model developed here is not only more accurate, in particular near the walls of the experiment, but in addition predicts accurate position probabilities, that span the complete detector plane. As a consequence, the additional probability information has already been implemented by other XENONnT scientists into their analyses.

The second project is the development of a Variational Autoencoder (VAE) for the optical identification of potential defects on the newly installed hexagonal mesh cathode electrode. For this purpose, several image processing algorithms from the fields of computer vision (CV) and machine learning (ML) have been implemented. Combinations of different CV and ML methods were compared among each other to achieve the best efficiency in detecting anomalies in images of the cathode. The implemented methods can also be used for the development of large-scale electrodes for future experiments, such as the planned XLZD observatory.

The final project included in this work is the derivation of the sensitivity of XENONnT for absorption signals of bosonic DM, such as axion-like particles (ALPs) and dark photons (DPs), in an extended energy region of interest (ROI) up to 1000 keV, using data from the first two science runs SR0 and SR1. Towards this goal, a detailed signal efficiency model was developed and validated in the relevant ROI. Furthermore, in-depth studies of all present backgrounds were performed and their spectra implemented in a statistical model that uses a binned profile likelihood ratio (PLR) test. The sensitivity projection derived in this work predicts world leading exclusion limits for the axio-electric coupling strength g_{ae} and the kinetic mixing parameter κ .

Zusammenfassung

Eine Großzahl astrophysikalischer und kosmologischer Beobachtungen deuten auf die Existenz von zusätzlicher Masse im Universum hin, die in etwa fünfmal so massereich ist wie baryonische Materie. Diese zusätzliche Materie wird gemeinhin als *Dunkle Materie* (DM) bezeichnet. Trotz zahlreicher Versuche sie nachzuweisen, ist die Teilchennatur dieser DM nach wie vor unbekannt. Eines der weltweit führenden Experimente zur direkten Detektion von DM ist das XENONnT Experiment, das als flüssig-Xenon zwei Phasen Spurendriftkammer (engl. TPC) im Laboratori Nazionali del Gran Sasso (LNGS) aufgebaut ist. Diese Arbeit ist unterteilt in drei Schwerpunkte, die im Rahmen von XENONnT bearbeitet wurden.

Der erste Schwerpunkt ist die Entwicklung und Implementierung eines Conditional Normalizing Flow (CNF) Modells zur Positionsrekonstruktion. In den ersten zwei Messphasen, SR0 und SR1, stütze sich XENONnT bei der Positionsrekonstruktion auf klassische Feed-Forward neuronale Netze, deren Vorhersage einem einzelnen Punkt in der horizontalen (x, y) -Ebene entspricht. Das hier vorgetestete CNF-Modell ist nicht nur genauer, insbesondere nahe der Wände des Experiments, sondern liefert darüber hinaus eine akkurate Wahrscheinlichkeitsverteilung der Position über die gesamte Detektor-Ebene. Diese zusätzliche Wahrscheinlichkeitsinformation ist bereits fester Bestandteil in verschiedenen Analysen von XENONnT-Daten.

Der zweite Schwerpunkt befasst sich mit der Entwicklung eines Variational Autoencoder (VAE) Modells, zur optischen Identifizierung potenzieller Defekte auf dem hexagonalen Gitter der Kathodenelektrode. Zu diesem Zweck wurden mehrere Bildverarbeitungsalgorithmen aus den Gebieten der Computer Vision (CV) und des Machine Learnings (ML) implementiert. Kombinationen verschiedener CV- und ML-Methoden wurden verglichen um die größte Effizienz zur Anomalieentdeckung zu erreichen. Die implementierten Methoden können auch für die Entwicklung großflächiger Elektroden für zukünftige Experimente wie beispielsweise das geplante XLZD-Observatorium genutzt werden.

Das letzte in dieser Arbeit behandelte Projekt befasst sich mit der Sensitivität von XENONnT gegenüber der Absorption von bosonischer DM, wie beispielsweise Axion-ähnlicher Teilchen (engl. ALP) oder dunkler Photonen (DP). Die Sensitivität wurde in einem erweiterten Energie-Interessenbereich (engl. ROI) von bis zu 1000 keV ermittelt, unter Verwendung von Daten aus den ersten beiden Messphasen von XENONnT, SR0 und SR1. Zu diesem Zweck wurde ein detailliertes Signaleffizienzmodell entwickelt und in der relevanten ROI validiert. Darüber hinaus wurden ausführliche Untersuchungen von allen vorhandenen Untergrundquellen durchgeführt, und deren Spektren wurden in ein statistisches Modell, auf Basis eines binned Profile Likelihood Ratio (PLR) Tests, integriert. Die in dieser Arbeit vorgestellte Sensitivitätsprojektion sagt weltweit führende Ausschlussgrenzen für die axioelektrische Kopplungskonstante g_{ae} und den kinetischen Mischungsparameter κ vorher.

Author's Contribution

This thesis presents research conducted in the context of the XENONnT experiment, which is a collaborative effort of more than 200 scientists. As such, no scientific result of XENONnT is the achievement of a single person, but rather accomplished as a team effort via collaboration. Below, I summarize my contributions to the work presented in this thesis.

- As member of the XENON collaboration, I spent multiple months at the Laboratori Nazionali del Gran Sasso (LNGS), working as a shifter. The tasks of a shifter include the maintenance and testing of the hardware and emergency response in case of an anomalous measurement by any of the many sensors monitoring the XENONnT experiment and auxiliary systems. Further, in early 2024, I wired the replacement anode together with Vera Hiu-Sze Wu, Julia Müller and Tiffany Luce, that has been installed in the detector in summer 2025.
- In [subsection 2.3.4](#), the signal corrections process of the XENONnT analysis pipeline is described. During my PhD project, I was leading the working group "Signals and Detector" for one and a half years. During that time, my duties included coordination, oversight and validation of the signal correction procedure for science run 2 (SR2) of XENONnT, in cooperation with my co-leaders Juehang Qin and Kexin Liu. As part of that work, I contributed to the data analysis software packages STRAXEN and CUTAX, and the simulation framework *fuse*, all of which are described in [subsection 2.3.6](#).
- Chapter 3 describes the Machine Learning (ML) based position reconstruction algorithms utilized in XENONnT. The models used in science run 0 (SR0) and science run 1 (SR1), of which I inherited the maintenance, were developed by Shixiao Liang and Andrii Terliuk. For SR2, I developed two ways of extending classical Neural Networks (NNs) to produce probabilistic predictions, as opposed to point estimates, and implemented a position reconstruction model that makes use of both of them in combination ([section 3.3](#)). Further, I worked in close cooperation with Juehang Qin to develop a Conditional Normalizing Flow (CNF) based position reconstruction model, which outperforms previous models and is since then the default used in XENONnT analysis ([section 3.4](#)). The development and validation of the CNF model is currently being prepared for submission as [1]. To validate the position reconstruction models, Shixiao Liang and Kexin Liu developed a data-driven approach that takes advantage of the electric field caused by the XENONnT perpendicular wires. I extended the data-driven validation to low energies, by developing a study of the position resolution using so-called hotspots in the detector ([section 3.5](#)). Finally, [section 3.6](#) lists two ways in which the additional information provided by the CNF is already used by other scientists of the XENON collaboration, and additionally, potential future use cases demonstrated by me.

-
- Chapter 4 is about the quality assurance of the replacement cathode of the XENONnT detector, with results published in [2]. My main contribution to this was the development of a Variational Autoencoder (VAE) ML model that is capable of identifying potential defects in images of the cathode, an idea initially tested for the predecessor experiment XENON1T by Jan Lommler and Daniel Wenz (section 4.3). In the course of this project, I tested and implemented computer vision (CV) and ML methods for image processing and segmentation for the optical inspection of the cathode. Alexey Elykov and Vera Hiu-Sze Wu, together with technicians and engineers from KIT, constructed a high voltage (HV) test setup to investigate the HV performance of the cathode, towards which I contributed data analysis and image processing software. The methods I developed and present in chapter 4 will be used in future KIT experiments aimed towards the development of electrodes for the XENON-LUX-ZEPLIN-DARWIN (XLZD) detector.
 - Chapter 5 examines the sensitivity of the XENONnT experiment to search for signals induced by the absorption of bosonic Dark Matter (DM), using SR0 and SR1 data. The preparation of the bosonic DM search was performed in parallel to the measurement of the spectrum of solar neutrinos from proton-proton fusion. For the neutrino search, which is currently being prepared for submission as [3], I contributed the optimization of the fiducial volume and the constraint of the neutron activated isotope background. Many intermediate results I could adapt either from [3], or from other XENONnT studies, cited at the appropriate place. In subsection 5.1.3, I present an estimate of the event building efficiency (EBE), that I developed in particular for the here presented study. To the cuts and cut acceptance described in section 5.2, my contribution was the extension of the validation of the cuts towards higher energies. I performed the computation and smoothing of the resulting acceptance, using methods developed by Yue Ma and Ananthakrishnan Ravindran. The material originating background model, explained in section 5.3, was developed by me, under guidance of, and using the simulation framework developed by, Maxime Pierre. The optimization of the fiducial volume was performed by me, using the adaptive kernel density estimation (AKDE) method first tested for this purpose by Alexey Elykov. The simulation of the lead backgrounds, detailed in section 5.4, was performed and analyzed by me, under the guidance and cooperation of Giovanni Volta. For the constraints of the neutron activated isotopes in section 5.5, I used methodology developed by Yue Ma. The details listed in section 5.6 were taken from the solar neutrino study, except for the ^{214}Bi skin background, which I constrained with the help of Maxime Pierre. Finally, the derivation of the sensitivity that is elaborated in section 5.7 was developed and validated by me, using the *alea* software, with the help of Giovanni Volta and Robert Hammann. The implementations made by me can be used to perform the fit of the statistical model to the recorded data, and compute the 90 % confidence level (CL) exclusion limits of the axioelectric coupling strength g_{ae} and the dark photon kinetic mixing κ .

Acknowledgements

In my journey as a PhD student I was accompanied by many great people, who helped me learn and discover all the fantastic things this universe has to offer. It would be impossible to list every single person who made an impact on me, but I wish to mention at least a few.

To begin with, I want to thank Prof. Dr. Kathrin Valerius for the supervision of my thesis, for the guidance and trust she offered me, and for always supporting my work. Further, I want to express my gratitude to Dr. Klaus Eitel, for directing and navigating me on all of my scientific and professional endeavors. Without the leadership of Kathrin and Klaus, this thesis would not have been possible.

I would also like to thank everyone at the Institute for Astroparticle Physics at KIT, former and current, for the fruitful years of discussions, insights and curiosities we shared: Dr. Alexey Elykov, Dr. Yanina Biondi, Dr. Francesco Toschi and Dr. Melih Solmaz. A special thanks I want to dedicate to my fellow PhD students, who endured the hardships, and shared the bright moments together with me and made my whole journey much more special: Dr. Melih Kara, Dr. Sukeerthi Dharani, Dr. Vera Hiu-Sze Wu, Alexis Michel and Tom Sonius. Thank you all. On the administrative side, I want to thank Anna Friedrich, Peichen Chen and Sabine Bucher, for always being available and bringing order to the chaos of bureaucracy.

Further, I want to extend my appreciation to all of the many people of the XENON collaboration. Working with so many different people from all over the world was an outstanding experience, and I can not count all of the deeply interesting things that I learned. In particular, I want to thank Dr. Maxime Pierre and Dr. Giovanni Volta, for their unwavering support, their ever-present guidance, and their thorough mentorship. In addition, I wish to thank Dr. Juehang Qin, for teaching me from across 7 time zones.

And finally, I want to thank my friends and family, who always helped me back up, whenever I fell down.

Contents

Abstract	ii
Author’s Contribution	iv
Acknowledgements	vi
Table of Contents	x
List of Tables	xi
List of Figures	xvi
Abbreviations	xvii
1 Introduction	1
2 The Search for Dark Matter with the XENONnT Experiment	3
2.1 Dark Matter	3
2.1.1 Evidences for Dark Matter	3
2.1.2 Particle Dark Matter	6
2.2 Direct Detection of Dark Matter	11
2.3 Dual Phase Liquid Xenon Time Projection Chambers and the XENONnT Experiment	14
2.3.1 Liquid Xenon as Detector Medium	15
2.3.2 The XENONnT Experiment	17
2.3.3 The Signal Formation Process in liquid Xenon	19
2.3.4 The Signal Correction Process of XENONnT	22
2.3.5 Data Organization and Jargon in XENONnT	25
2.3.6 The Software Landscape of XENONnT	27
3 Probabilistic Position Reconstruction using Conditional Normalizing Flows	31
3.1 Position Reconstruction in XENONnT	31
3.2 A short Introduction to Feed-Forward Neural Networks	34
3.3 Probabilistic Extension of Neural Networks	36

3.3.1	Obtaining Position Probability Distributions via One Hot encoded Classification	37
3.3.2	Obtaining Position Probability Distributions from Parameterized Likelihood Functions	40
3.3.3	Comparison and Combination of Probabilistic Extensions	41
3.3.4	The Skew Gauss Position Reconstruction Model	44
3.4	Position Reconstruction using Conditional Normalizing Flows	49
3.4.1	A short Introduction to Normalizing Flows	49
3.4.2	Model Definition and Training	50
3.4.3	Model Validation and Performance	51
3.5	Data driven Validation of Position Reconstruction	57
3.5.1	Single Electron Signals from Hotspots	57
3.5.2	S2 Signals from beneath the Perpendicular Wires	59
3.5.3	Result and Discussion	60
3.6	Applications of Probabilistic Position Reconstruction	62
3.6.1	Iterative Bayesian S2 Merging	62
3.6.2	Fuzzy Selection of Perpendicular Wire Events	64
3.6.3	Position Uncertainty Cut	65
3.6.4	Multi-Scatter Detection	67
3.6.5	Transversal Electron Diffusion	69
3.7	Conclusion and Outlook	71
4	Quality Assurance of the XENONnT Upgrade Cathode	73
4.1	The Upgrade Cathode of XENONnT	73
4.2	Leg Segmentation	77
4.2.1	A short Introduction to Computer Vision	77
4.2.2	Machine Learning based Pixel Segmentation	82
4.3	Machine Learning based Detection of Defects for the new XENONnT Cathode	89
4.3.1	A short Introduction to (Variational) Autoencoders	90
4.3.2	Performance of the VAE Defect Detection	92
4.4	High Voltage Test of the new XENONnT Cathode	98
4.4.1	Experimental Setup	98
4.4.2	High voltage Test Results	100
4.5	Conclusion and Outlook	102
5	The Search for Bosonic Dark Matter with XENONnT	105
5.1	Detector Response Model	106
5.1.1	Energy Reconstruction	106
5.1.2	Energy Threshold	107
5.1.3	Event Building Efficiency	108
5.2	Signal Acceptance	113
5.2.1	ER Band Selection	113
5.2.2	Single Scatter Selection	114
5.2.3	Derivation of the Cut Acceptance	115
5.2.4	S1 Width Cut	118
5.2.5	S1 Area Fraction Top Cut	119
5.2.6	S1 Max PMT Cut	119

5.2.7	Corrected S2 Area Fraction Top Cut	119
5.2.8	S2 Pattern Cut	120
5.2.9	S2 Width Cut	120
5.2.10	Cumulative Cut Acceptance	122
5.3	Material Radioactivity and Fiducialization	124
5.3.1	Material Originating Background	124
5.3.2	Validation of the Material Background Simulation using Multi- Scatter Events	126
5.3.3	Optimization of the Fiducial Volume	127
5.3.4	Uncertainty of the Fiducial Volume	130
5.3.5	Remaining Material Background after Fiducialization	133
5.4	The Decays of ^{214}Pb and ^{212}Pb	133
5.4.1	Lead Backgrounds: Origin and Mitigation	133
5.4.2	Total ^{214}Pb and ^{212}Pb Activities	134
5.4.3	The Challenge in Constraining Pb β - γ Events	135
5.4.4	Comparison of ^{212}Pb Full Chain Simulation and ^{220}Rn Data	136
5.4.5	Resulting ^{212}Pb and ^{214}Pb Spectra	138
5.5	Neutron Activated Isotopes	138
5.5.1	The Rate Constraint of Neutron-activated Xenon Isotopes on the Example of ^{131m}Xe	141
5.5.2	The Rate Constraint of ^{125}I Decays	142
5.5.3	The Resulting Background Rates from Neutron Activation	144
5.6	Other Backgrounds	146
5.6.1	^{85}Kr β Decay	146
5.6.2	^{124}Xe Double Electron Capture	147
5.6.3	^{136}Xe double β Decay	147
5.6.4	^{83m}Kr γ Emission	147
5.6.5	^3H β Decay	148
5.6.6	Accidental Coincidences	148
5.6.7	Solar Neutrinos	149
5.6.8	^{214}Bi Skin Background	149
5.7	The Sensitivity Projection for Bosonic Dark Matter in XENONnT	151
5.7.1	Summary and Formulation of the Signal and Background Mod- els	152
5.7.2	Statistical Framework	153
5.7.3	Sensitivity Projection	155
5.8	Conclusion and Outlook	157
6	Summary	161
6.1	Probabilistic Position Reconstruction using Conditional Normalizing Flows	161
6.2	Quality Assurance of the XENONnT Upgrade Cathode	163
6.3	The Search for Bosonic Dark Matter with XENONnT	164
A	S2 Hitpattern Simulation using Fuse	167
B	Regular Hexagonal Mesh Algorithm	171
C	Additional Figures on the Data Selection Cuts	175

C.1	S2 Single Scatter	175
C.2	S1 Width	177
C.3	S1 Area Fraction Top	178
C.4	S1 Max PMT	179
C.5	S2 corrected Area Fraction Top	180
C.6	S2 Pattern	181
C.7	S2 Width	182
D	Parameter Comparison between ^{212}Pb Full Chain Simulation and ^{220}Rn Calibration Data	185
D.1	S1 and S2 areas	185
D.2	S1 Parameters	187
D.3	S2 Parameters	188
D.4	Alternative S2s and SS Selection	189
E	Fits of the Rates of Neutron Activated Isotopes	191
F	Derivation of the Critical Region	197

List of Tables

2.1	Abundance of Xe isotopes	16
2.2	Parameters used in XENONnT analysis	27
2.3	Explanation of XENONnT cuts	28
3.1	Model benchmarks for OH models	39
3.2	Comparison between One Hot (OH) and parameterized models	41
3.3	Best fit parameters for the position resolution	62
4.1	Selection of different defect detection VAE models	94
5.1	Data used for energy calibration	107
5.2	Energy calibration parameters	107
5.3	Datasets used to compute the cut acceptance	117
5.4	single-scatter (SS) fraction of ^{214}Pb and ^{212}Pb simulations	138
5.5	Activities of ^{214}Pb and ^{212}Pb backgrounds	138
5.6	Decay energies and half lives of neutron-activated isotopes	140
5.7	Best fit activity and half-life of neutron-activated Xe isotopes	143
5.8	Expected rate of background events from neutron-activated isotopes	145
5.9	Expected activity per background component	152

List of Figures

2.1	The rotation curve of galaxy NGC5055	4
2.2	The CMB anisotropy map	5
2.3	The CMB multipole spectrum	6
2.4	The WIMP number density	9
2.5	Diagram of dark photon kinetic mixing	11
2.6	Current exclusion limits of the SI WIMP DM cross section	13
2.7	Current exclusion limits of g_{ae} and κ	14
2.8	Expected WIMP recoil spectrum for different target nuclei	15
2.9	The mean free path of radiation in liquid xenon	17
2.10	Rendering of XENONnT veto systems	18
2.11	Schematic of the XENONnT cryostat	19
2.12	Sketch of a TPC	20
2.13	Sketch of the quanta generation process	21
2.14	Quantum yields for NR and ER interactions	22
2.15	S1 and S2 distribution of ER and NR events	23
2.16	Comparison of uncorrected and corrected signal strength	24
2.17	Anticorrelation of light and charge yield	26
3.1	An example hitpattern and lightcurve	33
3.2	Sketch of the working principle of an MLP	35
3.3	Position reconstruction via OH models	38
3.4	Distance of uniform points to the center of a square	38
3.5	Resolutions of demonstration OH models	39
3.6	Prediction example of the Guassian loss model	41
3.7	Resolution of the Gaussian loss model	42
3.8	Prediction example of the high resolution OH model	42
3.9	Best matching analytical PDF for high resolution OH model	43
3.10	The training and validation loss of the SG model	45
3.11	Prediction examples of the SG model	46
3.12	Resolution of the SG model	47
3.13	Coverage of the SG model	47
3.14	Predicted parameter distributions of the SG model	48
3.15	Schematic of the working principle of a NF	50
3.16	CNF prediction example	51

3.17	Distribution of predicted positions of MLP and CNF models	52
3.18	Distribution of predicted radius close to the wall	53
3.19	Radial bias of the CNF and MLP position reconstruction models	54
3.20	Resolution of the CNF model	55
3.21	Example 1- σ contours of the CNF	55
3.22	Coverage of the CNF model	56
3.23	Spatial distribution of hotspot peaks	57
3.24	Fit of hotspot example bin	58
3.25	Comparison between hotspot and simulation derived resolution	59
3.26	Signal density near the perpendicular wires	60
3.27	Example fits of the signal density	61
3.28	Comparison of resolution derived by the perpendicular wires and via simulations	62
3.29	Comparison between data driven resolution and simulation driven res- olution	63
3.30	Comparison of data-driven, simulation-driven and model predicted resolution of the CNF model	64
3.31	Far wire selection via position uncertainty	65
3.32	Position uncertainty of simulation and calibration data	66
3.33	Correlation between S2 pattern, S2 single scatter and position uncer- tainty cuts	66
3.34	Spatial distribution of events rejected by position uncertainty	67
3.35	Position uncertainty of artificial MS events	68
3.36	Measurements of the transversal diffusion constant	69
3.37	Fit of the transversal diffusion constant	70
4.1	Designed and achieved fields of different LXe TPC experiments	74
4.2	XENONnT replacement cathode mesh	75
4.3	Examples of observed defects	76
4.4	Example base image of the mesh	76
4.5	Common image processing methods	81
4.6	Training data for the segmentation models	82
4.7	Working principle of a DT classifier	83
4.8	RFC classification of all pixels of the training image	84
4.9	MLP classification of all pixels of the training image	85
4.10	Pixel labels used for validation of the segmentation models	86
4.11	Prediction values of the validation image	87
4.12	Efficiency of the segmentation models	88
4.13	Match of hexagonal mesh using algorithm	88
4.14	Leg image examples	89
4.15	Working principle of an autoencoder	90
4.16	Sketch of the VAE anomaly detection	93
4.17	Examples of ways to counteract VAE model confusion	95
4.18	Examples of skip connections and fake defects	96
4.19	ROC curves of various VAE models	97
4.20	ROC curve of the baseline VAE anomaly detection using artificial de- fects	98
4.21	Schematic of the HV test setup	99

4.22	Correlation of electrical current and image brightness	101
4.23	Breakdowns observed on the cathode mesh	102
4.24	Survival probability of the XENONnT upgrade cathode	103
5.1	Energy resolution and bias	107
5.2	S1 and S2 detection efficiencies	109
5.3	Sketch of the XENONnT event building	111
5.4	EBE estimated via the peak shadow method distributed in S1 S2 space	112
5.5	EBE estimated via the peak shadow method as a function of energy .	112
5.6	Spectrum of the data available for the cut acceptance study	114
5.7	ER band selection	115
5.8	Main and alternative S2 strength for SS and MS events	116
5.9	Acceptance of the SS selection	117
5.10	Boundaries of the S1 width cut	118
5.11	Distribution of the normalized S2 width against the S2 area	121
5.12	S2 width cut acceptance	122
5.13	Cumulative cut acceptances	123
5.14	Smoothed cumulative cut acceptances	124
5.15	Spatial and energy distributions of the simulated material background	125
5.16	Material originating MS event rate compared to MS data	126
5.17	Best fit scaler for material isotopes	127
5.18	Rescaled material originating MS event rates	127
5.19	Comparison of material background distribution and AKDE fit	128
5.20	Distribution of mean AKDE fit including 16 %, 50 % and 84 % lines from bootstrapping	129
5.21	Optimized Fiducial Volume used in the DM search	130
5.22	Size of the FV in reconstructed, corrected and true coordinates	132
5.23	Spectrum of material background remaining inside FV	133
5.24	^{222}Rn activity in various LXe TPCs	134
5.25	Decay schemes of lead isotopes	135
5.26	Spectra of ^{212}Pb events from simulated decays and from calibration data	136
5.27	MS to SS ratio of ^{212}Pb events	137
5.28	Single scatter spectra of lead backgrounds	139
5.29	Event rate of neutron activated isotopes following 40 days after AmBe calibrations	141
5.30	Energy spectrum of the ^{131m}Xe decay	142
5.31	Fit of the decay of ^{131m}Xe	143
5.32	Fit of the decay of ^{125}I	144
5.33	Event rates expected from neutron activated isotopes as a function of time	146
5.34	Spatial distribution of the primary ^{214}Bi skin decay sites	150
5.35	Spatial and energy distributions of the simulated ^{214}Bi skin background	151
5.36	Background model of SR0	153
5.37	Background model of SR1	154
5.38	Visualization of the efficiency shaper	155
5.39	Visualization of the determination of the confidence intervals	156
5.40	Projected sensitivity to the axioelectric effect	157
5.41	Projected sensitivity to the kinetic mixing parameter	158

A.1	Mapping between number of electrons and S2 top area	168
A.2	Area distribution of the simulated hitpatterns used to train the position reconstruction models	169
A.3	Predicted position distribution of models trained with different optical maps	169
C.1	Main and alternative S2 areas for SR1	176
C.2	Acceptance of the SR1 SS selection	176
C.3	SR1 boundaries of the S1 width cut	177
C.4	Acceptance of the S1 width cut	177
C.5	Acceptance of the S1 AFT cut	178
C.6	Acceptance of the S1 max PMT cut	179
C.7	Acceptance of the S2 cAFT cut	180
C.8	Acceptance of the S2 pattern cut	181
C.9	Distribution of S2 width for SR1	182
C.10	Acceptance of the S2 width cut in SR1	182
C.11	S2 width cut acceptance, using default boundaries	183
D.1	Comparison of S1 and S2 areas between simulation and calibration . .	186
D.2	Comparison of S1 AFT between simulation and calibration	187
D.3	Comparison of S1 width between simulation and calibration	187
D.4	Comparison of S2 AFT between simulation and calibration	188
D.5	Comparison of S2 width between simulation and calibration	188
D.6	Comparison of the main and alternative S2 areas between simulation and calibration	190
E.1	Fit of the decay of ^{133}Xe in SR0	192
E.2	Fit of the decay of ^{133}Xe in SR1	192
E.3	Fit of the decay of ^{131m}Xe in SR0	193
E.4	Fit of the decay of ^{131m}Xe in SR1	193
E.5	Fit of the decay of ^{129m}Xe in SR0	193
E.6	Fit of the decay of ^{129m}Xe in SR1	194
E.7	Fit of the decay of ^{127}Xe in SR0	194
E.8	Fit of the decay of ^{127}Xe in SR1	194
E.9	Fit of the decay of ^{125}Xe in SR0	195
E.10	Fit of the decay of ^{125}Xe in SR1	195
F.1	MC data with an injected DM signal	198
F.2	Distribution of the test statistic $q(N)$	199
F.3	Critical region for the 90 % exclusion limit of various DM masses . .	199

Abbreviations

$0\nu\beta\beta$	neutrinoless double beta decay
$2\nu\beta\beta$	two neutrino double beta decay
AC	accidental coincidence
AE	Autoencoder
AFT	area fraction top
AKDE	adaptive kernel density estimation
ALP	axion-like particle
AUC	area under the curve
BDT	boosted decision tree
BR	branching ratio
BSM	Beyond the Standard Model
CDF	cumulative distribution function
CE	categorical crossentropy
CEνNS	coherent elastic neutrino-nucleus scattering
CIV	charge insensitive volume
CL	confidence level
CMB	Cosmic Microwave Background
CNF	Conditional Normalizing Flow
CNN	Convolutional Neural Network
CP	charge-parity
CV	computer vision
DAQ	data acquisition
DEC	double electron capture
DFSZ	Dine-Fischler-Srednicki-Zhitnitsky
DM	Dark Matter
DP	dark photon
DT	decision tree
EBE	event building efficiency
EC	electron capture
EE	extraction efficiency
ELBO	evidence lower bound
ER	electronic recoil
FDC	field distortion correction
FEA	free electron assumption

FP	false positive
FV	Fiducial Volume
GAr	gaseous argon
GXe	gaseous xenon
HV	high voltage
JIT	just-in-time
KDE	kernel density estimation
KL	Kullback-Leibler divergence
KSVZ	Kim-Shifman-Vainshtein-Zakharov
LCE	light collection efficiency
LNGS	Laboratori Nazionali del Gran Sasso
LSP	lightest SUSY particle
LXe	liquid xenon
LY	light yield
MC	Monte-Carlo
MISE	mean integrated square error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MS	multi-scatter
MSE	mean square error
MV	muon veto
NEST	Noble Element Simulation Technique
NF	Normalizing Flow
NFW	Navarro-Frenk-White
NN	Neural Network
NR	nuclear recoil
NV	neutron veto
OH	One Hot
PCL	power constrained limit
PDF	probability density function
PE	photoelectron
PLR	profile likelihood ratio
PI	photo ionization
PMT	photomultiplier tube
PQ	Peccei-Quinn
PTFE	Polytetrafluoroethylene
QCD	Quantum Chromodynamics
QY	charge yield
RFC	random forest classifier
RGB	red green blue
RGBA	red green blue α
RGMS	rare gas mass spectrometer
ROC	receiver operating characteristic
ROI	region of interest
RRPA	Relativistic Random Phase Approximation
RRS	Radon removal system
SBI	simulation based inference
SD	spin-dependent

SE	single electron
SEG	single electron gain
SG	Skew Gauss
SI	spin-independent
SM	Standard Model
SR0	science run 0
SR1	science run 1
SR2	science run 2
SS	single-scatter
SUSY	supersymmetry
TP	true positive
TPC	Time Projection Chamber
VAE	Variational Autoencoder
VUV	vacuum ultra violet
WIMP	Weakly Interacting Massive Particle
XLZD	XENON-LUX-ZEPLIN-DARWIN

Introduction

”All models are wrong, but some are useful” is a phrase attributed to British statistician George Box [4]. One of the most successful models in all of physics is the theory of general relativity, which for example simultaneously explains the perihelion movement of mercury [5] and the gravitational waves, the minuscule deformation of space itself, emitted by the merger of two black holes [6]. And while general relativity is one of the best validated models that have ever been introduced, there are numerous cosmological observations that suggest that a complete model of the universe is still missing something. In particular, what is missing is an additional mass, that is not only five times as abundant as the known baryonic matter, but that has so far eluded detection completely. The search for the particle nature of this additional matter, called Dark Matter (DM), is one of the greatest remaining puzzles in modern physics, as DM affects the structure of our universe at all observable scales [7, 8]. While it is still unknown what DM consists of, there exists a plethora of different theories aiming to explain its nature by introducing Beyond the Standard Model (BSM) particles, such as Weakly Interacting Massive Particles (WIMPs) [9], axions [10, 11] or so-called dark photons (DPs) [12].

One of the world leading experiments for the direct detection of DM particles, by searching for scattering of DM particles off target nuclei or electrons, is the XENONnT observatory. XENONnT is a liquid xenon dual-phase Time Projection Chamber (TPC), situated in the Laboratori Nazionali del Gran Sasso (LNGS) underground laboratory in Abruzzo, Italy. This thesis presents three projects that were performed in the context of XENONnT: The improvement of the position reconstruction algorithm, the quality assurance of the replacement cathode and the sensitivity for Dark Matter absorption signals from DPs and axion-like particles (ALPs).

First, [chapter 2](#) contains a general introduction to the topic of DM and particle candidates relevant to the work presented here. In addition, it describes the working principle of the XENONnT experiment, the signal formation process, as well as the analysis steps that go into the signal reconstruction. Furthermore, the XENONnT specific jargon and software is explained.

In recent years, physics research saw an increase in the usage of Machine Learning (ML) methods for a wide variety of tasks [13]. Although algorithms based on ML were already successfully utilized in XENONnT, ML is a rapidly developing discipline offering ever more insights that can be taken advantage of. One task utilizing ML in the XENONnT analysis is the position reconstruction, the enhancement of which is the focal point of [chapter 3](#). A main goal of this thesis is the extension of the position reconstruction process to generate a probabilistic prediction, as opposed to a point-like estimate. For this purpose, two probabilistic extensions of classical Neural Networks (NNs) and thereafter the Conditional Normalizing Flow (CNF) position reconstruction model are presented. One of the main points of this work is the development and the performance characterization of the CNF model, based on simulated data as well as a validation study performed on data recorded by the detector. The novel CNF model provides an improved position reconstruction, accurate position uncertainty information, and has now been established in latest XENONnT analyses.

In summer 2025, the XENONnT detector underwent an upgrade campaign, that included the replacement of the detectors electrodes. To ensure their smooth operation, the replacement electrodes were target of extensive quality evaluation. In [chapter 4](#), the quality control of the XENONnT replacement cathode is described. The part highlighted in this work is the computer vision (CV) and ML based automatic detection of potential weak points and defects of the electrode. Multiple methods from CV and ML were implemented for image processing and segmentation, culminating in a VAE that can identify anomalous regions in images of the cathode. The various techniques described in this chapter can also be adapted for the characterization of future large-scale electrodes, such as for the planned XLZD experiment.

Finally, the sensitivity to search for absorption signals of bosonic DM in the data of the first two science runs of XENONnT, SR0 and SR1, is presented in [chapter 5](#). The focus lies on the development of the signal and background models in a large extension of the energy region from 140 keV up to 1000 keV. The models presented in this thesis can be used to fit the SR0 and SR1 data and derive 90 % CL exclusion limits of the axioelectric coupling strength g_{ae} and the dark photon kinetic mixing κ . The projected sensitivity derived in this work predicts world-leading exclusion limits in the energy range from 140 keV to 1000 keV.

The Search for Dark Matter with the XENONnT Experiment

2.1 Dark Matter

This chapter summarizes the theoretical background of Dark Matter (DM), the evidences for its existence ([subsection 2.1.1](#)) and a selection of candidate particles ([subsection 2.1.2](#)). The current status and challenges in direct DM detection are described in [section 2.2](#). Since this work is based on the XENONnT experiment, this chapter concludes with an in-depth description of liquid xenon (LXe) based DM detection ([subsection 2.3.1](#)), the XENONnT detector ([subsection 2.3.2](#)), the signal formation ([subsection 2.3.3](#)) and signal reconstruction ([subsection 2.3.4](#)) processes, a description of the data organization ([subsection 2.3.5](#)) and finally an introduction to the XENONnT software environment ([subsection 2.3.6](#)).

2.1.1 Evidences for Dark Matter

Various astronomical observations point to the existence of some non-luminous, but gravitationally interacting constituent of our universe, which is commonly referred to as Dark Matter (DM). In 1933 Fritz Zwicky coined the term DM after analyzing the motion of individual galaxies in the Coma cluster [14]. He utilized the virial theorem, that links the total kinetic and potential energies of a system of particles bound together by conservative forces. What Zwicky noticed was that to explain the observed velocities, the cluster would need to be ~ 400 times heavier than assumed from measurements of its luminosity. Zwicky explained this anomaly by proposing an additional, yet unknown, mass component that does not emit light and is therefore called *dark*.

Later, around 1970, Vera Rubin performed a survey of the rotation curves of nearby galaxies and noticed that the velocity of stars does not follow the expected profile. From Newtonian dynamics, the velocity v of stars orbiting the galactic center is expected to follow a Keplerian profile, meaning to scale as $v \sim 1/\sqrt{r}$ for large radii r .

What Rubin observed, however, was that the velocity rather approaches a constant for stars outside the bulk region of the galaxy [15]. An example of such a measured rotation curve for the galaxy NGC5055 can be seen in Figure 2.1, reproduced using data from the SPARC database [16]. The black crosses show the average rotation speed of stars at a given radius from the galactic center, derived with photometry measurements taken by the Spitzer space telescope [17]. The discrepancy between expected and measured galactic rotation profiles can be explained by adding additional non-luminous matter to the galaxies. From the difference between observed and expected rotation curves, it can be deduced that galaxies are surrounded by a diffuse halo of this additional matter. One commonly assumed density distribution of this halo is the Navarro-Frenk-White (NFW) profile [18]

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s} \left(1 + \frac{r}{R_s}\right)^2}, \quad (2.1)$$

where the normalization factor ρ_0 and the scale radius R_s are free parameters that vary for each individual galaxy. Figure 2.1 shows the best fit for the mass contributions from the baryonic mass in the galactic disk and gas, as well as the contribution from DM assuming an NFW profile.

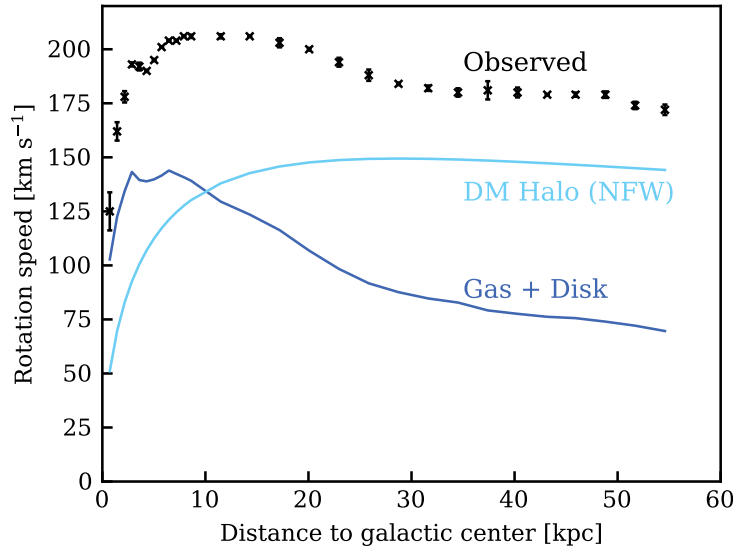


Figure 2.1: The rotation curve of galaxy NGC5055 using data from the SPARC database [16]. In black is the observed rotation speed as a function of the distance to the galactic center. The blue curves show the best fit for baryonic (dark blue) and DM (light blue) contributions, assuming an NFW density profile for the DM halo.

Indications about the properties of this new type of matter come from astrophysical observations, for example the collision of galaxy clusters. A famous example is the so-called bullet cluster 1E0657–56 [19]. The bullet cluster is the aftermath of the collision of two galaxy clusters approximately 150 million years ago. Using X-ray maps, it is possible to deduce the distribution of ordinary baryonic matter in the bullet cluster, while gravitational lensing can be employed to study the total mass distribution. The surprising result indicates that the majority of the mass of the bullet cluster is spatially separate from the luminous matter: While the ordinary matter remained near to the

site of the collision, the bulk of the mass seemed to have traveled on unhindered, i.e. collisionless, which indicates no, or limited interactions between DM and ordinary matter, and even between DM and itself [20].

The cosmological principle is an assumption that states that at the grandest scales, the universe is isotropic and homogeneous. Up until ~ 380000 years after the Big Bang, the universe was opaque and filled with an electron-proton plasma. The black body radiation in this hot early universe was strong enough to ionize any newly formed hydrogen atom immediately, ensuring the ambient photons were in a state of equilibrium with the baryons. As the universe expanded and the temperature fell, there came a point, called the recombination epoch, where the black body radiation decoupled from the plasma which allowed for neutral atoms to form. The photons released from this equilibrium transverse the universe since then and can be measured today as the so-called Cosmic Microwave Background (CMB) [21]. The CMB is the most perfect black body spectrum measured in nature, and following the cosmological principle should show the same temperature, at all positions across the sky. The latest and most accurate survey of the CMB was performed by the Planck satellite [22]. It shows the CMB as black body radiation of 2.725 K [23] with relative temperature anisotropies of the order of 10^{-5} . Figure 2.2 shows the map of those anisotropies, taken from the Planck Collaboration [24].

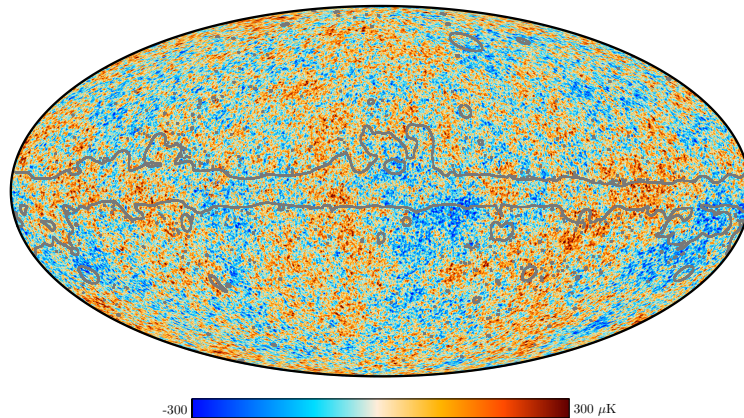


Figure 2.2: A Mollweide projection of the CMB anisotropy map, taken from [24]. It shows the deviation of the CMB temperature from the mean.

These minor anisotropies can map primordial density fluctuations in the baryon-photon fluid at the time of decoupling. In clusters of baryonic matter, the radiation pressure can oppose the gravitational attraction of masses, while for DM this is not the case because of missing interaction with photons. Therefore, regions of the universe behave slightly different, depending on the amount of DM they contain, and they directly encode this information into the CMB anisotropies [25].

Expanding the anisotropy map into a multipole spectrum can reveal the energy density of baryonic and dark matter. The multipole spectrum, as shown in Figure 2.3 taken from [7], measures the average temperature fluctuation as a function of the angular scale, where high values of the multipole l correspond to small angles. The acoustic peaks in the multipole spectrum can be excellently matched by the six parameter Λ CDM (Cold Dark Matter) model, one reason why this model is the current standard

to describe the energy content of our universe. The Λ CDM model describes the cosmological development of the universe, starting at the Big Bang, by dividing the total matter-energy distribution into three components: Ordinary baryonic matter, characterized by the density parameter Ω_b , DM, characterized by the density Ω_{DM} and *dark energy*, characterized by the cosmological constant Λ and the corresponding density parameter Ω_Λ . The best-fit values from [26] for the energy budget at present of the universe predicted by the Λ CDM model are $\Omega_\Lambda = 0.685$, $\Omega_{\text{DM}} = 0.264$ and $\Omega_b = 0.049$, in a flat universe ($\Omega_{\text{total}} = 1$).

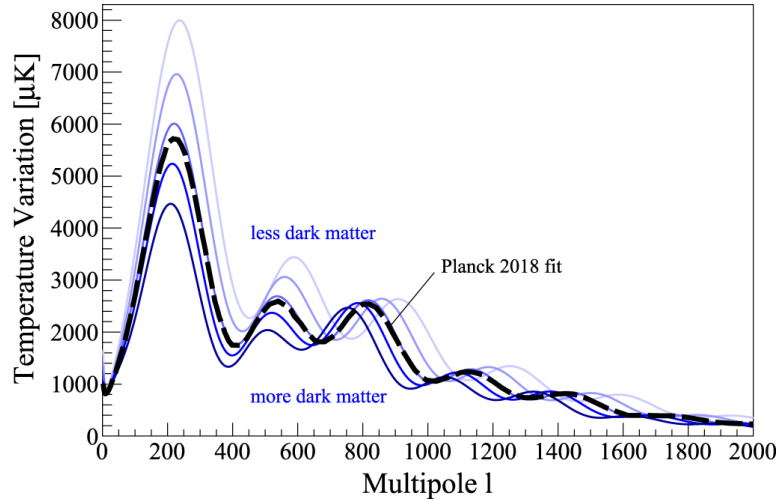


Figure 2.3: The CMB multipole spectrum, taken from [7]. The spectrum of temperature fluctuations of the CMB as function of angular scale, especially the acoustic peaks, depends on the content of DM in the universe (blue curves).

One final piece of evidence for the existence of DM comes from the large-scale structure of the universe, consisting of matter-dense *filaments* and empty *voids*. All of the structures visible in the present-day universe once started out as anisotropies in the baryon-photon fluid, and therefore in the CMB. The structure formation of the universe can be modeled using N-body simulations [27], and their results can be compared to the observations seen by sky survey experiments [28, 29]. In simulations that do not presuppose any DM, or where the simulated DM particles move relativistically, i.e. *hot dark matter*, the results do not match the structure measured by the experiments. Only by initiating the simulated universe with a large amount of non-relativistic, i.e. *cold*, and stable DM, can the observations be reproduced [30].

2.1.2 Particle Dark Matter

The gravitational effects outlined in the previous subsection 2.1.1 can be explained by a particle with the following features:

1. **Dark:** A potential DM particle can not interact with the electromagnetic force, to any so far measurable degree.
2. **Heavy:** Any DM particle must interact gravitationally, as all effects ascribed to DM are observations of gravity.

3. **Stable:** Since the mass contribution of DM has been constant since the Big Bang, any particle making up DM must not decay on cosmological time scales.
4. **Cold:** To explain the structural formation of galaxies in our universe, the majority of DM particles must have traveled slower than relativistic speeds at their time of decoupling.

No particle of the Standard Model (SM) of modern physics can fulfill all of these criteria at once, which opens the way for the search for physics BSM in the form of new particles.

No matter the particle, certain properties of DM can be constrained from astrophysical observations. Since this work focuses on the direct detection of DM, the parameters of interest are the local DM density ρ_χ and the velocity distribution of DM in the solar system $f(\vec{v}_\chi)$. The latest value for the local DM density is $\rho_\chi = (0.44 \pm 0.13) \text{ GeV}/c^2/\text{cm}^3$ [31], derived from the star survey data of the GAIA space probe [32]. However, the exact value of ρ_χ does not influence the expected spectrum measured by a direct detection experiment, but rather scales the overall event rate. More important is that the results of different direct detection experiments can be compared among one another, which requires them to use one agreed-upon canonical value for ρ_χ . The recommended convention for direct DM search, detailed in [33], is to use the Standard Halo model, with a value for the local DM density of $\rho_\chi = 0.3 \text{ GeV}/c^2/\text{cm}^3$. The second parameter is the velocity distribution $f(\vec{v}_\chi)$. In the Standard Halo model the velocity distribution is given as an isotropic Maxwell-Boltzmann distribution

$$f(\vec{v}_\chi) \propto \exp\left(-\frac{1}{2} \frac{|\vec{v}_\chi|^2}{\sigma_0^2}\right) \cdot \Theta(v_{\text{esc}} - |\vec{v}_\chi|), \quad (2.2)$$

where σ_0 is the velocity dispersion. The velocity dispersion and the the velocity of the sun around the galactic center, $v_c \approx 220 \text{ km/s}$, follow the relation $\sigma_0 = \sqrt{3/2} \cdot v_c$ [7, 8]. Θ is the Heavyside-Theta function, that limits the velocity distribution to the escape velocity of the Milky Way, which is usually taken as $v_{\text{esc}} = 544 \text{ km/s}$ [7, 8]. In addition, DM searches on earth need to take into account the movement of the earth around the sun. This orbit leads to an annular modulation in the expected rate of DM particles, with a maximum around June when the movement of sun and earth is parallel, and a minimum around December, when it is antiparallel. In case of a potential discovery of DM, this annular modulation could be used to further strengthen the claimed measurement.

WIMPs

The WIMP is a hypothetical particle, or rather a class of particles, that is favored by many theories to solve the DM puzzle, one of the reasons for its attention being that the particle density depends only on a single parameter. WIMPs are produced via a process called *freeze-out*, where some number changing process Γ is in competition with the expansion of the universe quantified by the Hubble constant H . The most simple case considers Γ to be annihilation and pair production of DM particles χ and SM particles

$$\chi + \chi \longleftrightarrow \text{SM} + \text{SM}. \quad (2.3)$$

In the early universe, there are high temperatures and densities, meaning $\Gamma \gg H$ and the WIMP particles are in equilibrium with their surroundings. The number density, n , the number of WIMPs per unit volume, is determined by the Boltzmann equation, and in this case follows

$$n_{\text{rel}} \propto T^3, \quad (2.4)$$

with T being the temperature of the universe. However, there comes a time when the universe cools down to $T < m_\chi$, making Equation 2.3 unidirectional and destroying the balance between pair production and annihilation. The total number of WIMPs decreases, determined by the non-relativistic Boltzmann equation:

$$n_{\text{non-rel}} \propto (m_\chi T)^{3/2} \exp(-m_\chi/T) \quad (2.5)$$

The freeze-out occurs when $\Gamma \ll H$, where the universe has expanded so much that annihilation also becomes negligible. From that point on the number density does no longer change, but is only diluted by the further expansion of the universe. The co-moving number density $Y = n/s$, where s is the entropy density of the universe, has the same value at the time of freeze-out as today. A more in-depth description of the freeze-out process can be found for example in [34]. We can see the WIMP density as a function of m_χ/T in Figure 2.4, with the thermal equilibrium on the left, the decrease driven by WIMP annihilation in the center and the freeze-out to the right.

For the quasi constant number of WIMPs remaining after freeze-out, the relic abundance Ω_χ , can be written as [36]

$$\Omega_\chi = \frac{T_0^3}{\rho_c} \frac{x_f}{M_P} \frac{1}{\langle \sigma v \rangle}, \quad (2.6)$$

where T_0 is the current temperature of the universe, ρ_c is the critical energy density, $x_f = m_\chi/T$ at freeze-out, M_P is the reduced Planck mass and $\langle \sigma v \rangle$ is the cross section that describes the number changing process $\Gamma = n\langle \sigma v \rangle$. The relic abundance of WIMP DM is dependent only on this cross section.

The so-called WIMP miracle is the fact that if we require the WIMP relic abundance from Equation 2.6 to amount for the whole DM content of the universe, as described in subsection 2.1.1, we are left with a cross section for the annihilation process that is at the scale of the weak interaction. However, as is discussed later in section 2.2, a vast range of potential WIMP interaction cross sections has already been excluded by experiments [34]. Theories that propose BSM particles that can be classified as WIMPs also have to motivate why the interaction cross section can be lower than the experimentally excluded limits, for example by extending the number changing process Γ . One popular extension of the SM that introduces WIMPs is supersymmetry (SUSY), which pairs each SM particle with a new supersymmetric partner [37]. In SUSY, a new quantum number, R , is introduced, with the conserved R -parity, $P_R = (-1)^{3B+L+2s}$, where B is the baryon number, L the lepton number and s the spin. Each SM particle has R -parity of $P_R = +1$ and each SUSY particle $P_R = -1$. Following from the conservation of the R -parity, the lightest SUSY particle (LSP) would be stable and thereby a candidate for WIMP DM [38]. While WIMPs receive a lot of attention from the DM community and are one of the main science channels of the XENONnT experiment, they are not the DM candidate in the focus of this thesis.

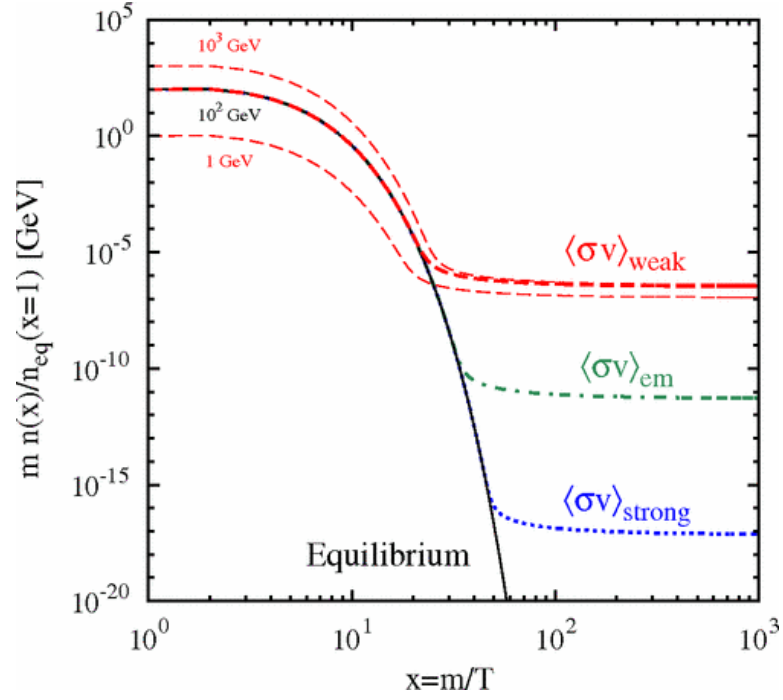


Figure 2.4: The WIMP number density as a function of time from [35]. In the early universe, meaning small values for m/T , the WIMPs are in thermal equilibrium and have a constant number density n_{eq} . As the universe cools down, WIMP production becomes energetically forbidden, while annihilation stays possible. Therefore, the number density decreases. Some time later, the WIMPs freeze out, leaving a constant relic abundance of WIMPs. The relic abundance depends on the strength of the interaction cross section σ in the number changing process Γ . Visualized in red, green and blue are different relic abundances that result from interaction cross sections of the scale of the weak force, the electromagnetic force and the strong force respectively.

Axions and Axion-like Particles

In the 1960s, it has been observed that the charge-parity (CP) symmetry is not conserved in the weak interaction [39]. There is, a priori, no reason to assume that the same should not be true for the strong interaction. In the Quantum Chromodynamics (QCD) Lagrangian, one can find the Chern-Simons term which is asymmetric under CP and therefore allows CP violation in the strong interaction [40]. The term contains an angle $\theta_{\text{QCD}} \in [0, 2\pi]$, that would in the case of $\theta_{\text{QCD}} = 0$ conserve CP symmetry. θ_{QCD} can be determined by experiments, one of which is the measurement of the neutron electric dipole moment d_n . However, the latest results of this measurement constrain the value of d_n , and set an experimental upper limit $\theta_{\text{QCD}} \lesssim 10^{-10}$ [41].

The search for an explanation of the apparent zero-ness of θ_{QCD} is known as the strong CP problem. One solution of the strong CP problem is the Peccei-Quinn (PQ) theory [42]. The PQ theory converts θ_{QCD} from a parameter to a field a , in such a way that the energy minimization automatically leads to conservation of CP symmetry. A consequence of the spontaneous symmetry breaking of the a -field is the introduction of a pseudoscalar Nambu-Goldstone boson, the so-called axion [10, 11]. The axion

mass m_a is

$$m_a \approx \frac{6 \times 10^6 \text{ GeV}}{f_a} \text{eV}/c^2, \quad (2.7)$$

where f_a is the decay constant. Given a suitable production mechanism and following relic density, the axion can serve as a candidate for particle DM.

The original axion from [10, 11] is already excluded by experiments and astrophysical observations [43]. Two common extensions of the QCD axion model are the Kim-Shifman-Vainshtein-Zakharov (KSVZ) [44] and Dine-Fischler-Srednicki-Zhitnitsky (DFSZ) [45] models. In both of these models the axion can interact with SM photons with a coupling strength $g_{a\gamma\gamma}$ [40]:

$$g_{a\gamma\gamma} = \frac{\alpha}{2\pi f_a} \cdot c_{a\gamma\gamma}, \quad (2.8)$$

where α is the fine structure constant, while $c_{a\gamma\gamma} = -1.92$ in the KSVZ model and $c_{a\gamma\gamma} = 0.75$ in the DFSZ model. This interaction is the main channel through which experiments are looking for QCD axions [46, 47].

However, the axion model can be further generalized, away from QCD and the CP problem to introduce so-called axion-like particles (ALPs). ALPs in general follow the same consideration as axions do, but they do not obey Equation 2.7, meaning their mass and scale parameter are independent of each other. An ALP can couple to electrons via the axio-electric effect with the coupling strength g_{ae} . The interactions look similar to those between photons and electrons, including axio-electric absorption, similar to the photo-effect. The cross section of the axio-electric effect is given by [48, 49]

$$\sigma_{ae} = \sigma_{pe} \cdot g_{ae}^2 \cdot \frac{3E_a^2}{16\pi\alpha m_e^2} \frac{1}{\beta} \left(1 - \frac{\beta^{2/3}}{3}\right), \quad (2.9)$$

where E_a and β are the energy and velocity of the ALP respectively, m_e is the electron mass and σ_{pe} is the cross section of the photo-electric effect. The search for ALPs via the axio-electric effect in XENONnT is one of the main science subjects of this thesis.

Dark Photons

One hypothesis for why dark matter has eluded detection despite decades of search is that it belongs to a *hidden sector*, a system of particles that interact among each other via a new massive gauge boson, a so-called dark photon (DP) A' [12, 50]. When particles of the hidden sector are charged under the SM hypercharge and this new dark charge, the DP can kinetically mix with SM photons γ , as outlined in the diagram in Figure 2.5, taken from [51]. This kinetic mixing provides a *dark portal* allowing physicists to probe the hidden sector.

A hidden sector with a heavy mediator could explain some astrophysical observations: The Fermi Large Area Telescope observed a yet unexplained excess of γ -rays originating from the galactic center, which could be explained by DM annihilation [52]. X-ray telescopes have observed a monochromatic line at 3.56 keV that might not match any atomic transition energies [53]. An explanation for this line includes inelastic self-scattering of DM transmitted via DP.

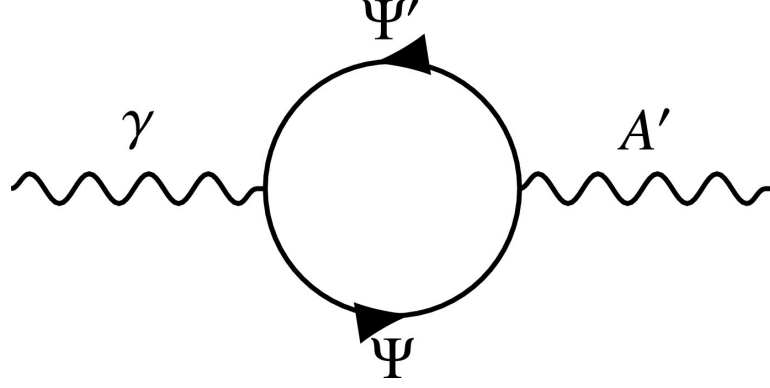


Figure 2.5: A diagram outlining the kinetic mixing of a DP A' and a SM photon γ , from [51]. The particles ψ, ψ' belong to the hidden sector and have to be charged under SM hypercharge and the new dark charge, propagated by the DP A' .

The DP is characterized by the kinetic mixing parameter¹ κ that quantifies the strength of the coupling between A' and γ . Assuming a minimal secluded model, meaning one where the new dark gauge group is the only one present in the hidden sector, κ can be approximated as [51]

$$\kappa \approx \frac{g_Y g_D}{16\pi^2} \log \left(\frac{m_\psi}{m_{\psi'}} \right), \quad (2.10)$$

with the SM electroweak coupling g_Y , the dark coupling constant g_D between A' and ψ and the masses $m_\psi, m_{\psi'}$ of the DM particle doublet of the kinetic mixing. Analogous to the SM photo-effect, a DP can be absorbed by an atom, depositing its energy, given by the DP mass $m_{A'}$. The cross section of the DP absorption is approximated as [55]

$$\sigma_{\text{DP}} \approx \kappa^2 \sigma_{\text{pe}} \frac{1}{\beta}, \quad (2.11)$$

where β is the DP velocity and σ_{pe} is the SM photo-effect cross section. The search for absorption signals of DPs is another one of the scientific subjects of this thesis.

2.2 Direct Detection of Dark Matter

The experimental measurement of the particle nature of DM is one of the grandest ventures of modern physics. There are in general three ways of searching for DM: Production, indirect search and direct search.

The search for DM production happens in particle colliders. In simple terms, after colliding high energy particles, physicists are looking in the cascade of resulting product particles for missing energies carried away by undetectable DM. A review of DM production experiments can be found for example in [56]. Indirect DM search is the process of looking for SM particle production from the annihilation or decay of DM particles. Such experiments are searching for anomalous excess of common decay products, like photons, near astronomical objects with a high expected density of DM. A review of indirect DM search is given for example in [57].

¹This work uses κ , following the convention of [54], while the mixing parameter is also called ϵ in literature [51].

This work focuses on the direct search for DM, where the base assumption is that DM can have some weak coupling to SM matter. Detectors use the passage of earth through the galactic DM halo to trigger interactions between DM particles and a SM target. The deposited energy is read out as a light, charge, or heat signal and experiments are specialized to collect one or two of those signals. A review of direct DM detection can be found for example in [7] and [8].

The rate of interactions measured by a direct detection experiment is $R = N\phi\sigma$, where N is the number of target atoms, ϕ is the incident flux of DM particles and σ is the specific DM SM interaction cross section. N is experiment dependent, ϕ can be derived from astrophysical observations in our galaxy described in subsection 2.1.2 and finally, σ depends on the specific DM model that is searched for. For scattering interactions, such as expected from WIMPs, that deposit some energy E_R , the spectral rate can be written as

$$\frac{dR}{dE_R} = \frac{\rho_\chi}{m_\chi m_T} \int_{v_{\min}}^{v_{\max}} \vec{v}_\chi f(\vec{v}_\chi) \frac{d\sigma}{dE_R} d\vec{v}, \quad (2.12)$$

with the DM velocity distribution $f(\vec{v}_\chi)$ from Equation 2.2, the local DM density ρ_χ and the DM mass m_χ , as well as the target mass m_T . The velocities $v_{\min}$ and $v_{\max}$ are the minimum necessary velocity for scattering and the maximum velocity of DM in the galaxy, respectively. The differential cross section for WIMP DM can be split into spin-dependent (SD) and spin-independent (SI) parts, as [7]

$$\frac{d\sigma}{dE_R} = \frac{m_N}{2\vec{v}_\chi^2 \mu^2} (\sigma_{\text{SI}} F_{\text{SI}}^2(E_R) + \sigma_{\text{SD}} F_{\text{SD}}^2(E_R)), \quad (2.13)$$

where m_N is the mass of the target nucleus and μ is the WIMP-target nucleus reduced mass. The form factors F_{SI} and F_{SD} characterize the dependence on the momentum transfer of the interaction and nuclear structure of the target: For small momentum transfers, the WIMP scatters coherently on all the nucleons of the target atom, but for larger momentum transfers, the WIMP's de Broglie wavelength becomes too small and the target nucleus only participates partially in the scattering process. The cross sections σ_{SI} and σ_{SD} are the SI and SD interaction cross sections at zero momentum transfer respectively. In the SI case, the cross section can be expressed as [7]

$$\sigma_{\text{SI}} = \sigma_n \frac{\mu^2}{\mu_n^2} \frac{(f_p Z + f_n (A - Z))^2}{f_n^2} = \sigma_n \frac{\mu^2}{\mu_n^2} A^2, \quad (2.14)$$

where σ_n is the WIMP-nucleon cross section, μ_n the WIMP-nucleon reduced mass and the factors f_p and f_n the coupling strength of WIMPs to protons and neutrons respectively. Assuming $f_p = f_n$, the SI interaction cross section simplifies to the second expression, which highlights the dependence on the square of the mass A of the target nucleus.

The current experimental best limits on the SI WIMP nucleus cross section are depicted in Figure 2.6. The gray shaded region is the so-called *neutrino fog*, which is reached when a direct detection experiment becomes sensitive enough to efficiently measure coherent elastic neutrino-nucleus scattering (CEvNS) from solar and atmospheric neutrinos. Since the flow of solar neutrinos through any earth-based detector

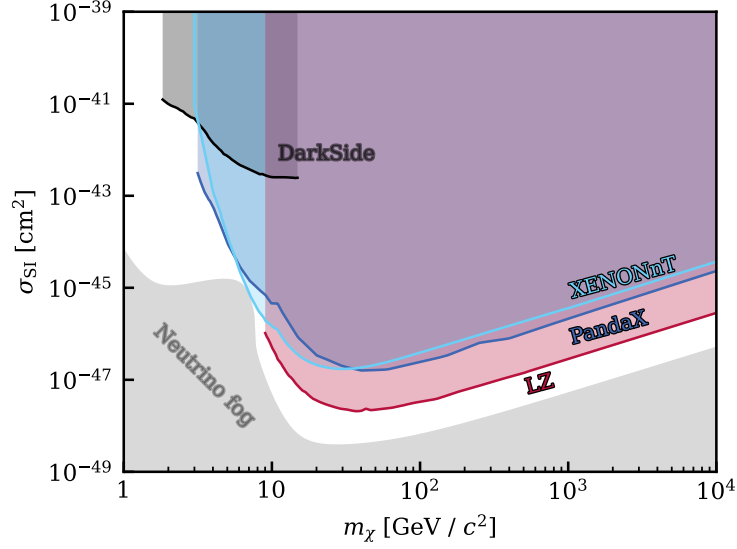


Figure 2.6: The current exclusion limits from direct detection experiments for the spin-independent WIMP nucleus scattering cross section, σ_{SI} , at 90 % CL. The gray shaded region is the neutrino fog, the region where potential DM signals are expected to be concealed by a high rate of coherent scattering from solar and atmospheric neutrinos [59]. Data provided by [60]. DarkSide: [61], PandaX: [62, 63], LZ: [64], XENONnT: [65, 66]

is immense, at that point, the neutrino signal will become a drowning background for any rare event search [58, 59].

The work presented in [chapter 5](#) describes the search for signals due to absorption of bosonic DM particles, such as ALPs and DPs. The rate computation for absorption signals differs from that based on particle scattering described in [Equation 2.12](#). In this work, the aim is to search for ALPs and DPs as DM, i.e. from the galactic halo as opposed to from local sources such as the sun. For this reason, it can be assumed that the velocities in [Equation 2.9](#) and [Equation 2.11](#) are non-relativistic, which simplifies the cross sections to

$$\sigma_{\text{ae}} \approx \sigma_{\text{pe}} g_{\text{ae}}^2 \frac{3m_{\text{a}}^2}{16\pi\alpha m_{\text{e}}^2} \frac{c}{v_{\text{a}}} \quad \text{and} \quad \sigma_{\text{DP}} \approx \sigma_{\text{pe}} \kappa^2 \frac{c}{v_{\text{A'}}}. \quad (2.15)$$

Assuming either particle accounts for all DM in the galaxy, the flux ϕ_i is $\phi_i = \rho_{\chi}/m_i \cdot v_i$ with $i \in \{\text{a}, \text{A'}\}$. Specifically to note is that the velocity of the DM particles cancels out in the event rate. The computation of the event rates is taken from [49], which follows [67]. The direct detection event rates for a target using atoms of atomic number A are

$$R_{\text{ae}} \simeq \frac{1.5 \times 10^{19}}{A} g_{\text{ae}}^2 \left(\frac{m_{\text{a}}}{\text{keV}/c^2} \right) \cdot \left(\frac{\sigma_{\text{pe}}}{\text{b}} \right) \text{kg}^{-1} \text{day}^{-1} \quad (2.16)$$

for the axio-electric effect and

$$R_{\text{dp}} \simeq \frac{4.7 \times 10^{23}}{A} \kappa^2 \left(\frac{\text{keV}/c^2}{m_{\text{A'}}} \right) \cdot \left(\frac{\sigma_{\text{pe}}}{\text{b}} \right) \text{kg}^{-1} \text{day}^{-1} \quad (2.17)$$

for the absorption of DPs, with the photoelectric cross section, σ_{pe} , in units of barns ($1 \times 10^{-24} \text{ cm}^2$).

The current best exclusion limits for g_{ae} and κ are shown in Figure 2.7. The current best limits for a mass range of approximately $[0, 100] \text{ keV}$ are provided by the XENONnT experiment, discussed in [54]. This thesis, in particular chapter 5 can be seen as a continuation and update to that study.

For a test mass of $m_i = 10 \text{ keV}/c^2$ ($i \in \{a, A'\}$) and the corresponding lowest excluded values for $g_{\text{ae}} = 2.45 \times 10^{-14}$ and $\kappa = 1.45 \times 10^{-15}$ from [54], the expected rate of dark absorption events can be computed from Equation 2.16 and Equation 2.17, using a mean atomic weight for natural Xe of $A = 131.3$. The resulting rates are $R_{\text{ae}} = 2.5 \times 10^{-5} \text{ kg}^{-1} \text{ day}^{-1} = 9.1 \text{ t}^{-1} \text{ y}^{-1}$ and $R_{\text{dp}} = 2.7 \times 10^{-5} \text{ kg}^{-1} \text{ day}^{-1} = 10.0 \text{ t}^{-1} \text{ y}^{-1}$. A detector of target mass 1 t would have to continuously measure for one year to expect about 10 events from the absorption of dark bosons, based on already excluded coupling constants at 90 % CL, all while disregarding finite detection efficiency. This is to highlight the rare nature of such events.

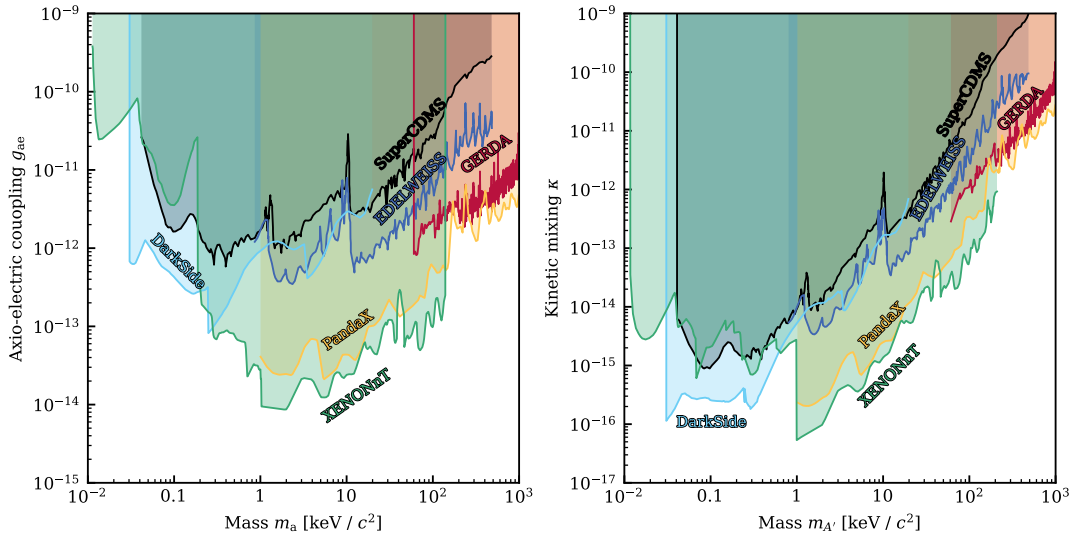


Figure 2.7: The current exclusion limits on the characteristic parameters g_{ae} (left) and κ (right) from direct detection experiments. Data provided by [60]. SuperCDMS: [68], EDELWEISS: [69], DarkSide: [70], GERDA: [71], PandaX: [72], XENONnT: [54, 66]. The shaded regions indicate coupling strengths that are excluded by the listed experiments with a 90 % CL.

2.3 Dual Phase Liquid Xenon Time Projection Chambers and the XENONnT Experiment

The main science goal of XENONnT is the search for WIMP DM via elastic scattering on the Xe nuclei, the latest result of which was shown in Figure 2.6. However, it is also sensitive to CE ν NS, and has reached the point of measuring solar neutrino interactions via CE ν NS [73]. Further, XENONnT has probed the edges of the neutrino fog for WIMP signals originating from WIMPs with masses of $m_\chi \sim 6 \text{ GeV}$ [74]. The experiment is also sensitive to lighter DM candidates by using subsets of data

like ionization only signals [66] or by considering astrophysical signal boosting processes like solar reflection [75]. In addition, bosonic DM can also be investigated in XENONnT, as performed in [54] and laid out in chapter 5. Other science channels investigated besides DM concern for example quantum collapse [76] or neutrinoless double beta decay ($0\nu\beta\beta$) studies [77]. Furthermore it is planned to measure the solar neutrino spectrum, and XENONnT is part of the supernova early warning system SNEWS [78].

2.3.1 Liquid Xenon as Detector Medium

LXe is an excellent detector medium to search for rare events, especially to search for WIMP DM, the primary science channel of XENONnT. Noble gases are scintillators, providing a light signal on a received energy deposit. Both liquid Argon and LXe are commonly used in direct detection experiments, such as XENON [79], LZ [80], PandaX [81], DarkSide [82] or EXO [83], or the currently planned next generation detector XENON-LUX-ZEPLIN-DARWIN (XLZD) [84].

Xenon is the heaviest of the stable noble gasses with an atomic number of $Z = 54$ and atomic weight $A = 131$. The high mass of the Xe nucleus is optimal for the energy transfer of WIMPs with masses often predicted in the GeV/c^2 to TeV/c^2 range. Further, as expressed by Equation 2.14, the SI interaction cross section scales with the target atomic mass squared. This is exemplified in Figure 2.8, where the expected recoil spectra between WIMPs and common direct detection target nuclei are shown, for an example WIMP mass of $m_\chi = 100 \text{ GeV}/c^2$ and cross section $\sigma_{\text{SI}} = 1 \times 10^{-47} \text{ cm}^2$. In the regime of low energy transfer, the rate for the Xe target is larger than that of the lighter targets. However, for increased energies, the form factor F_{SI} starts to strongly suppress the interaction rate.

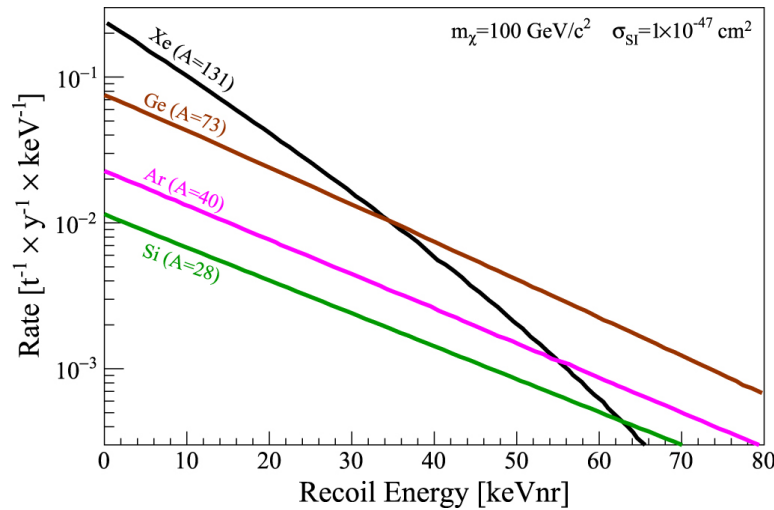


Figure 2.8: The recoil spectra, expected from the scattering of a WIMP of mass $m_\chi = 100 \text{ GeV}/c^2$ for a SI cross section of $\sigma_{\text{SI}} = 1 \times 10^{-47} \text{ cm}^2$ for different target nuclei, from [7]. At low recoil energies, the rate for a Xe target is higher than that using lighter target nuclei, but at higher energies the expected rate for Xe is suppressed by the form factor F_{SI} .

The natural abundance of Xe isotopes is shown in Table 2.1 [85]. Natural Xe contains isotopes with an odd number of neutrons, which also allows to probe the WIMP

Table 2.1: Natural abundances of Xe isotopes, from [85]

Isotope	Abundance [%]
^{124}Xe	0.095
^{126}Xe	0.089
^{128}Xe	1.910
^{129}Xe	26.401
^{130}Xe	4.071
^{131}Xe	21.232
^{132}Xe	26.909
^{134}Xe	10.436
^{136}Xe	8.857

spin-dependent interaction cross section. In addition, all of the naturally occurring Xe isotopes are stable, with the exception ^{124}Xe and ^{136}Xe , and potentially ^{134}Xe , reducing the intrinsic backgrounds. In addition, the rare decays of ^{124}Xe , ^{136}Xe , and ^{134}Xe provide science cases of their own.

The isotopes ^{124}Xe and ^{136}Xe can undergo double electron capture (DEC) and two neutrino double beta decay ($2\nu\beta\beta$) respectively. Both of these decays can be used to probe the nature of neutrinos, since observing any process of the two as neutrinoless would indicate neutrinos as Majorana particles. For this reason, both ^{124}Xe and ^{136}Xe are of scientific interest [86, 87]. In the SM ^{134}Xe could also decay via $2\nu\beta\beta$, however this decay has never been observed, and large-scale LXe experiments comprise a prime method to search for it [88].

Another great advantage of LXe as a detector medium is its self-shielding property. Figure 2.9 shows the mean free path of α , β and γ radiation in liquid xenon, for a given density of $\rho = 2.86 \text{ g/cm}^3$ [89, 90]. High energy γ -rays ($> 1 \text{ MeV}$) can travel multiple 10s of cm, but α and β particles will deposit their energy very quickly after entering LXe. This is taken advantage of in LXe detectors to shield an inner sensitive detector volume from radiation originating for example from detector materials surrounding the Xe target. The process of using the self-shielding of xenon to protect an inner sensitive volume is called *fiducialization*.

In addition, xenon has a relatively warm boiling point. At the operational pressure of XENONnT $p \sim 1.9 \text{ bar}$, the boiling point is $T_b \sim 177 \text{ K}$ [91], which makes the required cryogenics system much simpler as for example compared to Argon, with a boiling temperature of $T_b \sim 87.9 \text{ K}$. The density of LXe under these conditions is typically $\rho = 2.86 \text{ g/cm}^3$.

Finally, the average energy to produce detectable quanta, ion electron pairs or a scintillation photon, in xenon is $W = 13.7 \text{ eV}$ [92]. This number is lower than the equivalent in e.g. Argon, with $W_{\text{Ar}} = 19.5 \text{ eV}$ [93], which allows LXe detectors to probe events with a lower energy threshold. In addition, LXe is transparent for the vacuum ultra violet (VUV) light that is produced by the proper scintillation process, since the photons are emitted by a short-lived dimer as opposed to the naturally occurring unpaired atoms. More details on the signal formation process can be found in subsection 2.3.3.

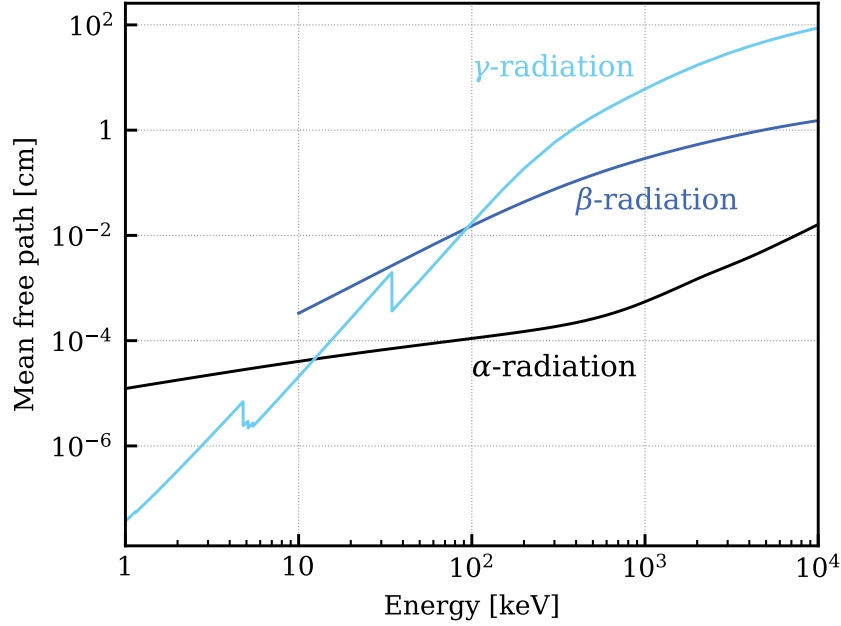


Figure 2.9: The mean free path of α , β and γ radiation in liquid xenon, for a density of $\rho = 2.86 \text{ g/cm}^3$. The data is extracted from [89, 90].

2.3.2 The XENONnT Experiment

XENONnT is an experiment for the direct detection of Dark Matter (DM), using a liquid xenon (LXe) dual-phase Time Projection Chamber (TPC), a detector concept first introduced by David Nygren in [94]. To shield it from backgrounds induced by cosmic radiation, the XENONnT experiment is located 3400 mwe below ground, in the Laboratori Nazionali del Gran Sasso (LNGS) underground facility in Abruzzo, Italy. It is the fourth iteration of the XENON project, a family of experiments dating back to the XENON10 detector [95] that was started to be developed in 2002. The latest experiment includes a three story service building that houses cryogenics infrastructure, data acquisition (DAQ) [96] and purification and distillation systems that remove contamination from the xenon [97]. The total LXe inventory of XENONnT is $\sim 8.6 \text{ t}$, with $\sim 5.4 \text{ t}$ of that being used as the active detector target volume. A more in-depth description of the detector and all of its auxiliary systems can be found in [79].

The XENONnT experiment consists of three nested detectors: A muon veto (MV), a neutron veto (NV) and a Time Projection Chamber (TPC) from outside to inside. The outermost layer, the MV is housed in a stainless steel tank of $\sim 10 \text{ m}$ diameter and height, containing $\sim 700 \text{ t}$ of demineralized water, see Figure 2.10. This water tank serves as a passive shield against radiation from the experimental environment. In addition, it is equipped with 84 Hamamatsu R5912ASSY photomultiplier tubes (PMTs) distributed on the inside of the tank that enable its usage as an active veto detector against muons [98]. Muons produced in cosmic ray showers can reach the detector even through the shielding provided by the underground laboratory setting. Via spallation or further cascades, they can introduce additional neutron radiation at any position in the experiment. To counteract any potentially produced neutrons, XENONnT rejects any data taken within certain time of the MV registering a muon signal.

Inside the MV is an additional section, specialized to veto neutrons: the NV equipped with 120 8 inch Hamamatsu R5912-100-10 PMTs. The water in the entire water tank is doped with gadolinium salt. The Gd has the purpose of capturing neutrons, which releases γ -rays, that can subsequently be detected via Cherenkov radiation. The tagging efficiency of this NV reached 82 % [99].

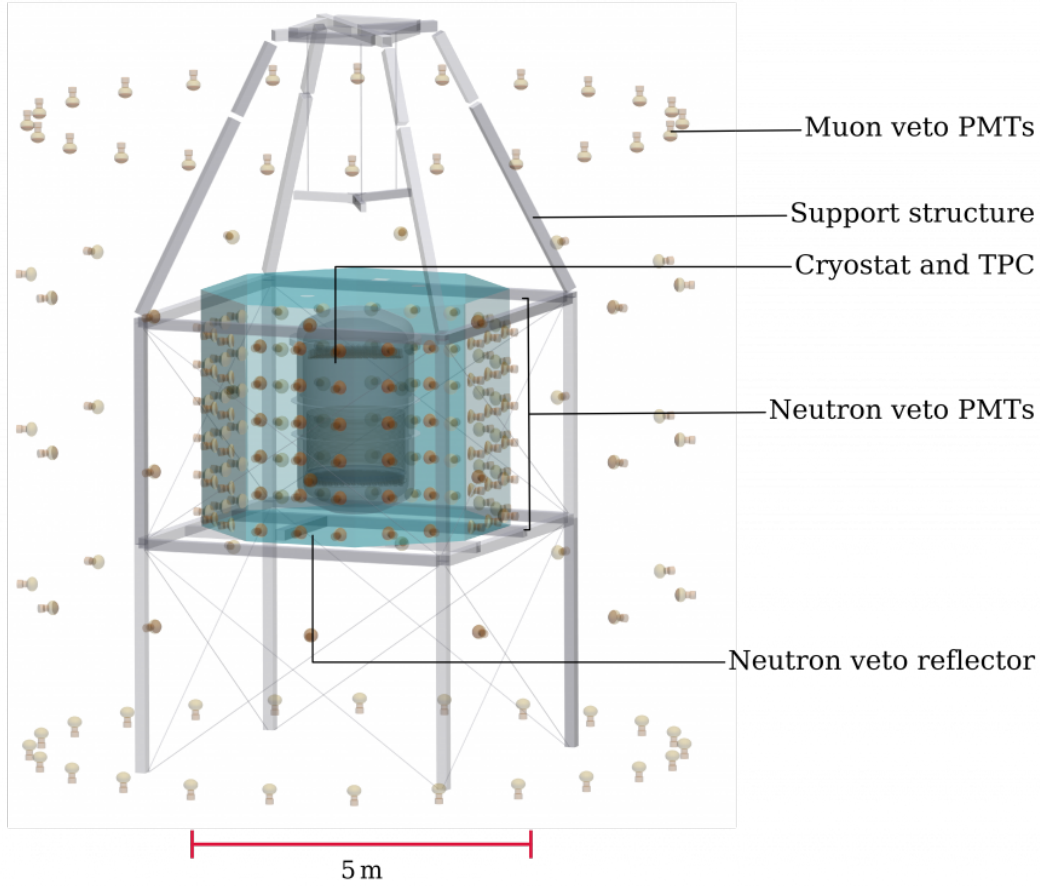


Figure 2.10: A rendering of the veto systems of the XENONnT experiment. The outermost layer is a water Cherenkov muon-detector that acts as a veto. Inside that is a neutron detector using Gd doped water, that serves as a secondary veto detector. From [100].

The heart of the experiment is the LXe dual-phase TPC, visualized in Figure 2.11. It is housed in a double walled stainless steel cryostat inside of the NV and was designed to be as light and radiopure as feasible [101]. The inside of the cryostat is lined with Polytetrafluoroethylene (PTFE) panels that have a high reflectivity for the wavelength of the Xe VUV scintillation light. The radius of the inside of the PTFE, in future referred to as the TPC radius, is on average $r_{\text{TPC}} = 66.4 \text{ cm}$. This number is an average, because the PTFE lining does not form a perfect circle. The top and bottom of the TPC are instrumented with photosensor-arrays using the Hamamatsu R11410-21 3 inch PMT, 253 at the top and 241 at the bottom, which form a total of 494. The circular R11410-21 PMTs are arranged in a hexagonal pattern, and since this is not space filling, the areas in between individual PMTs are also covered by PTFE.

Throughout the TPC there are five electrodes in two stacks, referred to as bottom grids and top grids in Figure 2.11. The bottom stack consists of a cathode and a bottom

screening electrode, that is designed to protect the bottom PMT array from high voltage. The top stack consists of a gate electrode, an anode and a top screening electrode, the latter of which serves the same purpose as the bottom screening electrode, but for the other PMT array. The other three electrodes span two electric fields: the drift field E_{drift} between cathode and gate and the extraction field E_{ext} between gate and anode, whose purpose is explained in [subsection 2.3.3](#). Inside the cryostat walls, but also covered by the PTFE lining, is a field cage made of copper field shaping elements, that ensures highest homogeneity of the above electric fields [102]. All electrodes that were implemented in XENONnT during the periods relevant for this thesis consist of parallel wires spanned across a frame, all made of stainless steel. The anode and gate have wires of 0.216 mm diameter and a pitch of 5 mm between individual wires, while the cathode has 0.304 mm diameter wires with a wire pitch of 7.5 mm. Additionally, the gate and anode electrodes are supported by 2 and 4 wires, respectively, of diameter 0.304 mm each, that are mounted perpendicular to the main electrode wires. This was done in a later step just before installation, to counteract field distortion due to wire sagging, driven by gravity and electrostatic force. The distance between cathode and gate is 148.6 cm and the distance between gate and anode is 0.8 cm. The coordinate system commonly used in XENONnT sets the center of the gate as the coordinate origin, disregarding of any potential sagging.

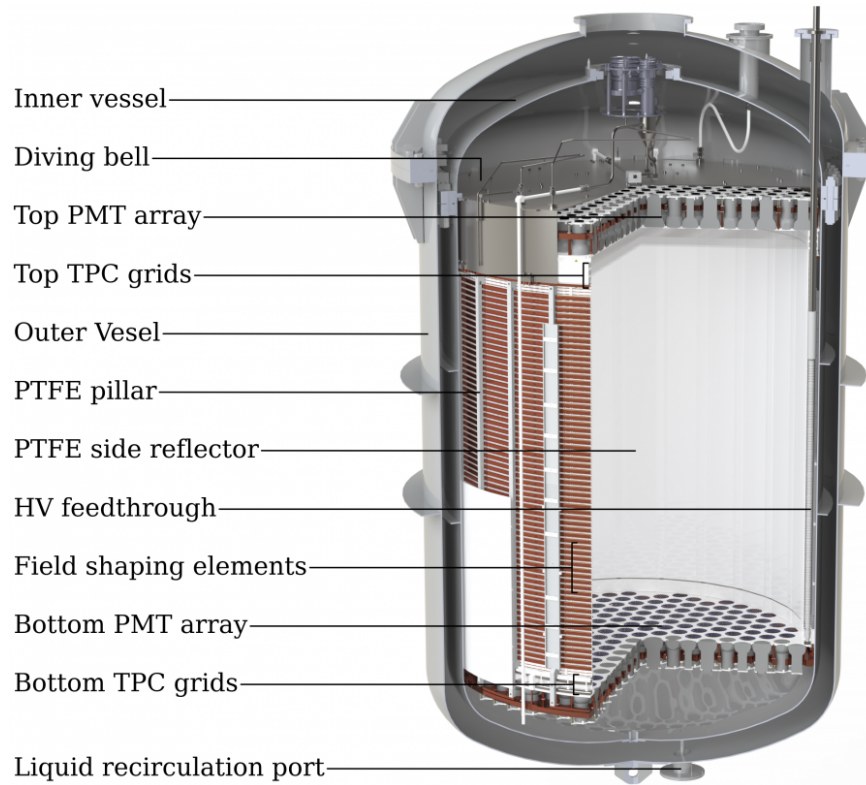


Figure 2.11: A schematic of the double-walled cryostat of the XENONnT experiment. From [100].

2.3.3 The Signal Formation Process in liquid Xenon

The TPC is filled with a large volume of LXe and a smaller volume of gaseous xenon (GXe) on top of that, with the liquid gas interface typically in between the

gate and anode. A not to scale sketch of the XENONnT TPC is shown in [Figure 2.12](#), highlighting the parts that are significant for the signal formation process.

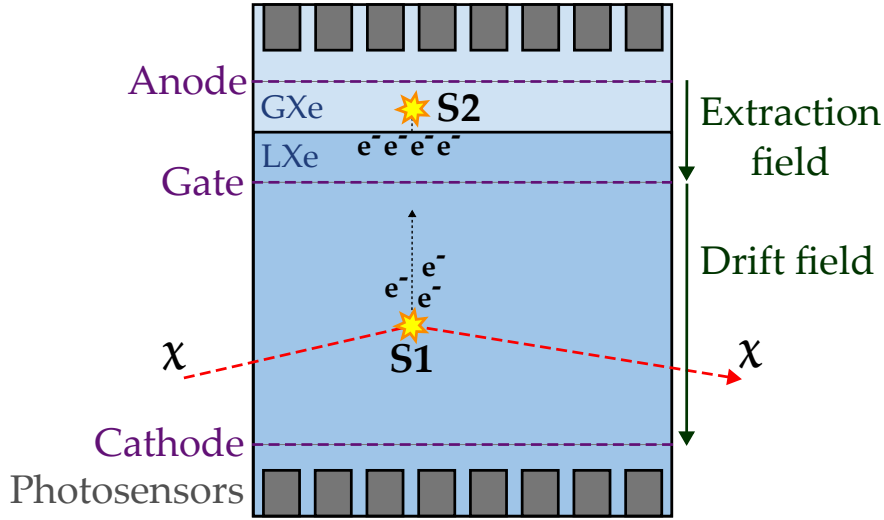


Figure 2.12: A sketch (not to scale) of a LXe TPC and the signal formation process. When a particle deposits energy in the TPC, the Xe atoms will scintillate (S1) and ionize. The ionization electrons are drifted along the drift field to the top of the TPC, where they are extracted into the GXe by the extraction field. There they are accelerated, leading to a secondary scintillation pulse (S2). The photons from the S1 and S2 are collected by photosensors, in the case of XENONnT photomultiplier tubes (PMTs).

When a particle scatters off one of the atoms in the target volume, it causes the Xe atoms to scintillate and to ionize. The scintillation light is collected by the PMTs and forms the so-called S1 signal. The freed electrons can either recombine with one of the ions, thereby adding to the S1 signal, or they can drift along the drift field between the cathode and gate to the top of the detector. The extraction field between the gate and the anode serves two purposes: The first is to extract the drifted electrons from the liquid to the gaseous phase, and the second is to accelerate the extracted electrons, creating a secondary scintillation pulse. This secondary pulse is also caught by the photo-sensors and is referred to as S2 signal.

The individual S1 and S2 pulses are referred to as *peaks* in the XENONnT analysis. This is to distinguish them from *events*, which refers to the pairing of an S1 and S2. The initial energy of an event E_R can be determined by the strength of the S1 and S2 signals, which is outlined below. The 3D position of an event in a LXe TPC can be fully reconstructed from the combination of S1 and S2: The vertical depth, the z -position, can be derived from the electron drift speed and the time difference between S1 and S2. The horizontal position (x, y) can be determined by the pattern that the S2 creates in the top PMT array. An in-depth description of the (x, y) position reconstruction follows in [chapter 3](#), the improvement and extension of which is one of the main outcomes of this thesis.

The quanta formation process is visualized in [Figure 2.13](#). The produced quanta are either free electrons or VUV photons of wavelength $\lambda = 178 \text{ nm}$. The VUV light is emitted by a diatomic excimer, that can decay via two channels, a singlet and a triplet state, with lifetimes of 4 ns and 27 ns, respectively [[103](#)]. The energy required to form

either quanta, W , is commonly assumed to be $W = 13.7 \text{ eV}$ [92]. The total energy of the recoil, can be written as

$$E_R = \frac{W}{L}(n_{\text{ex}} + n_{\text{ion}}), \quad (2.18)$$

with the number of excimers n_{ex} and the number of electron-ion pairs n_{ion} . The parameter L is the Lindhard quenching factor that accounts for non-detectable energy carried away via atomic motion [104–106]. The Lindhard factor depends on the type of interaction causing the excitation and ionization, which is described later. The ratio of the numbers of initial excimers to electron-ion pairs can be described by the empirical function

$$\frac{n_{\text{ex}}}{n_{\text{ion}}} = \alpha E_{\text{drift}}^{-\zeta} \cdot (1 - e^{-\beta E_R}), \quad (2.19)$$

where α , β and ζ are parameters determined by measurement [106]. The ratio depends on the electric drift field, E_{drift} , and the recoil energy, E_R .

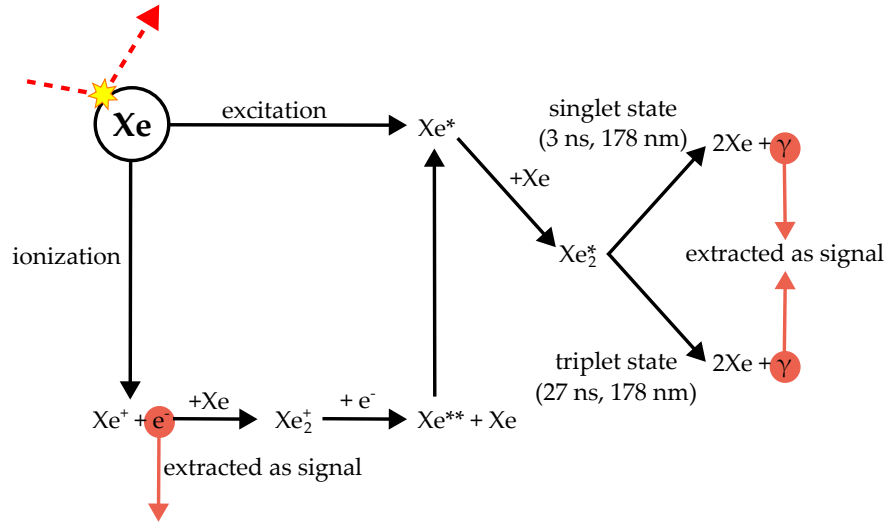


Figure 2.13: A sketch of the quanta generation after an energy deposit in xenon, inspired by [107, 108]. After a recoil (red dashed) the Xe atoms in the immediate environment become ionized or excited. Ionized atoms can recombine with ionization electrons. The remaining electrons get extracted by the electric field. The excited atoms and the recombined atoms both lead to diatomic excimers Xe_2^* that decay via a triplet or singlet state, releasing VUV light with a wavelength of $\lambda = 178 \text{ nm}$.

The actually detectable quanta are the total number of emitted photons, including the ones produced via recombination of ions, $n_\gamma = n_{\text{ex}} + r \cdot n_{\text{ion}}$, and the number of free electrons that do not recombine with an ion $n_e = (1 - r) \cdot n_{\text{ion}}$. The recombination rate r can be parameterized with the Thomas-Imel model [109]

$$r = 1 - \frac{\ln(1 + n_{\text{ion}}\varsigma)}{n_{\text{ion}}\varsigma}, \quad (2.20)$$

where ς encodes the dependence on the electric drift field and is again empirically determined by measurement [106].

In addition, there is a dependence on the type of interaction that caused the energy deposit E_R . Incoming particles can interact in one of two ways with the Xe atoms,

either with the electron cloud via electronic recoil (ER), or with the Xe nucleus via nuclear recoil (NR). As intuitively expected, an ER interaction leads to a higher ratio of electron-ion pairs, as compared to an NR interaction. This thesis concentrates on the search for bosonic DM, which is expected to interact with LXe via ER. The Lindhard factor L for ER interactions can be assumed to be 1. The relation of total energy and produced quanta, Equation 2.18, simplifies to

$$E_R = W \cdot (n_\gamma + n_e). \quad (2.21)$$

The light yield (LY) and charge yield (QY), n_γ and n_e per energy, for ER and NR interactions for electric fields of $E_{\text{drift}} \in \{20, 200\}$ V/cm, computed using the Noble Element Simulation Technique (NEST) software [110], is shown in Figure 2.14. A higher electric field leads to an increase in QY and decrease in LY, based on suppressed recombination.

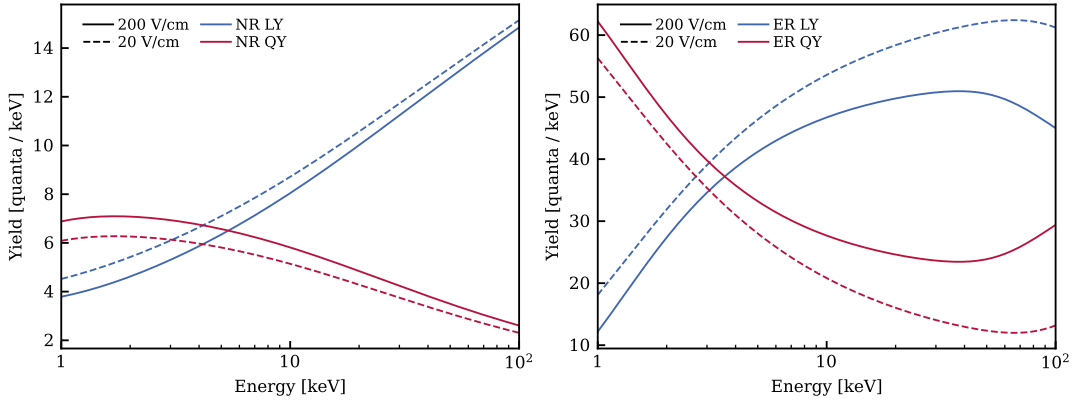


Figure 2.14: Quantum yields of NR (left) and ER (right) interactions in LXe, computed using the NEST software [110] for two different drift field values.

The difference in yields between ER and NR is taken advantage of in XENONnT. Figure 2.15 shows the corrected S1 and S2 strength of two populations of events. A description of the meaning of corrected signal strength and the units follows in subsection 2.3.4. Shown in blue are events caused by the β -decay of ^{212}Pb , a daughter of ^{220}Rn which is used as a calibration source. These events are ERs. The events shown in golden originate from an AmBe calibration source, which provides neutrons via an (α, n) process, meaning those events are NR. The populations follow individual bands, one for ER and one for NR, whose separation increases with increasing E_{drift} . The ER and NR bands are used to remove certain background events. For example WIMP events are expected to interact via NR. Therefore the search for WIMP DM removes the ER band from the analysis space, reducing the background by a significant margin.

2.3.4 The Signal Correction Process of XENONnT

Due to various reasons, the generated photons and electrons from an event can be lost before they are measured, which means that the measured values $S1$ and $S2$ do not equal the the number of quanta (after recombination) n_γ and n_e . The signal correction process aims to introduce a quantification ϵ of the lost quanta to generate *corrected*

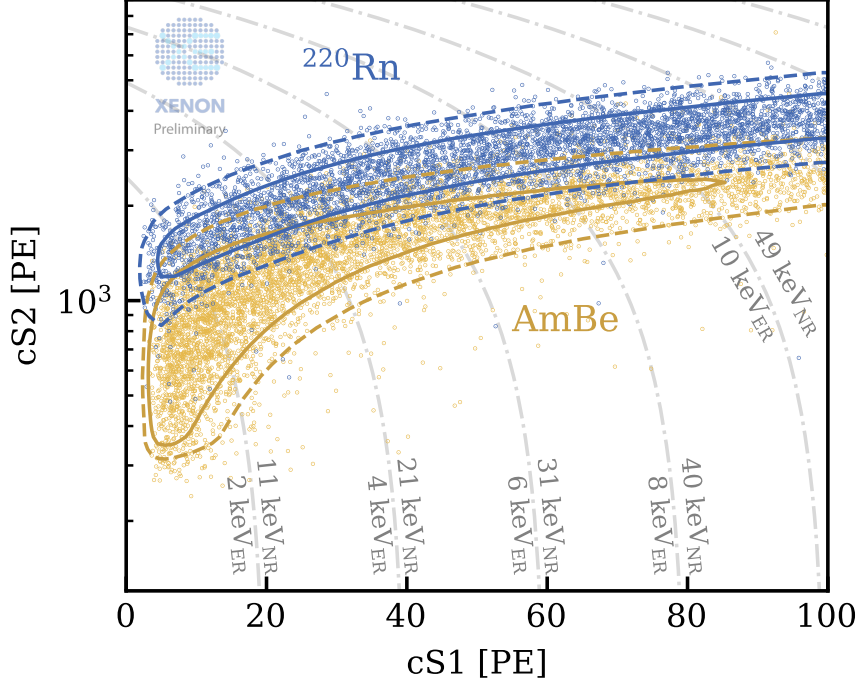


Figure 2.15: Corrected S1 and S2 signal strength for different calibration data, from [111]. The ^{220}Rn calibration leads to ER events (blue), while the AmBe calibration leads to NR events.

signal strengths $cS1 = \epsilon_{S1} \cdot S1$ and $cS2 = \epsilon_{S2} \cdot S2$, such that Equation 2.21 can be converted to

$$E_R = W \cdot \left(\frac{cS1}{g_1} + \frac{cS2}{g_2} \right), \quad (2.22)$$

where g_1 and g_2 are the detector dependent scintillation and ionization yield. All variables in Equation 2.22 can be determined by measurement.

The correctness of the absolute value computed via Equation 2.22 is not the main goal of the correction process, since a deviation from the true value can be handled by an energy bias correction. Rather, the correction process is meant to counteract dependencies of signal strength on the detector condition. In simple terms, an event of energy E_R should produce the same, bar finite resolution, $cS1$ and $cS2$, no matter where in the detector or to what point in time it was measured. Temporal and spatial non-uniformities in the detector response are corrected via the correction functions $\epsilon = \epsilon(t, \vec{r})$. The full correction process of XENONnT is also documented in [112]. Figure 2.16 shows a comparison between the distributions of the uncorrected and corrected signal areas for ^{83m}Kr calibration data. A signal area, or commonly simply *area*, is the area under the peaks waveform as the measure of the signals strength.

Since both signals are at the end measured by PMTs, their efficiencies share a PMT parameter ϵ_{PMT} , that models the probability of actually measuring the signal of a photon hit. It contains the quantum efficiency and charge collection efficiency of the PMTs. In XENONnT, the area is given in units of photoelectrons (PE), the number of electrons freed from the PMT photo-cathode on the initial absorption of the signal photon. VUV photons have the chance of freeing two PE from the photocathode of the PMT, called

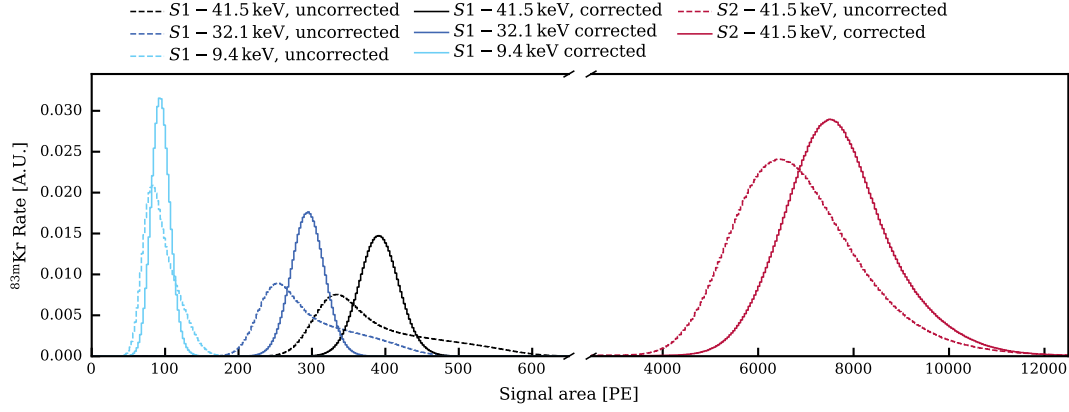


Figure 2.16: A comparison between uncorrected $S1$ and $S2$ area and the corrected $cS1$ and $cS2$ area for various energies, from [112].

the double-photo-electron probability and which is $\sim 20\%$ in XENONnT [113].

The $S1$ signal has to be corrected for two effects, next to the PMT efficiency: The detector has a certain light collection efficiency (LCE), the probability of a photon created at some position (x, y, z) in the TPC to be detected by one of the PMTs. This effect is dominated by the detector geometry and can be assumed to be uniform in time and highly non-uniform in space. Furthermore, many LXe detectors have measured backgrounds of spurious single to few electron $S2$ s, that have as of yet no conclusive origin [114, 115]. One explanation is photo ionization (PI) of impurities by the VUV photons. This explanation is supported by the fact that the $S1$ signal strength shows anti-correlation to the rate of these spurious small $S2$ s. For the correction, this means that the $S1$ signal needs an additional correction that handles the absorption of $S1$ photons by impurities in the LXe, that depends on the cleanliness of the detector, and therefore on the time of measurement. Combining all of these effects, the $S1$ correction can be expressed as

$$\epsilon_{S1}(t, x, y, z) = \epsilon_{PMT} \cdot \epsilon_{LCE}(x, y, z) \cdot \epsilon_{PI}(t). \quad (2.23)$$

The LCE (PI) correction ϵ_{LCE} (ϵ_{PI}) is assumed to be temporally (spatially) uniform.

The correction of the $S2$ signal is slightly more complicated. During the drift of the electrons from the initial interaction site to the liquid-gas interface, they can attach to electronegative impurities dissolved in the LXe. This effect can be characterized by the so-called electron lifetime τ_{life} , the average time that an electron survives inside the TPC without getting caught by an impurity. The probability of an electron surviving to the top of the TPC can then be written as $\epsilon_{\text{life}} = \exp(-t_{\text{drift}}/\tau_{\text{life}})$, where t_{drift} is the drift time of the electron. As this is dependent on the concentration of electronegative impurities in the detector, this correction is time dependent. As a side note, the electron lifetime seems to be uncorrelated to the PI strength, indicating that the electronegative impurities are in general distinct from the impurities causing PI.

After reaching the liquid-gas interface the electrons have to be extracted into the gas phase by the extraction field between gate and anode. This extraction has a certain field dependent extraction efficiency (EE) ϵ_{EE} , and for XENONnT is $\sim 50\%$ [116].

The EE is highly sensitive to minor changes in the extraction field, in such a way that it changes slightly after a reramping of the anode voltage, making this correction time dependent.

The extracted electrons are accelerated in the gas phase, emitting S2 photons via electroluminescence. The number of produced photons per electron is called the secondary scintillation gain, or the single electron gain (SEG), f_{SEG} . Parallel to the S1, there is also an LCE for the S2 photons. These two effects, SEG and S2 LCE, can not be distinguished without special dedicated calibrations. If the light received per electron changes by some percentage, it could be caused by a change in the SEG, or the LCE. For this reason, the spatial non-uniformities in SEG and S2 LCE are combined in the so-called S2 (x, y) map $M_{\text{S2xy}}(x, y)$.

The SEG depends, like the EE, on the extraction field. Temporal variations in the SEG and EE are combined in the efficiency ϵ_{SEGEE} . The ϵ_{SEGEE} is slightly spatially non-uniform, in part caused by the perpendicular wires across the gate and anode. The spatial dependence of ϵ_{SEGEE} is handled by separating the detector plain into two different partitions with individual correction factors each.

Finally, the S2 correction factor can be written as

$$\epsilon_{\text{S2}}(t, x, y, t_{\text{drift}}) = \epsilon_{\text{PMT}} \cdot M_{\text{S2xy}}(x, y) \cdot \epsilon_{\text{SEGEE}}(x, y, t) \cdot \exp(-t_{\text{drift}}/\tau_{\text{elife}}). \quad (2.24)$$

One important parameter in XENONnT analysis is the area fraction top (AFT), the fraction of photons measured by the top PMT array. For S2 signals, this AFT is approximately $\sim 70\%$. However, the remaining 30 % of photons have to transverse the LXe and are thereby subject to absorption by PI impurities. This time dependent reduction in effective S2 LCE is also included in the ϵ_{SEGEE} correction factor by design.

The values of g_1 and g_2 from Equation 2.22 can be determined by measuring the $cS1$ and $cS2$ areas of mono-energetic events and comparing the charge signal per energy to the light signal per energy, as is shown in Figure 2.17 [112]. The g_1 and g_2 values are extracted as slope and offset of a linear fit of the relative charge and light signals.

2.3.5 Data Organization and Jargon in XENONnT

Registered signals are classified as either S1s or S2s, or ignored, if they match neither classification. In XENONnT, S1 and S2 signals are referred to as *peaks*, while the pairing of multiple peaks, but at least one S1 and one S2, is called an *event*, that if reconstructed correctly, corresponds to a physical energy deposit in the TPC. Common peak or event parameters including simplified explanations are shown in Table 2.2. For a parameter like the *area* that both S1 and S2 peaks posses, it is common to specify the *S2 area* of an event.

The measurements are divided into *runs*, periods of data taking with a duration in general between 30 min and 3 h. Ambient parameters, like the LXe purity, the temperature or the electrical field configurations are assumed to be constant within one run. Each run has associated tags, that provide information about what happened during

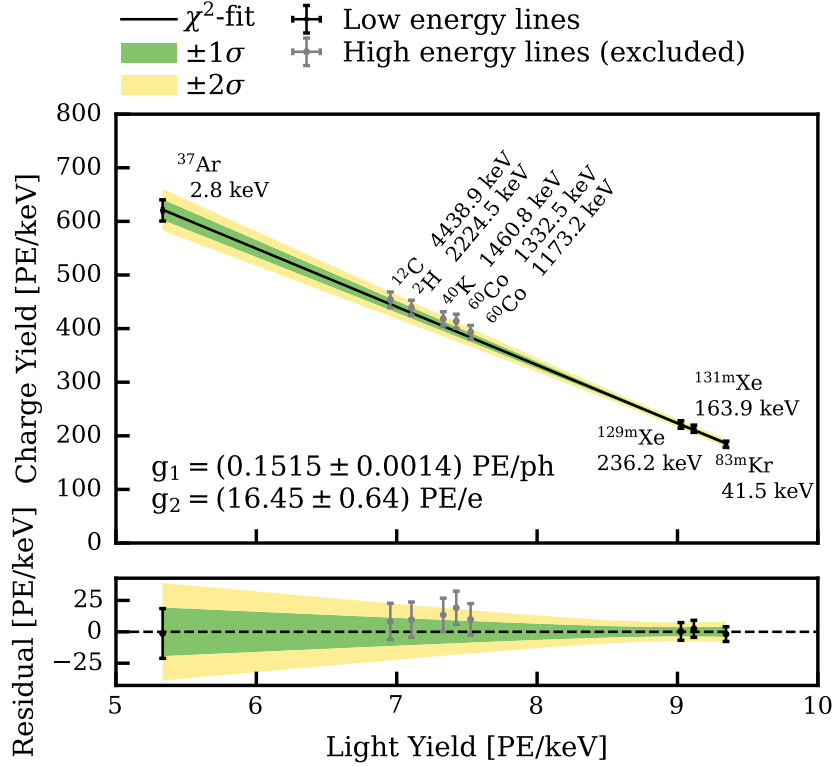


Figure 2.17: The anti-correlation of light and charge signals for mono-energetic lines, from [112].

the run, for example a ^{220}Rn -run is a run where a ^{220}Rn source was introduced into the detector for calibration purposes. A *background run* is a run without any purposely introduced additions, meaning it measures the background only state of XENONnT. These background runs are the basis of any science search. More information about the calibration can be found for example in [112].

Several runs are grouped into so-called *science runs* (SRs). To date there have been three science runs in XENONnT: science run 0 (SR0) from 2021-05-01 to 2021-12-10, science run 1 (SR1) from 2022-05-19 to 2023-08-08 and science run 2 (SR2) from 2023-10-12 to 2025-03-25.

Data Selection

Data selection in XENONnT is done via so-called *cuts*. A cut in this context is an algorithm with a boolean output that depends on event parameters. An easy example is the Fiducial Volume (FV) cut, that rejects events based on their position, to fully utilize the self-shielding properties of LXe that were detailed in subsection 2.3.1. The FV-cut rejects all events deemed too close to the detector walls, and thereby greatly reduces the background contribution from detector material radiation. A cut has in general one of two purposes: Either it is meant to ensure data quality and reject low quality, read probable to be background, events, or a cut is meant to select a very specific population of events, for example events caused by calibration sources. An in-depth description of the XENONnT cuts can be found in [112].

Table 2.2: A selection of peak and event parameters commonly used in XENONnT analysis.

Parameter	Unit	Explanation
Time	ns	Start time of the peak, or of the main S1 for events
Type	-	The type of a peak, S1, S2 or other
Area	PE	Integrated signal strength
Width ($k\%$)	ns	Temporal width of the central $k\%$ of the area of a peak. Most often 50 %
Area fraction top (AFT)	-	Fraction of area measured by the top PMTs
Rise time	ns	The time between 10 % and 50 % of the total area of a peak
Position	cm	The reconstructed position of an S2 peak
Main S1 (S2)	-	The S1 (S2) peak with the greatest area within an event
Alt S1 (S2)	-	The S1 (S2) peak with the second greatest area within an event
Drift time	ns	The time between main S1 and main S2 of an event

In general, there are four types of cuts: Detector veto cuts are cuts that reject data recorded during a trigger of the MV or NV veto detector or during time where the DAQ was unable to properly record data. Accidental coincidence (AC) cuts target the incorrect pairing of uncorrelated S1s and S2s. Such AC events possess for example an anomalous drift time or reconstructed energy. The other categories are S1 and S2 quality cuts that target unphysical events, for examples ones where some of the signal was lost or where a PMT is likely to have malfunctioned. A selection of cuts used in XENONnT is listed in Table 2.3.

2.3.6 The Software Landscape of XENONnT

Strax, straxen and cutax

The XENONnT experiment uses the STRAXEN (Streaming analysis for XENONnT)² [117] software package for data processing, which is built as an extension of the more general STRAX (Streaming analysis for xenon experiments)³ [118] software. Both are open-source packages written in the python language, that utilize common scientific python packages like numpy⁴, scipy⁵, pandas⁶ and matplotlib⁷. In addition, many of the data processing steps in STRAXEN are accelerated using the numba⁸ package for just-in-time (JIT) compilation. The focus on *streaming* analysis comes from the fact that XENONnT uses a triggerless DAQ [96], which provides the recorded data as a continuous stream, as opposed to a stack of easily countable individual events.

²<https://github.com/XENONnT/straxen>

³<https://github.com/AxFoundation/strax>

⁴<https://numpy.org/>

⁵<https://scipy.org/>

⁶<https://pandas.pydata.org/>

⁷<https://matplotlib.org/>

⁸<https://numba.pydata.org/>

Table 2.3: A selection of cuts used in XENONnT, vaguely grouped into AC cuts (top), S1 quality cuts (center) and S2 quality cuts (bottom).

Name	Explanation
AC BDT	Rejection of mispaired S1s and S2s based on a boosted decision tree (BDT)
Shadow	Rejection of mispaired S1s and S2s based on correlation to previous large signal
Ambience	Rejection of mispaired S1s and S2s based on adjacent signals
S2 width	Rejection of events where the width of the S2 does not match expectation from diffusion and drift time
S1 width	Rejection of S1s with larger than expected width
S1 max PMT	Rejection of S1s where one PMT contributes an unusually large amount to the area
S1 AFT	Rejection of S1s where the area distribution between top and bottom does not follow what is expected from the detector geometry
S1 single scatter	Rejection of multi-scatter events identified by matching the main S1 to the alt S2
S2 AFT	See S1 AFT
S2 pattern	Rejection of S2s where the hitpattern does not match expectations given the position
S2 single scatter	Rejection of multi-scatter events identified via the alt S2

STRAXEN divides the full chain of data processing into smaller, and therefore simpler to handle, individual steps using a modular approach with so-called plugins. The tasks of individual plugins are for example classifying peaks into S1s or S2s, pairing S1s and S2s to events or correcting an event parameter for known detector effects. For example, the process of reconstructing the position of an S2 signal described in [chapter 3](#) is implemented as such a plugin.

Each plugin can have different auxiliary parameters, for example time dependent input variables. Auxiliary parameters are handled by the so-called *context*, that stores uniquely identifiable versions of each additional input. A context also handles data storing, caching, and loading, in a way that the average user does not have to worry about them. All of this makes it extremely easy to share analysis tasks among collaborators, as using the same context automatically ensures that everyone is working on the exact same data and analysis algorithms.

The general data processing of XENONnT utilizes a third package called CUTAX, that is not openly accessible. CUTAX handles cuts and selections in data, described in [subsubsection 2.3.5](#). CUTAX inherits the plugin structure from STRAX and STRAXEN. That means that each individual cut is also independent and modular, and that the versioning information as well as auxiliary parameters are stored in the STRAXEN context, and again easily shareable among collaborating analysts.

Simulations with fuse and GEANT4

Monte-Carlo (MC) simulations play an important role in the analysis of data in XENONnT, for example when the modeling of an effect requires knowledge about some underlying ground truth, such as the training of the position reconstruction models described in [chapter 3](#). Two main frameworks are employed to simulate data in XENONnT, GEANT4 (geometry and tracking) [119] and *fuse* (framework for unified simulation of events)⁹ [120].

GEANT4 is a C++-based software package, that focuses on realistic propagation of particles through user-defined media, including processes from atomic physics and particle physics such as scattering and atomic excitation. Because the modeling of physics in GEANT4 is so detailed, a simulation can take hours, and even up to months, of CPU time. A fully detailed realization of the XENONnT geometry is implemented in GEANT4. The input of a GEANT4 simulation consists in general of a list of primary particles, their initial positions, and their energy or momentum. The software offers the possibility of accessing every intermediate physical process from the creation of the initial particle until every subsequently produced particle has either come to rest or left the simulated world. However, most commonly the result of a GEANT4 simulation is a list of ensuing energy deposits and their respective positions. A common example of using GEANT4 is the simulation of radioactive decay in one of the detector components, capturing the complex spatial and energetic distributions of the resulting events.

Fuse is a simulation framework implemented in the *python* programming language that inherits the plugin structure from the STRAX framework. Each step of a *fuse* simulation is implemented as its own plugin, and the *fuse* plugins are divided into three general categories: the microphysics processes, the detector physics processes and the PMT and DAQ plugins. The microphysics plugins handle the quanta generation, sketched in [Figure 2.13](#), translating an energy deposition at a given position into a number of generated electrons and photons. The detector physics plugins, which are themselves split into S1 and S2 plugins, take these electrons and photons and simulate the full signal formation process that is described in [subsection 2.3.3](#), including electron drift, secondary scintillation and photon propagation. The results of the detector physics simulation are PMT hitpatterns, the distribution of light signals received by the PMT arrays. The PMT and DAQ plugins simulate the response of the PMTs to those light signals and the read-out performed by the DAQ system, culminating in simulated electrical signals that resemble the data that is recorded by the experiment. The format of the simulations is such, that the results can be immediately processed by the STRAXEN software, which ensures that also the signal reconstruction process is the same for simulated and recorded data.

The plugin structure of *fuse* allows for a modular approach to the data simulation, where the simulation process can be initiated at any given point in the plugin hierarchy. For example, if the underlying energy depositions do not matter for a given study, but only the photons and electrons, the microphysics plugins can be skipped completely. The most basics input to a *fuse* simulation is a list of deposited energies

⁹<https://github.com/XENONnT/fuse>

and their corresponding positions, such as is provided for example by a preceding GEANT4 simulation. The combination of a GEANT4 simulation, followed by a *fuse* simulation is referred to as a *full-chain* simulation. A full-chain simulation models the energies, positions and resulting recorded signals as detailed as possible given the best knowledge of the detector and the underlying physics. In general, a full-chain simulation is only necessary when the spatial distribution of the studied events is complex, as is the case for example for detector material originating radiation, or if the energy spectrum is complex, for example in the β - γ cascade of ^{214}Pb decays.

Alea

The *alea*¹⁰ [121] software is an open-source python package for statistical inference, including implementations of likelihood-tests described in [122], the power constrained limit (PCL) explained in [123] and following statistical recommendations from [33]. The base statistical measure is the profile likelihood ratio (PLR)

$$q(x|s) = -2 \ln \frac{\mathcal{L}(x|s, \theta_{\max}(s))}{\mathcal{L}(x|\hat{s}, \hat{\theta})}. \quad (2.25)$$

Here, the likelihood $\mathcal{L}(x|s, \theta)$ is the likelihood function of observing the data x , given a statistical model that is characterized by the nuisance parameters θ and signal-parameters s . The parameters $\hat{s}, \hat{\theta}$ are those that maximize $\mathcal{L}$, defining the *best-fit* model. The nuisance parameters $\theta_{\max}(s)$ are those that maximize $\mathcal{L}$ for a fixed value of s . Assuming large statistical samples, the PLR, q , is distributed as a χ^2 distribution with one degree of freedom, following Wilks theorem [124], which simplifies for example the construction of confidence intervals.

Alea utilizes human-readable configuration files that contain all of the potential sources that contribute to $\mathcal{L}$, such as background and signal models, as well as their parameters s and θ . In addition, detector specific effects, for example the exposure, signal efficiency or energy reconstruction bias can be included in the construction of the statistical model. The spectra of complex backgrounds can be supplied as probability density functions (PDFs) via template files, that contain a normalized rate for the parameter space. For the optimization of the PLR, parameters can be left free, or can be constrained by some auxiliary measurement, information that is also included in the configuration file. In addition, *alea* supports both binned and unbinned PLR analyses.

Furthermore, *alea* offers the possibility of generating toy MC data from the statistical model defined in the configuration file. Such toy MC data is used for example to compute the sensitivity of a given configuration towards some scientific parameter, or to evaluate the goodness-of-fit between a model and data. A more in-depth description of the *alea* software can be found in [125].

¹⁰<https://github.com/XENONnT/alea>

Probabilistic Position Reconstruction using Conditional Normalizing Flows

In [subsection 2.3.4](#), multiple effects were introduced that can alter the size of the S1 and S2 signals, that depend on the event location inside the detector. As they display a position dependence, the correct accounting of these effects is intrinsically intertwined with the accuracy of the position reconstruction. In addition, accurate knowledge of the event position is absolutely necessary to exploit the self-shielding properties of LXe described in [subsection 2.3.1](#), which leads to the calculation of the fiducial volume, as done in [section 5.3](#). This chapter describes the position reconstruction process of XENONnT, starting with the base principles and leading to the introduction of a novel Machine Learning (ML) method that enables probabilistic predictions of event positions.

The difficulties encountered when reconstructing the spatial coordinates of an energy deposit inside the sensitive LXe volume are outlined in [section 3.1](#). A quick introduction of the working principles of NNs is provided in [section 3.2](#), after which [section 3.3](#) describes how to extend such NNs to predict probabilistic outputs. Thereafter, in [section 3.4](#), the novel Conditional Normalizing Flow (CNF) position reconstruction model is introduced. Finally, [section 3.5](#) focuses on data driven validation of the CNF model, and [section 3.6](#) outlines the applications of the probabilistic output.

3.1 Position Reconstruction in XENONnT

In a LXe TPC, the electrons that carry the charge signal and emit the S2 photons drift from the initial interaction site to the liquid-gas interface at the top of the detector. Due to minor field inhomogeneities or transversal diffusion of the electrons, the position at the liquid interface is not necessarily in a perfect vertical line above the initial energy deposit [[112](#)]. Position reconstruction in this context refers to the inference of the horizontal (x, y) position of the drifted electrons at the liquid interface of the TPC. This is notably not the (x, y, z) -vertex of the original event, and therefore in this chapter the *position* will always refer to the (x, y) coordinates after the electron drift.

In XENONnT, the position is derived from the S2 top *hitpattern*, which is the distribution of light signals in the top PMT array caused by an S2 signal [112]. The position can be inferred via a profile likelihood ratio (PLR) minimization, as done for example in [126]. Given N total top PMTs, an observed S2 top hitpattern $\vec{h} \in \mathbb{R}^N$ and an expected S2 top hitpattern $\vec{g}(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^N$, the likelihood function can be decomposed into the product of per PMT probabilities, p_i :

$$\mathcal{L}(\vec{h}|\vec{g}(x, y)) = \prod_{i=1}^N p_i(h_i|g_i(x, y)), \quad (3.1)$$

where h_i is the observed area of PMT i , and $g_i(x, y)$ the expected area of PMT i , given an event at position (x, y) . The probability function, p_i , for a PMT to observe h_i photoelectrons given g_i expected photoelectrons can be approximated as a Poisson distribution

$$p_i(h_i|g_i(x, y)) = \frac{g_i(x, y)^{h_i} \exp(-g_i(x, y))}{h_i!}. \quad (3.2)$$

In the PLR test, the metric to minimize, q , is the negative logarithm of the ratio of two likelihoods, with the denominator being the global best fit. The equivalent of a best-fit estimator in the position reconstruction task is $\mathcal{L}(\vec{h}|\vec{h})$, the likelihood of observing the hitpattern $\vec{h}$, given $\vec{h}$ is also expected. This leads to the test statistic

$$\begin{aligned} q &= -2(\ln \mathcal{L}(\vec{h}|\vec{g}(x, y)) - \ln \mathcal{L}(\vec{h}|\vec{h})) \\ &= 2 \sum_{i=1}^N \left(g_i(x, y) - h_i + h_i \ln \frac{h_i}{g_i(x, y)} \right), \end{aligned} \quad (3.3)$$

which has to be minimized to locate the optimum position (x, y) .

One major challenge for this approach is the derivation of the expected hitpattern $\vec{g}$. The expectation per PMT can be described by $g_i(x, y) = h \cdot f_i(x, y)$, with the total number of observed photons $h = |\vec{h}|$ and the so-called *lightcurve* $f_i(x, y)$. The lightcurve describes the proportion of all detected photons that PMT i observes, given an event at the location (x, y) . These lightcurves are commonly derived from extensive photon tracking simulations, and can be, given certain assumptions, refined using recorded data [126].

An observed hitpattern $\vec{h}$ and a lightcurve $\vec{f}(x, y)$ for a given example position $(x, y) = (-15 \text{ cm}, -15 \text{ cm})$ are shown in Figure 3.1. A PLR position reconstruction algorithm performs a form of gradient descent along (x, y) , thereby varying $\vec{f}(x, y)$, to derive the minimum of Equation 3.3.

To compare different position reconstruction algorithms and decide among them, a metric to quantify their prediction capabilities is required for comparison. One way to define the precision of the reconstruction of the position is via the distribution of the reconstruction error, the spatial distance between the true positions and the predicted positions. For a bivariate normal distribution, an ellipse defined by the standard deviation contains $\sim 39.3\%$ of the cumulative distribution function (CDF)¹. The resolution

¹In the same way that an interval $[-\sigma, \sigma]$ contains $\sim 68\%$ of the CDF of the one dimensional normal distribution

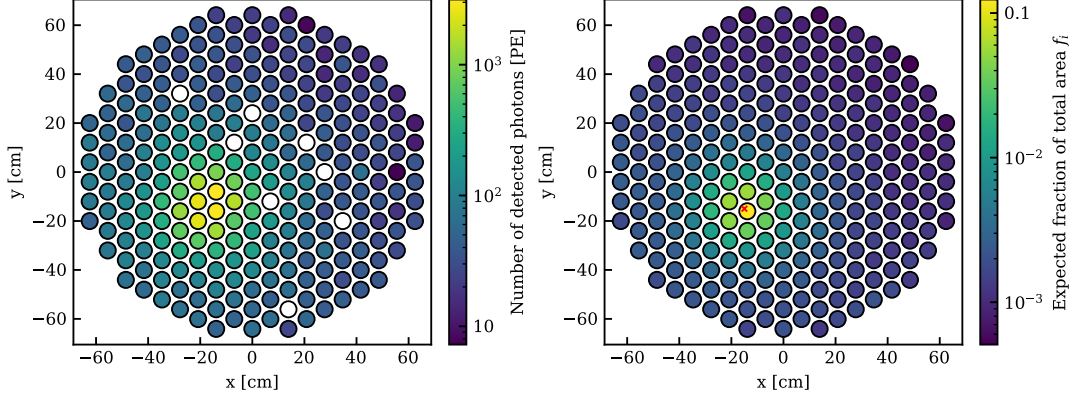


Figure 3.1: Left: An example of an S2 top hitpattern, $\vec{h}$. The hitpattern is the distribution of recorded photons per PMT. PMTs without any recorded photons (white) were turned off during the measurement of this hitpattern. Right: An example of a lightcurve $\vec{f}(x, y)$ derived for a position $(x, y) = (-15 \text{ cm}, -15 \text{ cm})$, highlighted in red. The lightcurve equals the expected fraction of the total measured area per PMT.

ρ is defined as the 0.393th quantile of the spatial distance between true position and reconstructed position:

$$\rho := q_{\sigma_{2D}} \left(\sqrt{(x_{\text{true}} - x_{\text{pred}})^2 + (y_{\text{true}} - y_{\text{pred}})^2} \right). \quad (3.4)$$

Here, $(x_{\text{true}}, y_{\text{true}})$ is the true position of the events and $(x_{\text{pred}}, y_{\text{pred}})$ is the position predicted by the position reconstruction algorithm. The function $q_{\sigma_{2D}}$ is the quantile function, that returns the quantile of a list of values corresponding to the standard deviation of a two dimensional normal distribution. A lower resolution means less prediction error and in general a better reconstruction model.

The Newman-Pearson lemma [127] states that the most powerful test to compare statistical hypotheses is the minimization of the PLR, given in Equation 3.3. However, this is only true if the likelihood in Equation 3.1 is accurate, which requires perfect knowledge of the behavior of photons inside the detector. But due to being derived primarily from simulated data, the PMT probabilities p_i and the lightcurves $f_i(x, y)$ are at best approximations of the actual underlying physical nature. In addition, finding the minimum of Equation 3.3 is often non-trivial and time consuming [1]. For this reason, the XENONnT experiment uses various ML methods to derive the position from the S2 top hitpattern [112].

To train ML models to predict the position it is necessary to have a training data set consisting of hitpatterns produced by S2s with a known true position. This can only be done by using simulated data. Whenever an ML model is trained in the following sections, this is done on data simulated using the *fuse* package introduced in subsection 2.3.6. The simulated S2 signals are uniformly distributed in the horizontal plane of the detector, and uniform in $\log_{10}(A_{\text{top}})$, where A_{top} is the top area, the sum of the PMT signals measured by the top PMT array. A more detailed description of the S2 signal simulation can be found in Appendix A.

3.2 A short Introduction to Feed-Forward Neural Networks

Merriam Webster defines Machine Learning (ML) as "a computational method [...] that enables a computer to learn to perform tasks by analyzing a large dataset without being explicitly programmed" [128]. In the most general sense, this definition encompasses a wide variety of algorithms, from simple χ^2 -minimization to the highly complex large language models of recent years [129]. Most commonly associated with the expression ML are the so-called feed-forward Neural Networks (NNs) [130]. An NN is composed of different stages, known as *layers*: an input layer that receives some input $x \in \mathbb{R}^n$, an output layer that yields a prediction $y \in \mathbb{R}^m$, and a number of hidden layers that link these two. Each layer consists of *neurons*, that represent mathematical operations to be performed on the input, and *connections* between those neurons that define the input of the successive layer, hence the name neural network. The amount and nature of layers, neurons and connections is called the *architecture* of an NN.

One of the simplest and most common NN architectures is the Multi-Layer Perceptron (MLP), where each neuron of one layer is connected to every neuron of the following layer, as visualized in Figure 3.2. Each neuron represents a number and each connection is characterized by two numbers, called weight w and bias b , which are multiplied and added to the preceding number, respectively. Let $\vec{x}_i \in \mathbb{R}^{d_i}$ be the numerical values of the neurons at layer i . In the most simple variation, the values of the neurons of layer $i + 1$, $\vec{x}_{i+1} \in \mathbb{R}^{d_{i+1}}$, are computed via

$$\vec{x}_{i+1} = \mathbf{w}_i \cdot \vec{x}_i + \vec{b}_i \quad (3.5)$$

with weights $\mathbf{w}_i \in \mathbb{R}^{d_i \times d_{i+1}}$ and biases $\vec{b}_i \in \mathbb{R}^{d_{i+1}}$. In general, an MLP with only the linear matrix multiplication from Equation 3.5 functions as a linear regressor. To introduce non-linearity into the relations between the layers, and therefore between input and output, each layer is commonly equipped with an activation function $f : \mathbb{R} \rightarrow \mathbb{R}$. Including the activation function, Equation 3.5 becomes

$$\vec{x}_{i+1} = f(\mathbf{w}_i \cdot \vec{x}_i + \vec{b}_i), \quad (3.6)$$

where f is applied element-wise to each entry of its input vector.

Common activation functions include the linear function $f_{\text{linear}}(x) = x$, the hyperbolic tangent function $f_{\text{tanh}}(x) = \tanh(x)$, the sigmoid function $f_{\text{sigmoid}}(x) = (1 + e^x)^{-1}$, the rectified linear unit (ReLU) $f_{\text{ReLU}}(x) = \max(0, x)$ or the exponential linear unit (ELU) $f_{\text{ELU}}(x) = x$ for $x > 0$ and $f_{\text{ELU}}(x) = \alpha(e^x - 1)$ for $x \leq 0$ [131–133]. The primary goal of an activation function is to introduce nonlinearities into the network. The correct choice of activation function depends on many things, like the desired range of parameters or the sensitivity of the gradient to outliers in the data. A more in-depth review on different activation functions can be found for example in [134].

Finally, for an MLP with k layers, the prediction $\vec{y}$ can be derived from the input $\vec{x}$ via

$$\vec{y}_{\text{pred}} = f_k(\mathbf{w}_k \cdot \dots f_2(\mathbf{w}_2 \cdot f_1(\mathbf{w}_1 \cdot \vec{x} + \vec{b}_1) + \vec{b}_2) \dots + \vec{b}_k). \quad (3.7)$$

The mathematical relations between individual neurons are also sketched out in Figure 3.2.

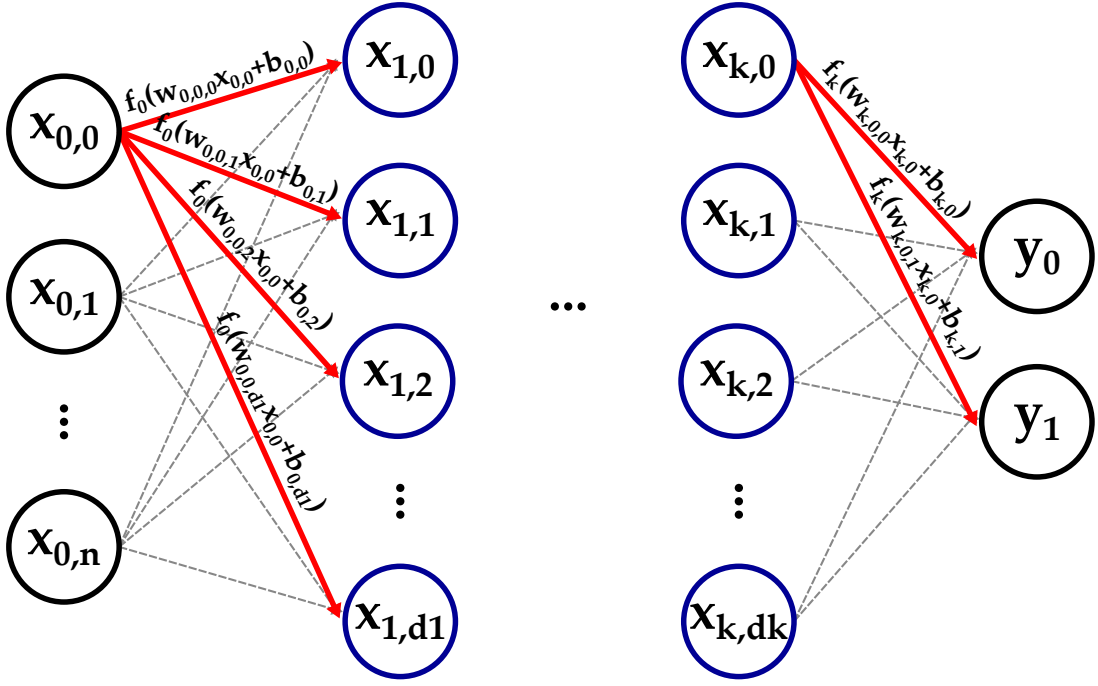


Figure 3.2: A sketch of the working principle of an MLP. The input layer x and output layer y are shown in black. Between those there are k hidden layers (blue) of dimensionality d_i . Each neuron of layer i has a connection to layer $i + t$ (gray and red). The connections multiply the values of $\vec{x}_i$ with a weight matrix $\mathbf{w}_i$ and add the bias term $\vec{b}_i$. This term is then enclosed by an activation function f_i .

During the training procedure, an MLP is provided with a set of n_{train} training samples, $X = \{\vec{x}\}$, and their corresponding labels, $Y = \{\vec{y}_{\text{truth}}\}$. A label represents the true values that the model attempts to approximate with its prediction, $\vec{y}_{\text{pred}}$. The so-called *loss function* $\mathcal{L}(y_{\text{truth}}, y_{\text{pred}})$ is the mathematical relation that is optimized during training, by varying the model parameters $\{\mathbf{w}, \vec{b}\}$. Common loss functions include the mean square error (MSE), or the categorical crossentropy (CE). The choice of loss function has to be tailored specifically to the problem at hand, for example, the MSE is suited for regression tasks, while the CE is suited for classification. The loss function is minimized by adjusting the weights and biases of the model utilizing end-to-end gradient descent. Different realizations of the gradient descent are called optimizers. Popular optimizers are for example the Adam [135] or AdamW [136] algorithms.

A *step* is the basic building block of the training procedure, during which the model iterates over one *batch* of n_{batch} training samples to adjust the model parameters. After $\lceil n_{\text{train}}/n_{\text{batch}} \rceil$ steps, the model has seen all n_{train} available training samples, which is called one *epoch*. A model is considered fully trained when the loss function converges to a constant value. A model that is not trained until loss convergence is called *undertrained*. However, the opposite is also possible: A model with sufficient parameters can learn the training data by heart, instead of learning the underlying data topology that the training data is meant to represent. An ML model that can predict input from its training set very well, but can not generalize to data outside the training set is called *overtrained*. To detect and be able to counteract overtraining, it is common to analyze the model performance on a *validation set* after each epoch. The

validation set contains data that is not used to train the model, but that is selected from the same general population as the training set. If the loss function averaged over the validation set, or *validation loss*, no longer decreases, the model is no longer learning generalized information about the underlying data. In this case, if the training loss is continuing to decrease, the model is overtraining.

Generally, the risk of overtraining increases with the number of total parameters $n_{\text{par}} = |\{\mathbf{w}, \vec{b}\}|$ that are available to the model. Although recommendations for the relation between model size n_{par} and size of the available training set n_{train} exist, these are rules of thumb as opposed to strict laws, because of the no-free-lunch theorem. The no-free-lunch theorem states that there is no general step by step recipe one can follow to attain the perfect optimization algorithm for a generic problem, but rather every optimization has to be adjusted specifically for the underlying problem [137]. Overall, it can be useful to keep the following in mind [138, 139]:

- A problem with $n_{\text{train}} < n_{\text{par}}$ is *underdetermined* because there exists a guaranteed degeneracy among the model parameters.
- To minimize the risk of overtraining, a problem should be ten to fifty times overdetermined, meaning $n_{\text{par}} \leq n_{\text{train}}/50$.
- For classification problems, there should be at least $\mathcal{O}(10)$ samples per class available, meaning the training set should also be diverse enough to represent the full data population and not only a fraction

Aside from increasing the number of training samples, common methods to counteract overfitting include *weight dropout* and *regularization*. Weight dropout refers to the practice of setting randomly selected weights of one layer to zero during training. The desired result of this method is to force the information represented by that weight to be encoded in other weights, effectively decreasing the number of independent parameters of the model. Regularization is the addition of a penalty term to the loss function, that scales with the sum of the model weights. The penalty term favors smaller weights in the gradient descent, making the model overall less expressive and less likely to learn the training data by heart.

3.3 Probabilistic Extension of Neural Networks

Building on the theoretical background of ML presented in [section 3.2](#), the focus of the following sections shifts to the details of the ML-based implementation of the position reconstruction process in XENONnT and the improvements developed during this thesis. For the first two measurement campaigns, SR0 and SR1, the position reconstruction in XENONnT utilized a Graph Neural Network [140], a Convolutional Neural Network (CNN) and an MLP [112], all of which are feed-forward NNs. The following [section 3.3](#) presents two different ways of transforming the point-like output of such a NN into a probability density function (PDF) of the position likelihood, spanning the full horizontal plane. Both of these methods are applicable to any arbitrary feed-forward NN, and are demonstrated on the example of an extended MLP.

The focus of this work is primarily to provide additional information in the form of the position PDF, as opposed to an improvement of the position reconstruction accuracy.

3.3.1 Obtaining Position Probability Distributions via One Hot encoded Classification

There is a general distinction between classification and regression tasks in ML. In a regression task, the NN is trained to predict a numerical output corresponding to a given input, analogous to a least-squares fit. On the other hand, a classification model maps the input variables to one of several distinct classes. For classification, it is common to constrain the model output such, that the sum of the output across all possible classes equals one. This way, the model prediction can be interpreted as a pseudo-probability of the input belonging to a certain class.

The position reconstruction is a regression task, converting the hitpattern $\vec{h}$ to a predicted position (x, y) , where the output of the model can not trivially be transformed into a pseudo-probability. The idea behind the One Hot (OH) classifier is to rephrase the position reconstruction task from a regression problem into a classification problem, and then using standard ML classifier methods to attain pseudo-probabilities. This is done by separating the output space, in this case the (x, y) -plane, into individual bins of a certain size and treating each bin as a distinct class.

The OH model is trained as a typical classifier on the categorical crossentropy (CE) of predictions and labels, $\mathcal{L}_{\text{CE}}(y_{\text{truth}}, y_{\text{pred}}) = -y_{\text{truth}} \log y_{\text{pred}} - (1 - y_{\text{truth}}) \log(1 - y_{\text{pred}})$. The labels are adjusted according to a One Hot encoding [141], hence the model name: Each position $(x, y)_i$ of the training data is transformed into an array $Y_{\text{OH}} \in \mathbb{R}^n$ with

$$Y_{\text{OH}} = \begin{cases} 1 & \text{if } (x, y)_i \text{ in bin } j \\ 0 & \text{else} \end{cases}, \quad (3.8)$$

where n is the number of bins dividing the horizontal plane. The final layer of a OH model is a *softmax* layer, the function of which is to ensure that all output values are between 0 and 1, and that the sum of the output equals 1.

The predicted position is taken as the center of that bin with the highest output value. An example of the reconstruction is illustrated in Figure 3.3, using two demonstration models, trained on $n_{\text{train}} = 10000$ samples for 20 epochs each, with bin widths of 10 cm and 5 cm. Both models have two hidden layers of 60 and 20 nodes, while the size of the output layer of each model is the number of bins in the horizontal plane. Figure 3.3 demonstrates the prediction process of an OH model, with the left panel showing the hitpattern of the top PMT array, that serves as input for the position reconstruction models, and the center and right plots showing the predicted probability per bin of the two different OH models. Finally, the black and red crosses in each show the true and predicted positions respectively.

The finite binning of a OH model leads to an obvious lower limit on the position resolution ρ : Assuming the model always predicts the correct bin, and that the true positions are distributed uniform in the (square) bin, the minimum achievable resolution of a OH model is $\rho_{\text{min}} \approx 0.354 \cdot L$, where L is the side length of one bin.

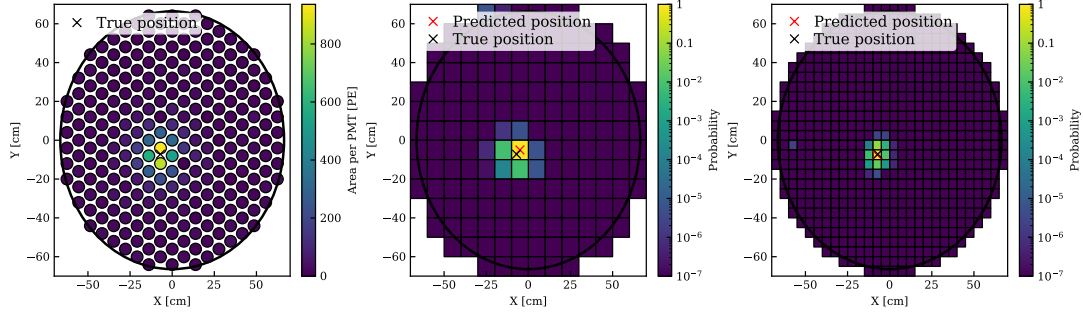


Figure 3.3: The position of an example hitpattern (left) reconstructed by a 10 cm (center) and 5 cm (right) OH model. The bin with highest predicted probability is the bin containing the true position for both models. The predicted positions are shown in red, the true position is shown in black.

The minimal resolution $\rho_{\min}$ was derived via a toy MC study, the result of which is visualized in Figure 3.4: Inside a square of side length $L = 1$ cm, 1×10^8 positions were generated at random, and their spatial distance to the center of the square was calculated. The distribution of these distances is shown in Figure 3.4, with the addition of the mode, the median and the 39.3 % percentile in red dotted, dashed and solid, respectively. Since the resolution ρ is based on the 39.3 % percentile, as defined in Equation 3.4, the minimum resolution of an OH position reconstruction model is 0.354 times the bin width, which corresponds to 39.3 % of the distance distribution.

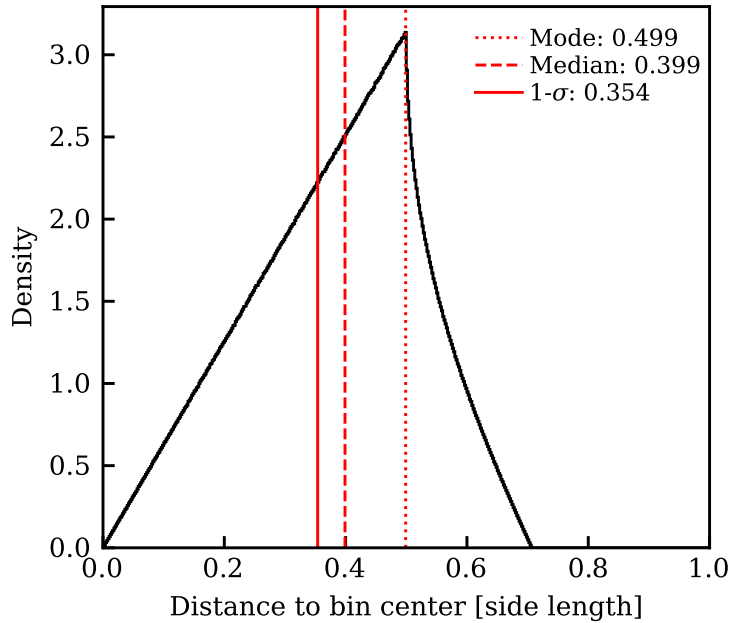


Figure 3.4: The distribution of the distance of randomly generated points inside a square, to the center of the square. The distance is converted to a fraction of the side length of the square. In addition, the mode (red dotted), the median (red dashed) and the 39.3-percentile (red solid) of the distribution are shown.

Figure 3.5 shows the resolution ρ as a function of the area of the S2 hitpattern in the top PMT array for the 10 cm model and the 5 cm model. A horizontal line at the theoretical minimal resolution for an OH model $\rho_{\min}$ is shown as a dashed line, to

Table 3.1: Parameter and timing benchmarks of two example OH position reconstruction models. The timing was derived on a non-optimized setup using an AMD Radeon 610M graphics card. The absolute timing values are therefore not representative, but the relative differences are.

Bin width	10 cm	5 cm
Number of parameters	19904	29228
Time to train for 20 epochs	14 s	23 s
Time to infer 1000 positions	70.3 ms	71.4 ms
Minimum resolution $\rho_{\min}$	3.54 cm	1.77 cm

highlight that the resolution of both models does indeed converges to that value for increasing S2 signal strength.

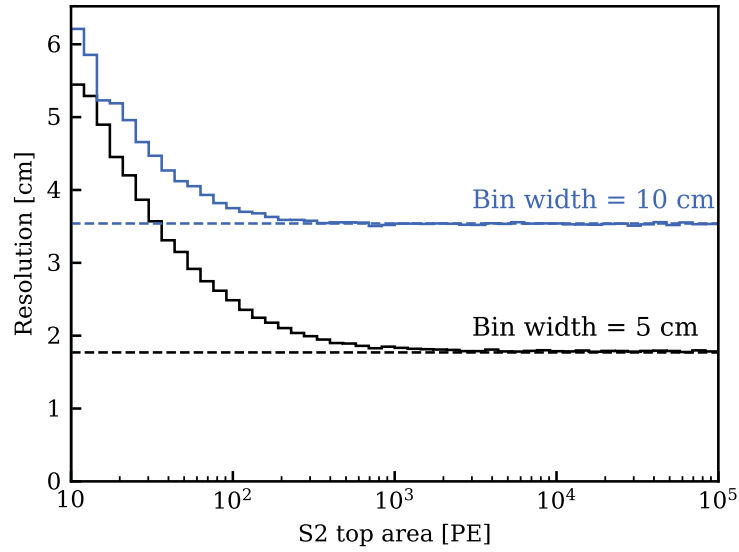


Figure 3.5: The resolution of two example OH models, using bin widths of 10 cm (blue) and 5 cm (black). The model resolution converges to the theoretical minimum of $\rho_{\min} = 3.54$ cm (10 cm model) and $\rho_{\min} = 1.77$ cm (5 cm model) for increasing S2 areas.

The number of parameters, minimal resolution, and timing information for the two demonstration models is shown in Table 3.1. It is trivial that a model with a smaller bin width L will perform better from a resolution perspective. The fact that the training time nearly doubles is not that important of an argument for either model, since the training procedure has to be performed only once to get a functioning model. The more important time measure is the inference time, in which the two models do not differ significantly. The main issue is the exploding number of parameters that scale linear with the number of bins and inverse squared with the bin width $n_{\text{par}} \propto 1/L^2$.

As shown in [112], the resolution of the position reconstruction models used in SR0 and SR1 reach a resolution of the order of $\mathcal{O}(1 \text{ mm})$, for highest S2 areas. To achieve a similar resolution, the required bin width of an OH model would be $L = \rho_{\min}/0.354 \approx 2.82 \text{ mm}$. Dividing the circle of the TPC into squares of side length 2.82 mm requires 177640 bins. Constructing an OH model with this number of output bins and two hidden layers of 60 and 20 nodes, which is the same number used in the two example

models above, leads to an architecture with $n_{\text{par}} \approx 3.7 \times 10^6$ parameters. A model that big runs great risk to overfit the data, is slow to train and is very heavy to load into memory for smaller machines as the XENONnT DAQ. While all of these are not insurmountable challenges, it is shown in the following sections that there are better and more simple ways to attain a probabilistic output for the position reconstruction.

3.3.2 Obtaining Position Probability Distributions from Parameterized Likelihood Functions

A different way of making the output of a NN probabilistic is by defining the output values $\vec{\theta}$ as the parameters of a pre-defined PDF $f(\vec{\theta}(\vec{h}))$. The basic example for position reconstruction would be to replace the point-prediction, $\vec{y}_{\text{pred}} = (x, y)_{\text{pred}}$, with the parameters of a bivariate Normal Distribution, $\vec{y}_{\text{pred}} = (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho_{xy})_{\text{pred}}$:

$$\mathcal{N}_2(x, y | \mu_x, \mu_y, \sigma_x, \sigma_y, \rho_{xy}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho_{xy}^2}} \quad (3.9)$$

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{\left[\frac{(x - \mu_x)^2}{\sigma_x^2} - 2\frac{\rho_{xy}(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2}\right]}{[1 - \rho_{xy}^2]}, \quad (3.10)$$

where μ_x and σ_x (μ_y and σ_y) are the mean position and standard deviation in x (y) respectively, and ρ_{xy} is the correlation between x and y .

The physical meaning of the parameters $\vec{\theta}$ is determined via the loss function, which means that to make the prediction resemble the chosen PDF, the NN is trained on the negative log-likelihood of that PDF. For a bivariate normal, the negative log-likelihood is

$$\mathcal{L}_{\mathcal{N}_2} = \frac{1}{2} \left(\log[\sigma_x^2\sigma_y^2(1 - \rho_{xy}^2)] + \frac{\left[\frac{(x - \mu_x)^2}{\sigma_x^2} - 2\frac{\rho_{xy}(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2}\right]}{(1 - \rho_{xy}^2)} \right). \quad (3.11)$$

The prediction $\vec{y}_{\text{pred}}$ has 5 variables, meaning the output layer of a parameterized model using a bivariate Gaussian loss has also 5 entries, which is significantly less than the output layer of the OH model, meaning that using the same two hidden layers as in the OH models in the previous section, the parameterized model has only 16565 parameters. The time to train on $n_{\text{train}} = 10000$ samples for 20 epochs is 11.3 s and the inference time for 1000 samples is 53.1 ms, both faster than even the smaller one of the OH models. To illustrate this point, a smaller model is trained for 20 epochs, and an example of the reconstruction of this bivariate Gauss model is shown in [Figure 3.6](#), with the input hitpattern on the left, the predicted normal distribution in the center and zoomed-in to the right. The true and predicted positions are indicated in black and red respectively.

The resolution ρ as a function of the S2 top area is shown in [Figure 3.7](#). While at small S2 area, the parameterized model shows a worse resolution than the OH models, at larger areas the resolution is better than that of the bigger of the two OH models. It can be expected that with extended training, this Gaussian loss model will outperform any OH model of similar size.

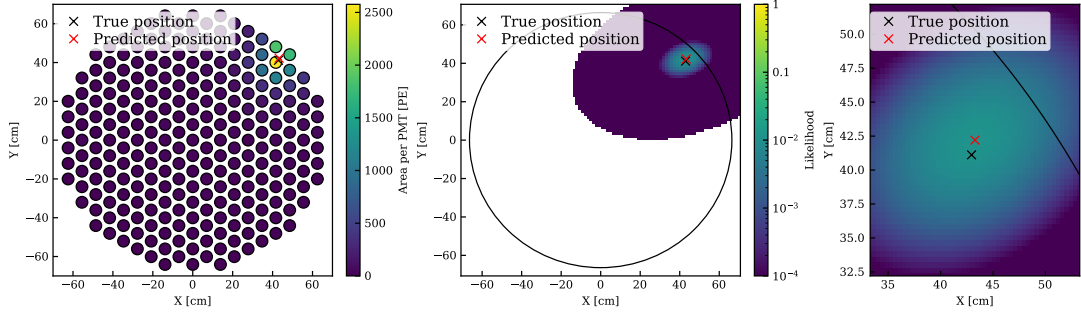


Figure 3.6: A position reconstruction using a Gaussian loss model. Left: The input hitpattern that serves as input for the NN. Center: The normal distribution, predicted by the Gaussian loss position reconstruction model. Right: A zoom-in of the PDF shown in the central panel. In addition, the true and predicted positions are portrayed in black and red respectively. The black circle indicates the edge of the TPC.

Table 3.2: The advantages and disadvantages of the OH model and the parameterized model.

One-Hot model	Parameterized model
Allows modeling of arbitrary probability distributions	Only pre-defined PDF
Resolution scales trivially with bin width	Performance dependent on model architecture
Major increase in number of parameters	Minor increase in number of parameters
Output probability distribution not necessarily tractable	Output PDF is analytical function

3.3.3 Comparison and Combination of Probabilistic Extensions

To summarize the above [subsection 3.3.1](#) and [subsection 3.3.2](#), a parameterized output extension reaches a better resolution than an OH model of comparable size, while being faster in both training and inference. However, the OH model offers one clear advantage in that it does not presuppose any PDF apriori, while the parameterized model needs a fixed analytical PDF hard built into the loss function. The output of the OH model can mimic any arbitrary underlying distribution, however computing statistical variables like the standard deviation is much easier with a pre-defined PDF, than it is for an arbitrary distribution. [Table 3.2](#) summarizes the advantages and disadvantages of either model.

Since the main disadvantage of the parameterized model stems from presupposing a PDF that does not necessarily reflect reality, and the OH model can provide arbitrary distributions, an easy solution is to use the OH arbitrariness to determine the underlying PDF, and utilize the such determined PDF as a loss for a parameterized model. To determine that analytical function which best models the underlying position probability distribution, a high resolution OH model is trained, with the goal of comparing its output to different analytical PDFs. The high resolution OH model has a bin width of 2.5 mm and 3 hidden layers with 180, 120 and 60 nodes each, which results in a model with $n_{\text{par,OH}} \approx 13 \times 10^6$ parameters. The dataset used for training contains only

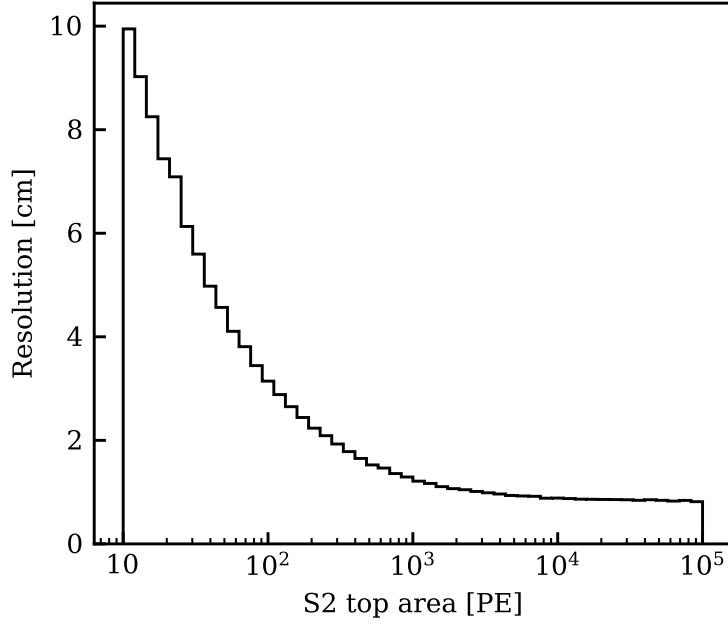


Figure 3.7: The resolution ρ derived with Equation 3.4 of the Gaussian loss model as a function of the S2 top area.

$n_{\text{train}} \approx 1.4 \times 10^6$ samples, making this model underdetermined.

An example prediction can be seen in Figure 3.8, with the input hitpattern on the left and the prediction in the central and right columns. The prediction seems to have multiple clusters of bins with a high probability. The high resolution model is more accurate than the small demonstration models shown in subsection 3.3.1, but it does not perform as well as possible, due to the disproportionate ratio between model parameters and training samples, making this high resolution OH model is overtrained. However, this does not reflect an error in the training procedure, since the model is not supposed to predict the PDF of new data, but rather learn the underlying PDF of the data it was trained on.

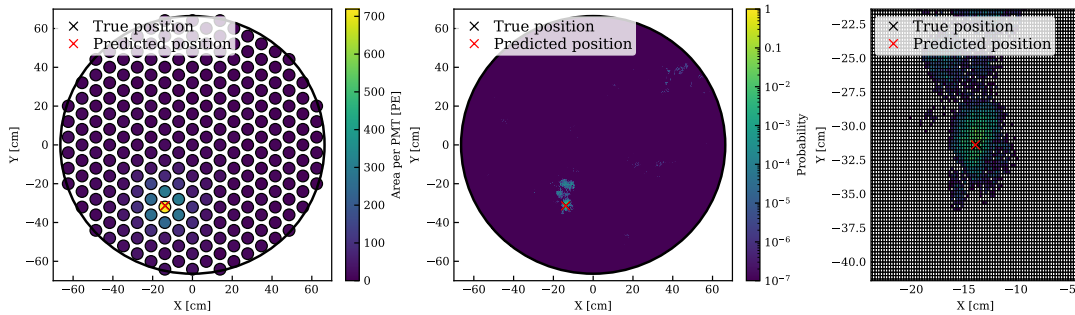


Figure 3.8: The position of an example hitpattern (left) reconstructed by the high resolution OH model. The predicted probability per bin is shown in the center and zoomed-in on the right.

The procedure to determine the best matching analytical function is as follows:

1. The high resolution OH model is used to predict the PDFs, $P_{\text{OH}}(x, y)$, of a set

of hitpatterns.

2. The distributions $P_{\text{OH}}(x, y)$ are translated such that the mode is at $(x, y) = (0, 0)$ and cut to 10 cm squares around the mode, to trim data and simplify computing.
3. Each 10 cm square is normalized to make the sum of all bins equal to 1.
4. 1000 positions are randomly sampled according to the distribution of the square:
 $\vec{r}_{\text{sample}} \sim P_{\text{OH}}$.
5. A maximum likelihood fit is performed on the sampled positions for a selection i of pre-defined PDFs to obtain a best-fit distribution $P_{\text{fit},i}$.
6. The negative log of the MSE between best fit and OH output is computed as
 $f_i = -\log(P_{\text{OH}} - P_{\text{fit},i})^2$, where i is the index of the different analytical PDFs.

The pre-defined PDFs are a bivariate Gaussian, a Skew Gauss (SG) distribution, a Cauchy distribution and a symmetric Laplace distribution and an asymmetric Laplace distribution. Figure 3.9 shows a map of the best matching analytical function, meaning the one with the highest derived value of f_i , as a distribution of the radius squared and the S2 top area. For the vast majority of parameters, the best matching analytical PDF is a 2-dimensional SG distribution, in particular for hitpattern of higher S2 top area, where it can be expected that the high resolution OH model can predict the position probability distribution the best. However, in general all of the analyzed PDFs did match the sampled positions very well.

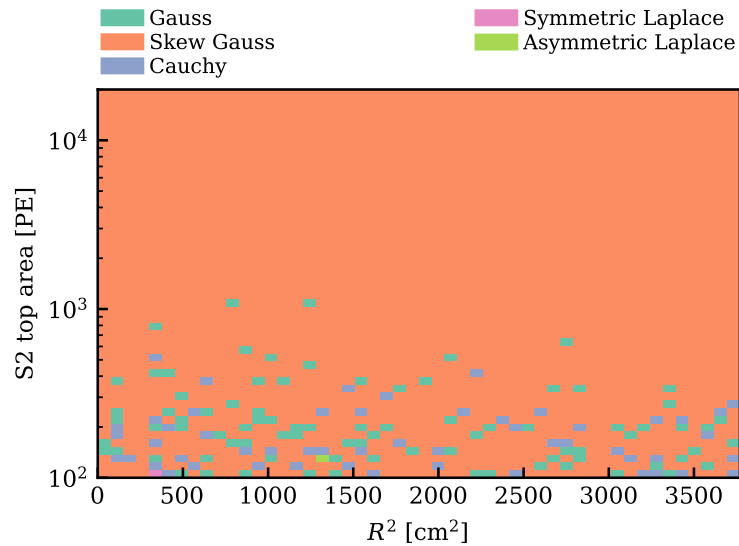


Figure 3.9: The analytical function with the highest f_i as a distribution of radius squared and S2 top area. The SG function is favored for the majority of the parameter space, and in particular for the high S2 top area region, where it can be assumed that the fits perform the best.

3.3.4 The Skew Gauss Position Reconstruction Model

Based on the insights described in [subsection 3.3.3](#), a parameterized model utilizing the SG likelihood is trained to perform position reconstruction. The definition of the bivariate Skew Gauss (SG) distribution is taken from [142]:

$$P_{\text{SG}}(x, y | \mu_x, \mu_y, \sigma_x, \sigma_y, \rho, \delta_x, \delta_y) = 2 \cdot \mathcal{N}_2(x, y) \cdot \Phi(\delta_x(x - \mu_x) + \delta_y(y - \mu_y)), \quad (3.12)$$

where $\mathcal{N}_2$ is the bivariate normal distribution from [Equation 3.9](#) and $\Phi(x)$ is the CDF of the *one dimensional* normal distribution with zero mean and unity standard deviation:

$$\Phi(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right]. \quad (3.13)$$

The parameters μ_i, σ_i ($i \in \{x, y\}$) and ρ are the same as in the bivariate Gaussian, the mean position, standard deviation and correlation parameters respectively. The parameters δ_x and δ_y define the skewness of the distribution. The negative log-likelihood of this distribution is

$$\mathcal{L}_{\text{SG}} = \mathcal{L}_{\mathcal{N}_2} - \log(\Phi(\delta_x(x - \mu_x) + \delta_y(y - \mu_y))), \quad (3.14)$$

where $\mathcal{L}_{\mathcal{N}_2}$ is the negative log-likelihood of the unskewed normal, shown in [Equation 3.11](#).

The SG position reconstruction model is an MLP using 8 hidden layers of 180, 120, 80, 60, 40, 30, 20 and 7 neurons respectively, leading to a total number of $n_{\text{par}} = 86473$ parameters. The SG model is trained for 1000 epochs on 1.8×10^6 samples and validated during training on $n_{\text{val}} = \times 10^5$ samples, using [Equation 3.14](#) as loss function.

[Figure 3.10](#) shows the training and validation loss of the SG model, where after around 500 epochs the validation loss converges and the model does not improve anymore. During training, the learning rate of the gradient descent is reduced by a factor of 0.8 each time the validation loss did not decrease for the past 20 epochs, until a minimum learning rate of 10^{-4} is reached. The epochs where the learning rate was decreased are well visible in as sudden decreases of the training loss.

Examples of the prediction done by the SG model are shown in [Figure 3.11](#), for an S2 close to the TPC wall (top row) and for an S2 from the central bulk volume of the TPC (bottom row). The distribution of the bulk S2 looks rather symmetrical, while the edge S2s position probability distribution is elongated perpendicular to the TPC wall in radial direction. Both PDFs exhibit no apparent skewness.

[Figure 3.11a](#), which shows a hitpattern with smaller S2 area and an original position located near the TPC wall, has a much wider position uncertainty distribution compared to [Figure 3.11b](#), which shows the prediction of an S2 with greater area and originating from the TPC bulk. This is a trend that can be observed for all position reconstruction models: The greater the S2 top area, the better the resolution, and the position reconstruction model is more accurate in the TPC bulk, than near the walls.

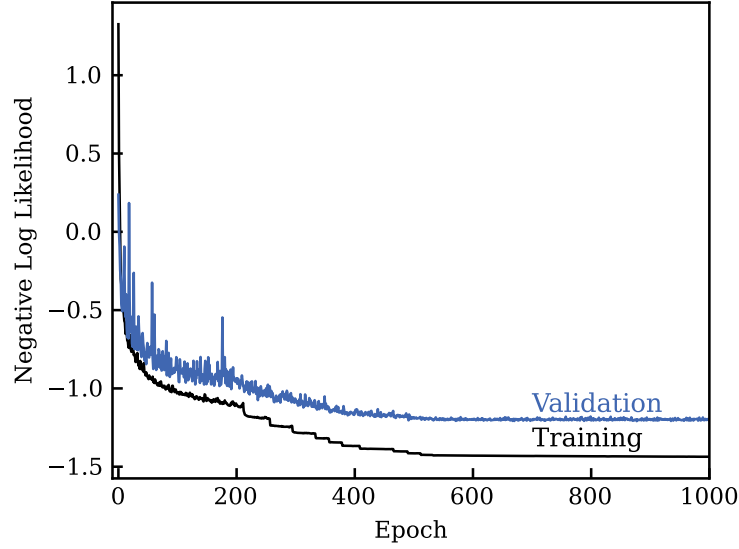


Figure 3.10: The training and validation loss of the SG model. The model was trained for 500 epochs while gradually reducing the learning rate by a factor of 0.8 every time the validation loss did not decrease for 20 consecutive epochs. Both, the validation and training loss converged.

Figure 3.12 shows the average resolution ρ as a function of the S2 top area, separated into TPC bulk S2s with a radius $R \leq 60$ cm and TPC wall S2s with radius $R > 60$ cm. The SG model is compared to the MLP position reconstruction model from [112]. In the TPC bulk, the SG model has approximately the same resolution as the MLP used in SR0 and SR1, and for TPC wall events, the SG model performs even better than the MLP across all S2 top areas. It is to note here, that the main purpose of the probabilistic extension is not to improve the resolution ρ , but to provide additional information without degrading the models performance.

While the SG model predicts information on the position uncertainty for each individual S2, this information needs to be validated. To analyze whether the predicted position uncertainty actually represents reality, the coverage is computed and shown in Figure 3.13. The coverage relates predicted uncertainty to observed uncertainty of the positions: For some nominal coverage f_{nom} , in the horizontal plane that contains the fraction f_{nom} of the full CDF predicted by the SG model is computed, which forms an elliptical region around the predicted position. Following this, it is checked whether the true position lies inside this region or not. Note that for each S2, the region containing f_{nom} is different. The empirical coverage is the number of S2s, where the true position lies within the area computed from the respective predicted PDF divided by the total number of considered S2s $f_{\text{emp}} = n_{\text{inside}}/n_{\text{total}}$. An estimator that predicts the position uncertainty correctly will have $f_{\text{nom}} = f_{\text{emp}}$, indicated by the black dashed line in Figure 3.13. The SG model is overcovering by up to 5 % for the most part, which means it is overestimating the actual uncertainty of the predicted position. In general the empirical and nominal coverages are very close, meaning that the PDF provided by the SG model predicts the position uncertainty well.

Finally, the spatial distribution of the parameters, σ_x , σ_y , ρ , δ_x and δ_y , that determine

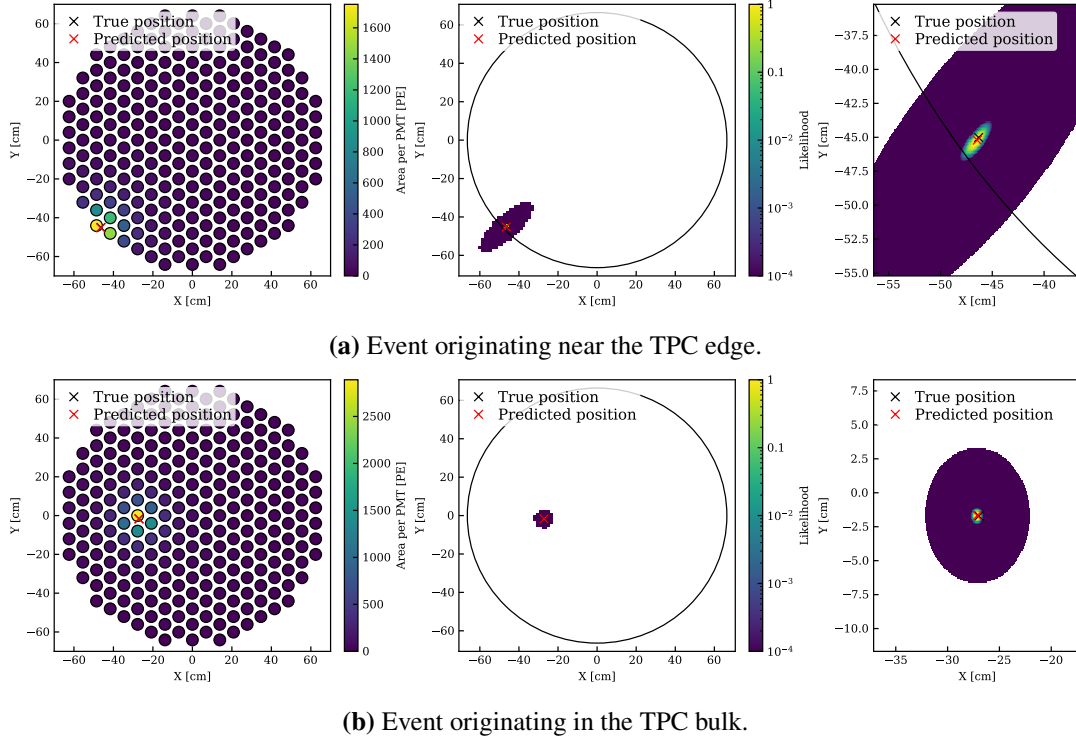


Figure 3.11: Example predictions of the SG model. The position of an example hitpattern (left) reconstructed by the SG model. The predicted PDF is shown in the central column, and zoomed-in on the right. The true position is shown in black and the predicted position is shown in red. One example originated from the edge of the TPC (top) and one from the central bulk of the TPC (bottom). For the edge example, the predicted PDF is elongated in radial direction. For the bulk example, the predicted PDF is nearly circular.

the shape of the PDF predicted by the SG model is visualized in [Figure 3.14](#). If $\sigma_x > \sigma_y$, then the width of the PDF is greater in x -direction than in y -direction, meaning that the PDF is elongated parallel to the x -axis, and the opposite, meaning an elongation along the y -direction, is true for $\sigma_x < \sigma_y$. In addition, the rotation of the PDF is quantified by ρ , where a positive value corresponds to a PDF rotated in clockwise direction and a negative value in counter clockwise direction. In combination, the distributions in [Figure 3.14a](#) and [Figure 3.14b](#) visualize that the ellipses of the SG PDFs are elongated in radial direction perpendicular to the TPC wall, similar to the example shown in [Figure 3.11a](#). In the bulk of the TPC the predicted PDFs are nearly circular. This can be interpreted as the model being certain about the position in the TPC bulk, bar some resolution driven by the number of detected photons, but at the edge of the TPC there exists an uncertainty mainly in radial direction which could be caused by absorption and reflection effects at the TPC walls. The skewness, characterized by δ_x and δ_y , is biased towards the nearest TPC wall, meaning that the PDF will fall quicker at the wall, and the tails of the distribution are elongated towards the TPC center. This suggests that the SG model learned that the TPC wall poses a hard limit for the predicted positions.

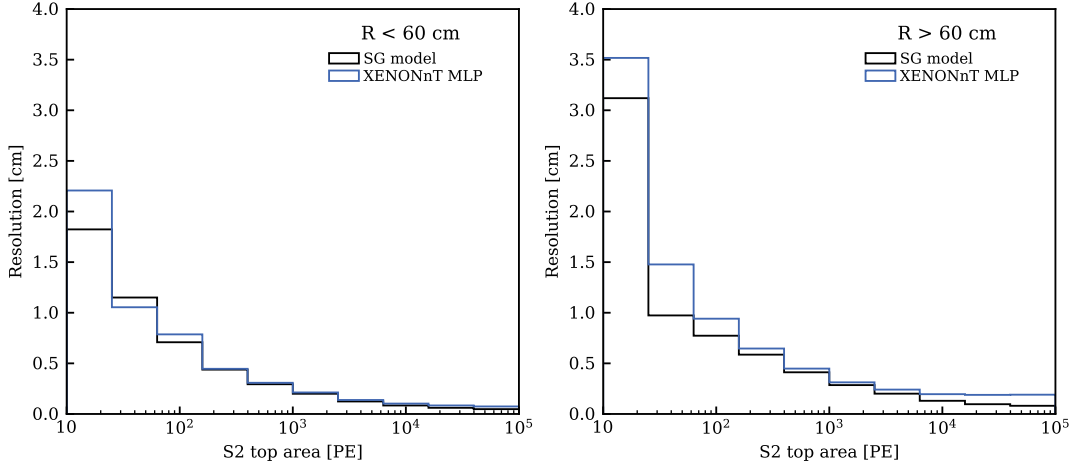


Figure 3.12: The resolution of the SG model compared to the MLP model from [112], separated into TPC bulk with $R \leq 60$ cm (left) and TPC wall with $R > 60$ cm (right) regions. The SG model is slightly more accurate, in particular for high S2 top area events at the TPC edge.

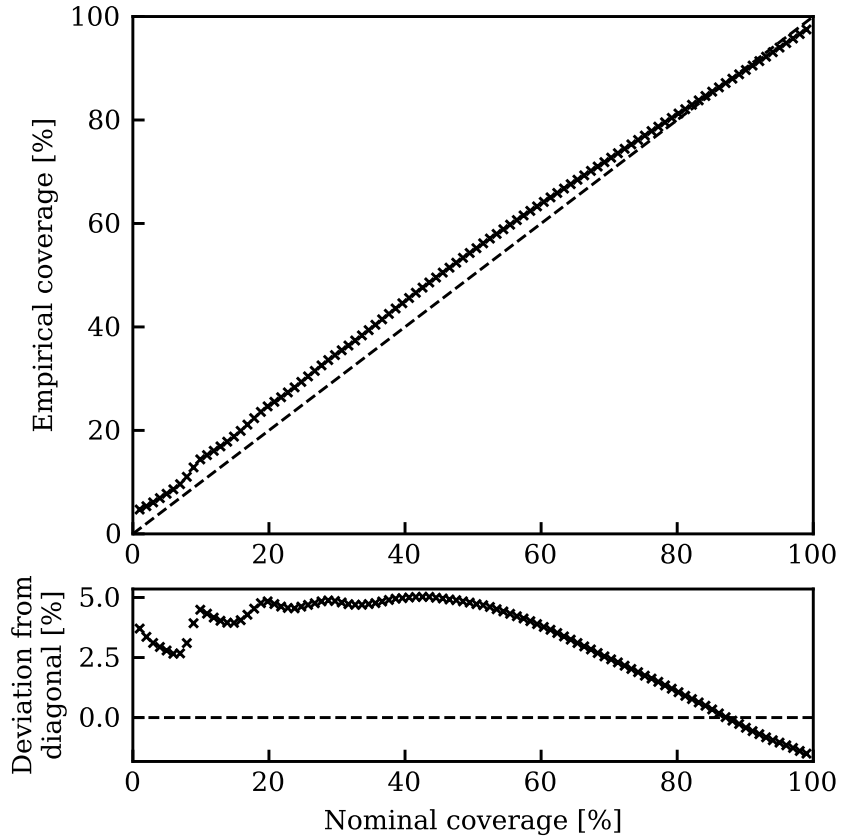


Figure 3.13: The coverage of the SG model (top). The nominal coverage defines a spatial region that contains some fraction of the PDF predicted by the model. The empirical coverage is the fraction of positions where the true position of the S2 lies within that region. For a perfect estimator, the empirical and nominal coverage are the same, which is visualized by the black dashed line. The bottom row shows the deviation from the SG predictions from the perfect estimator in percent points.

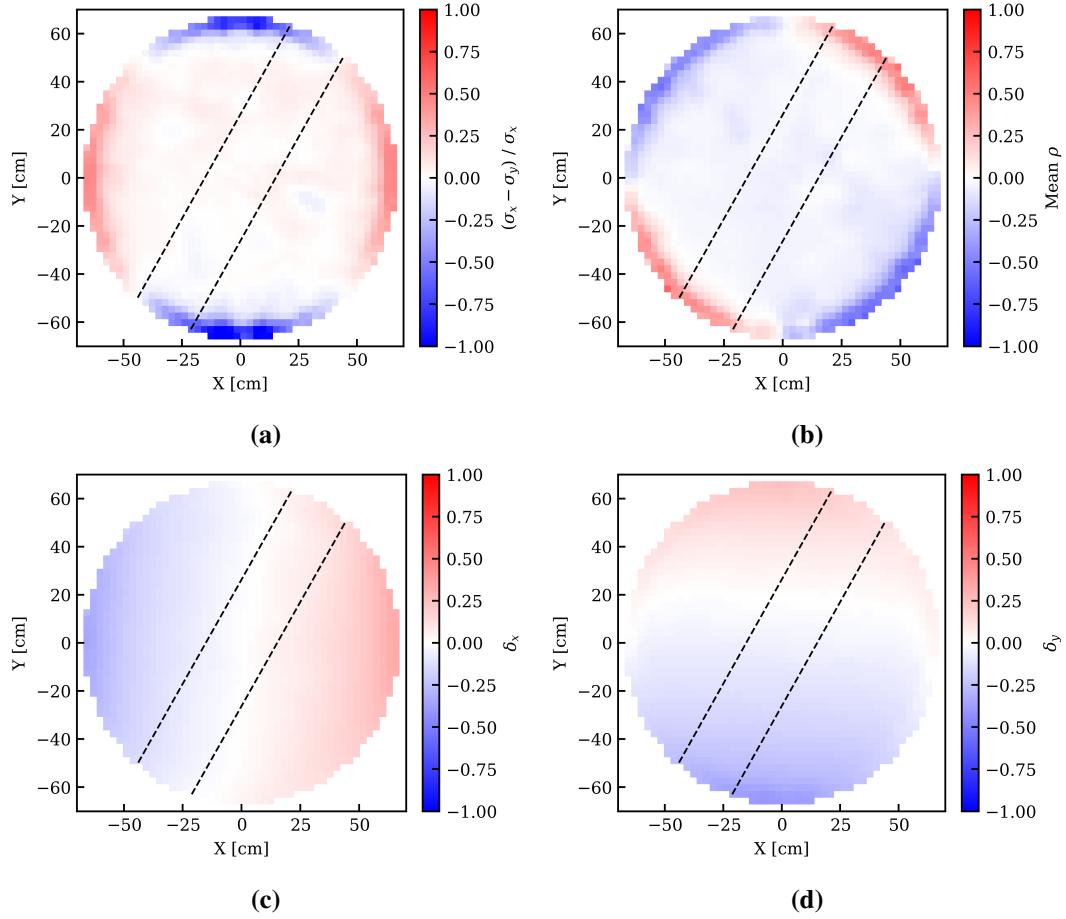


Figure 3.14: Distributions of various parameters of the predicted SG PDFs across the horizontal detector plane. The perpendicular wires of the anode are shown in black dashed. (a): The relative ratio of the width parameters σ_x and σ_y . A negative value (blue) corresponds to an ellipse elongated parallel to the y-axis. (b): The correlation parameter ρ . A positive value (red) corresponds to an ellipse elongated in a 45° angle. (c): The skewness in x direction δ_x . A positive value (red) corresponds to a distribution skewed to the right. (d): The skewness in y direction δ_y . A positive value (red) corresponds to a distribution skewed to the top.

3.4 Position Reconstruction using Conditional Normalizing Flows

One issue of the above described SG model is that it presupposes an analytical PDF that is not necessarily a correct model of the underlying position uncertainty. The One Hot model detailed in [subsection 3.3.1](#) is PDF-agnostic, but it comes with a heavy increase in model parameters, that complicates the whole training and testing process. Both models are extensions of NNs. Another way to make the position reconstruction probabilistic is to exchange the NN model by a different ML approach that inherently incorporates PDFs. One such category of ML models are so-called Normalizing Flows (NFs).

3.4.1 A short Introduction to Normalizing Flows

A Normalizing Flow (NF) is a type of ML model that differs from the classical NN framework introduced in [section 3.2](#). Whereas a NN relies on matrix multiplication and non-linear activation functions, a NF utilizes bijective mappings, called *flows*, to transform a simple and known base distribution into a complex and unknown target distribution.

Let $Y \sim p_Y(y)$ be an arbitrary random variable, with an unknown PDF, p_Y , named the target distribution, and $Z_0 \sim p_0(z)$ a random variable with a simple and known PDF such as a standard normal, called the base distribution. By finding a bijective and differentiable, and thereby also strictly monotonic, function f with $Y = f(Z)$, the target distribution, p_Y , can be determined via the change of variables formula

$$p_Y(y) = p_0(f^{-1}(y)) \left| \frac{\partial f^{-1}(y)}{\partial y} \right|, \quad (3.15)$$

where f^{-1} denotes the inverse function of f and $|\partial f^{-1}/\partial y|$ is the Jacobian determinant of f^{-1} . The challenge is finding a suitable function f , where the Jacobian, the inverse and Jacobian of the inverse can be computed efficiently, and that is flexible enough to actually map between the random variables such that $Y = f(Z)$ holds. A simple way of making a function more expressive is by chaining multiple uncomplicated functions in succession as a composite function, which is particularly useful for NFs, since a composite function $f = f_1 \circ f_2 \circ \dots \circ f_k$, where each f_i is invertible is again invertible, and its Jacobian determinant is the product of the Jacobian determinants of the constituents. Because of this, NFs can be constructed as the composition of simple, invertible, differentiable functions f_i that are made flexible via some nuisance parameters. Those nuisance parameters are the NF equivalent to the weights and biases of the NN in the sense that they are the parameters that are tuned during the training process [[143–146](#)]. The NF inference process is visualized in [Figure 3.15](#), where the NF model is trained such that p_k resembles p_Y .

Since the whole structure of the NF is invertible, it can *flow* in both directions: Samples of the complex target distribution, $p_Y(y)$, can be generated by drawing samples from p_0 and propagating them through the flow. Furthermore, given a sample y from the target distribution, its probability under p_Y can be computed by propagating it in

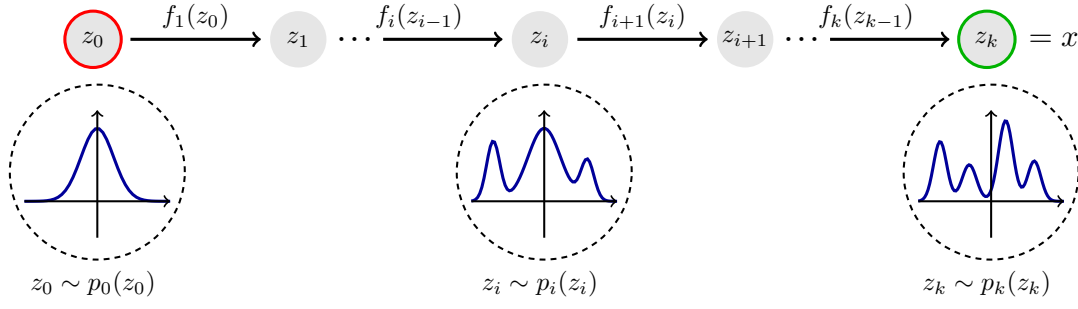


Figure 3.15: A schematic of the working principle of a NF, taken from [147]. A simple base distribution p_0 is transformed via bijective flows f_i into a complex target distribution p_k .

the inverse direction of the NF and evaluating the probability of $p_0(f^{-1}(y))$, which is easy to compute by design.

Originally, the NF was proposed as a form of unsupervised learning to model complex target distributions and draw new samples from them. However, the NF can easily be extended to work on conditional PDFs as a Conditional Normalizing Flow (CNF), accepting conditional variables [148]. In the case presented in the following, the CNF is trained to answer the question “What is the probability of the S2 having originated in position (x, y) , given the condition that the hitpattern $\vec{h}$ was observed?”

3.4.2 Model Definition and Training

The Conditional Normalizing Flow (CNF) position reconstruction model used in SR2 of XENONnT and subsequent runs is implemented and trained using the *flowjax* software [149]. The base distribution is a bivariate Gaussian and the flow layers f_i are coupling layers [148] using rational quadratic splines [150] as bijections. The parameters of the bijective transformations are controlled by an MLP, consisting of 3 hidden layers with 128 neurons each. The CNF has 5 coupling layers, and the input $\vec{h}'$ consists of the normalized 253 entry hitpattern, concatenated with the logarithm base 10 of the S2 top hitpattern, resulting in a model with $n_{\text{par}} = 338550$ parameters. The loss function is the expectation value across all training samples of the negative logarithm of the predicted PDF p_k , which is equivalent to minimizing the Kullback-Leibler divergence (KL) between p_K and p_Y [1]

$$\mathcal{L}_{\text{CNF}} = -\mathbb{E}_{\vec{r} \sim p(\vec{r}), \vec{h}' \sim p(\vec{h}'|\vec{r})} [\log(p_k(\vec{r}|\vec{h}'))]. \quad (3.16)$$

Here, $\vec{r}'$ denotes the transformed position in the horizontal detector plane. A position transformation is used because the CNF output is meant to be a PDF across an unbounded 2D plane, while all of the learning samples are bounded in the position by the extent of the TPC walls. The transformation is defined by

$$\vec{r}' = \tanh \left(2(1 - \epsilon) \cdot \left(\frac{\vec{r}}{2(r_{\text{TPC}} + 20 \text{ cm})} + \epsilon \right) \right) + \epsilon, \quad (3.17)$$

with the untransformed position $\vec{r}$, the TPC radius $r_{\text{TPC}} = 66.4 \text{ cm}$ and a small value for numerical stability $\epsilon = 10^{-7}$.

The optimizer used for the gradient descent is the AdamW algorithm [136], with a maximum learning rate of 6×10^{-4} , a learning rate reduction factor of 0.15 and a weight decay factor of 8×10^{-4} . The model was trained for 150 epochs on 9.8×10^6 samples with S2 top areas log-uniform between 10 PE and 1×10^5 PE, generated with *fuse* using an updated optical simulation setting constructed for XENONnT SR2. In addition, 9.8×10^5 artificial multi-scatter (MS) S2s were included in the training data, to provide the model with samples of high position uncertainty for learning. These MS signals are constructed by adding the hitpatterns of two S2s from the original training set together. The true position of an MS is assumed to be the position of the bigger (in area) of its two constituents.

An example of the CNF position reconstruction is shown in Figure 3.16, with the input hitpattern to the left, the log-posterior in the center and zoomed-in on the right. The true simulated position is shown in black and the mode of the posterior PDF, which is taken as the predicted position, is shown in red. The predicted position matches the true position very well.

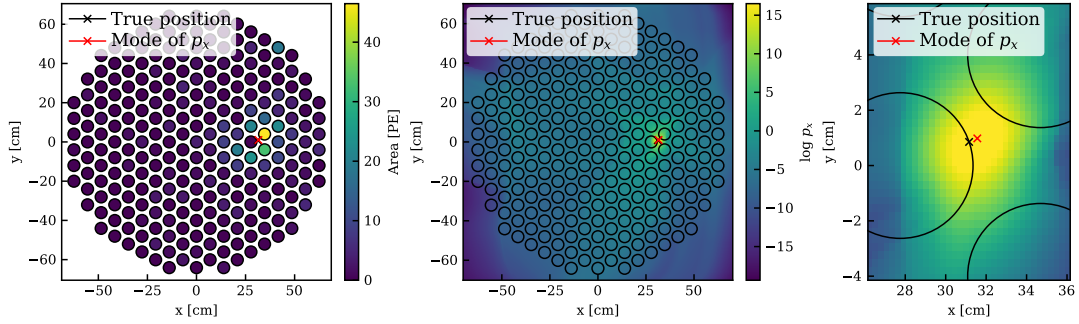


Figure 3.16: An example of the CNF position reconstruction. The input hitpattern can be seen on the left, the predicted PDF in the horizontal plane in the center and zoomed-in on the right. The true and predicted positions are shown in black and red respectively.

3.4.3 Model Validation and Performance

To properly validate and compare different models, an MLP position reconstruction model is trained using the same settings and training procedures as the CNF model. Figure 3.12 shows that the resolution of an MLP is approximately equal to that of an SG model, which allows conclusion drawn from the comparison of CNF and MLP to be extended to the SG model too. The MLP and CNF models presented in this section are the default position reconstruction models used in SR2 of XENONnT, with the CNF being the default algorithm.

The performance of the CNF and MLP models is evaluated on 2×10^5 S2s simulated with the same settings as the training data. Figure 3.17 shows the distribution of positions predicted by the MLP (left) and CNF (right). Overall, the predictions are uniform in the horizontal plane, reflecting the uniform distribution of the true positions of the simulated hitpatterns. However, the MLP is predicting a higher density of S2s in a ring just inside of the TPC wall, and a lack of events just at the wall. This is highlighted in Figure 3.18, where the distribution of predicted radius squared is shown in addition to the simulated truth. The CNF predicts events closer to the wall, whereas the MLP clusters events in a ring a small distance inside from the wall.

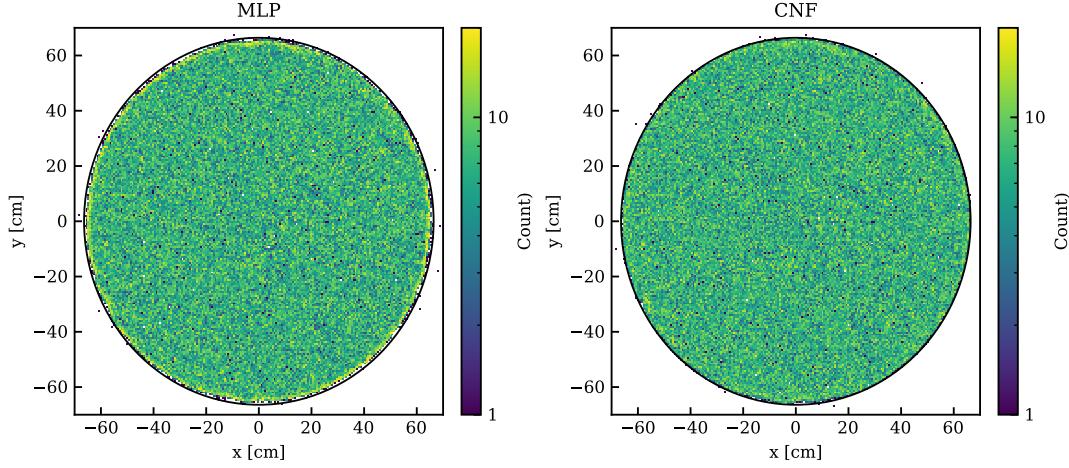


Figure 3.17: The distribution of positions predicted by the MLP (left) and CNF (right) models. The simulated true positions were uniform in the horizontal plane.

Figure 3.19 shows the radial bias of the two position reconstruction models as a function of radius squared (left) and azimuthal angle θ (right), where a positive (negative) value corresponds to a bias in inward (outward) direction. Near the TPC wall both models show a tendency to bias the position inwards, which can be explained by both models having learned that the TPC walls pose hard limits on the possible positions. Thereby if the model is wrong, it is only wrong in one direction, not equally in both. Note that neither model has a hard limit on the predicted radius baked in, which is exemplified by there existing individual events predicted outside of the TPC wall in Figure 3.17. For the central bulk of the TPC the models both show a very slight bias in outward direction of 0.06 cm on average for the MLP and 0.05 cm on average for the CNF. The right hand diagrams of Figure 3.19 in particular show that the prediction of both models is rotationally symmetric, as expected.

The resolution ρ of both models is shown in Figure 3.20, separated into the TPC bulk below $R \leq 60$ cm (left) and TPC wall above $R > 60$ cm (right). The resolution is fitted, using a least squares minimization implemented in the scipy package [151], with the following empirical function

$$\rho(A_{\text{top}}) = \frac{a}{\sqrt{A_{\text{top}}}} + b. \quad (3.18)$$

The best fit values and the uncertainties from the fit are written into Figure 3.20, and for the TPC bulk summarized in Table 3.3. Both models show a very similar resolution, with the CNF being slightly more accurate than the MLP for S2s near the TPC wall.

The main purpose of introducing the CNF model is not to improve the resolution ρ , since for most position reconstruction models ρ is already close to the optimal attainable level described in section 3.1. Rather, the CNF model provides additional information about the position uncertainty, and as shown in Figure 3.20, it does so without degrading the model accuracy and even improves the resolution for S2s near

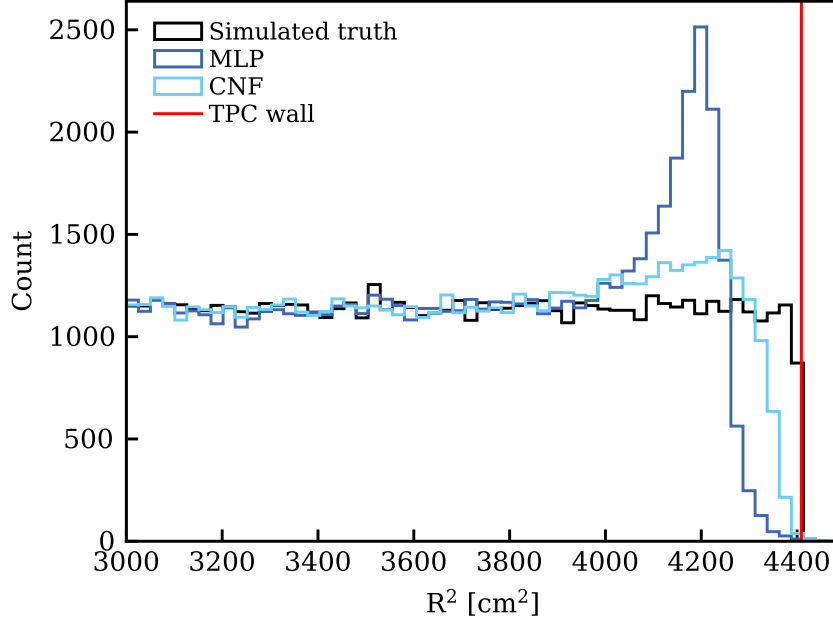


Figure 3.18: The distribution of predicted radius squared for the MLP (dark blue) and CNF (light blue) models. The red line shows the TPC wall at $r_{\text{TPC}} = 66.4$ cm, and the black line shows the simulated true positions.

the TPC walls. The additional information provided by the CNF comes in form of the probability distribution across the full horizontal plane for each position.

However, in XENONnT, the position is predicted for every recorded S2 peak, and storing the full PDF information for each of them would require unfeasible large amounts of data storage. To save on data storage, the information about the position uncertainty is broken down into the form of a contour in (x, y) space, that contains approximately one standard deviation of the predicted CDF. Finding the area in the horizontal plane that covers $k\%$ of the CDF predicted by the CNF is rather simple: Due to the monotonic nature of the flows, if a contour $\vec{c}_0$ covers $k\%$ of the base distribution, then the propagation through the flow $\vec{c} = f(\vec{c}_0)$ covers the exact same percentage of the target distribution. For a bivariate standard normal base distribution, the contour $\vec{c}_0$ is a circle for which the radius r_k can be determined by the inverse CDF of the χ^2 -distribution with two degrees of freedom

$$r_k = q_{\chi^2_2}(k). \quad (3.19)$$

The function q is the quantile function, which is equal to the inverse CDF, of the χ^2 distribution. The contour in the base distribution, $\vec{c}_0$, is approximated as a polygon with N corners. Two example contours in the posterior distribution, $\vec{c} = f(\vec{c}_0)$, are shown in Figure 3.21, using $N = 4$ and $N = 16$ corner polygons. To deduce the optimum number of corners, a coverage study was performed, the result of which is depicted in Figure 3.22. The definitions of nominal and empirical coverage are described in subsection 3.3.4. Simplified, the nominal coverage is what is asked of the model to be predicted, and the empirical coverage is what is observed. Using a triangle or tetragon underestimated the actual position uncertainty, while using an octagon was already surprisingly accurate. For XENONnT SR2 it was decided that the CNF is to predict a polygon with $N = 16$ corners. Figure 3.22 shows that this achieves an

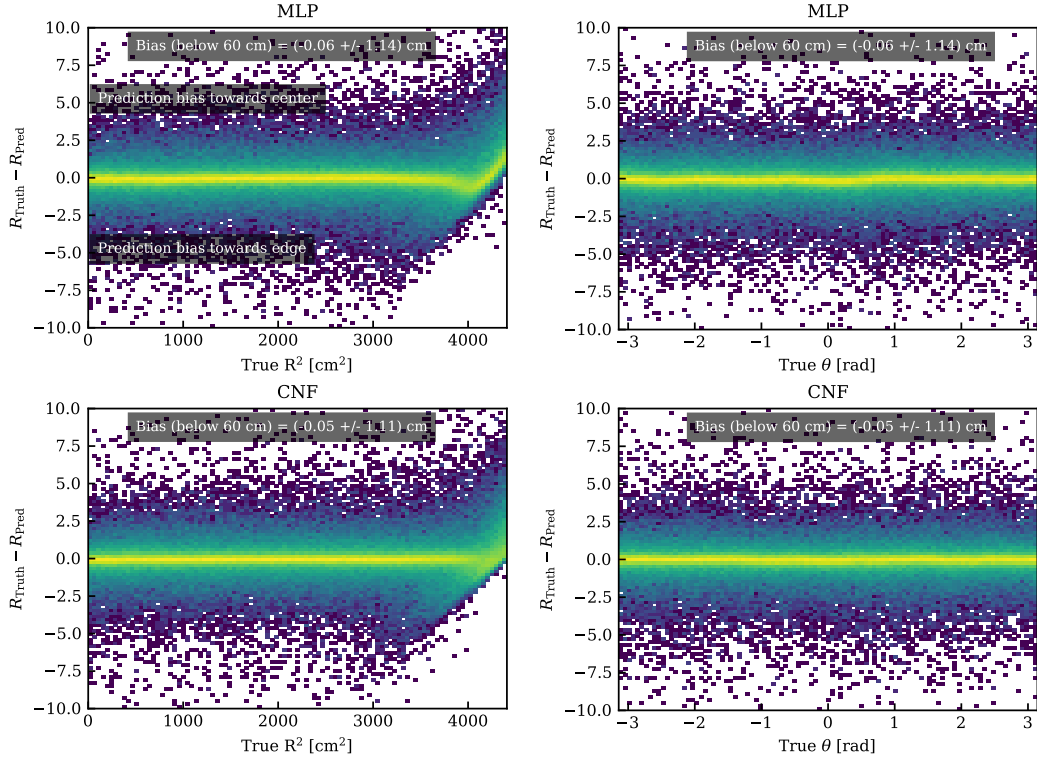


Figure 3.19: The radial bias of the MLP (top) and CNF (bottom) position reconstruction models. Left: The radial bias as a function of the true simulated radius squared. Right: The radial bias as a function of the true simulated azimuthal angle, using only events with $R \leq 60$ cm. The average biases for events inside of 60 cm and their standard deviation are written onto the figures.

extremely accurate coverage, with a maximum overcoverage of only 2 %, meaning the position uncertainty predicted by the CNF resembles the actual underlying uncertainty very well.

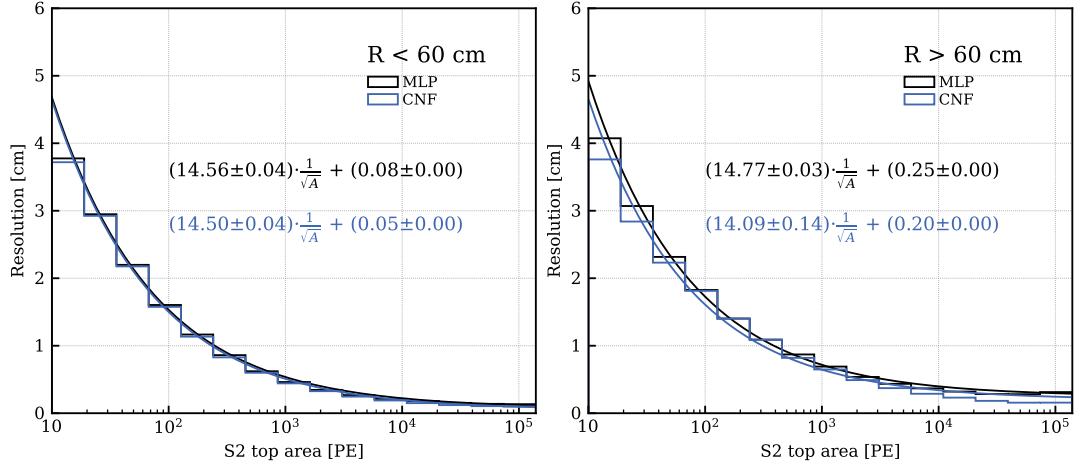


Figure 3.20: The resolution ρ as a function of the S2 top area for the CNF and MLP models, separated in TPC bulk below $R \leq 60$ cm (left) and TPC wall above $R > 60$ cm (right). The resolution has been fitted using Equation 3.18, and the best-fit parameters are included in the figure.

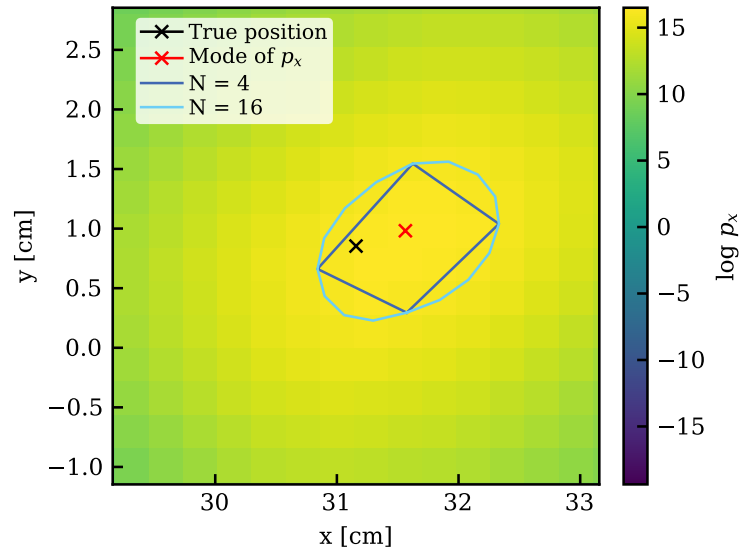


Figure 3.21: Two example contours, approximating the $1\text{-}\sigma$ region of the PDF predicted by the CNF model, using the same example hitpattern as Figure 3.16. The true position is shown in black, the predicted position in red. The two contours in dark and light blue approximate the area that contains one standard deviation of the predicted CDF, using polygons with $N = 4$ and $N = 16$ corners respectively.

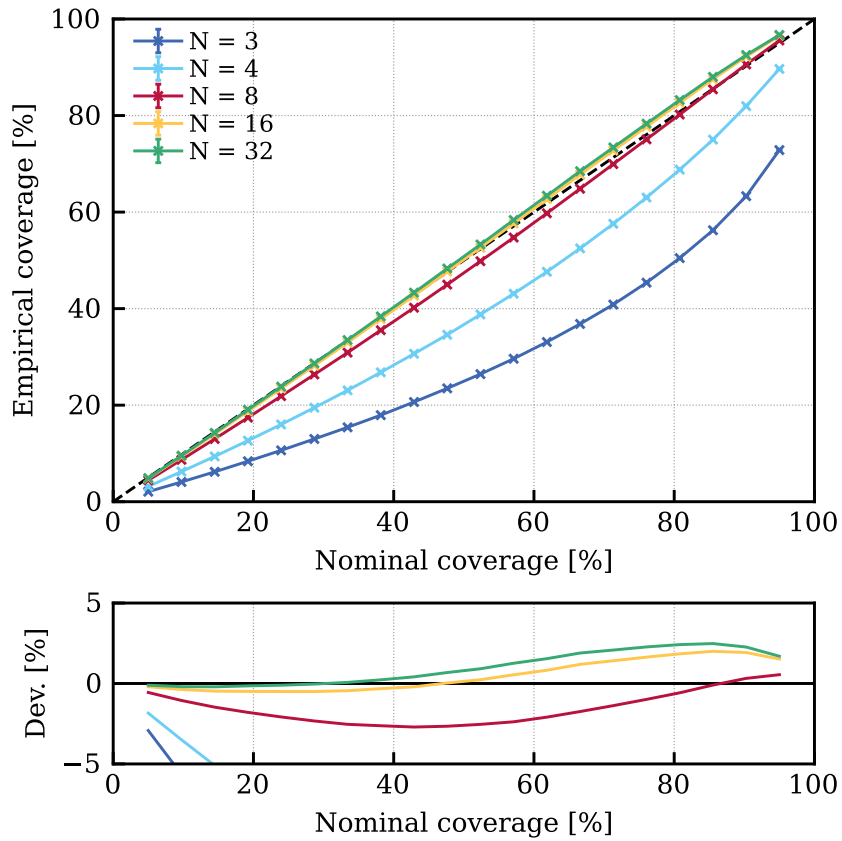


Figure 3.22: The coverage of the CNF model given different numbers of polygon corners N . The nominal coverage defines a spatial region that contains some fraction of the PDF predicted by the model. The empirical coverage is the fraction of positions where the true position of the S2 lies within that region. For a perfect estimator, the empirical and nominal coverage are the same, which is visualized by the black dashed line. The bottom panel shows the deviation from a perfect estimator in percentage points.

3.5 Data driven Validation of Position Reconstruction

All position reconstruction algorithms using ML techniques detailed in the previous sections are examples of simulation based inference (SBI) [152], and simulating a hitpattern produced by an S2 signal is a common task for simulation software such as GEANT4 [119]. With known values for photon absorption rate, reflectivity of the involved surfaces and photon acceptance by the PMTs, it is possible to make accurate simulations of the expected S2 top hitpattern, given a known initial position. The position reconstruction task poses the inverse problem: Finding the initial position of the S2, given an observation of a hitpattern and (approximately) known values for all optical parameters. This is done using, for example, the PLR method described in section 3.1, however, this involves tedious and non-trivial minimization of the PLR. A modern solution to such inverse problems where one direction, here the simulation of hitpatterns, is rather easy compared to the other direction, here the inference of the initial position, is SBI. In SBI an expressive estimator such as an ML model is trained on a multitude of examples of the easy direction of the problem to learn the underlying distribution of the hard direction of the problem, *without the need to formulate an explicit likelihood*. For this reason, SBI is sometimes also called likelihood-free inference. However, the whole idea of SBI hinges on the correctness of the simulations. This section presents a validation of the predictions made by the CNF position reconstruction model that is independent of the simulations and fully data-driven.

3.5.1 Single Electron Signals from Hotspots

The XENONnT experiment observed seemingly random bursts of single electron (SE) induced S2 signals, from a small set of fixed locations in the horizontal plane. These bursts are referred to as hotspots [65] and as of yet, there is no conclusive explanation for their origin. Figure 3.23 shows the distribution of S2 signals produced by SEs, selected via an S2 top area range between 10 PE and 40 PE. The position with the maximum amount of SEs, the hotspot, is marked in red and shows more than two orders of magnitude more SE activity than the rest of the TPC.

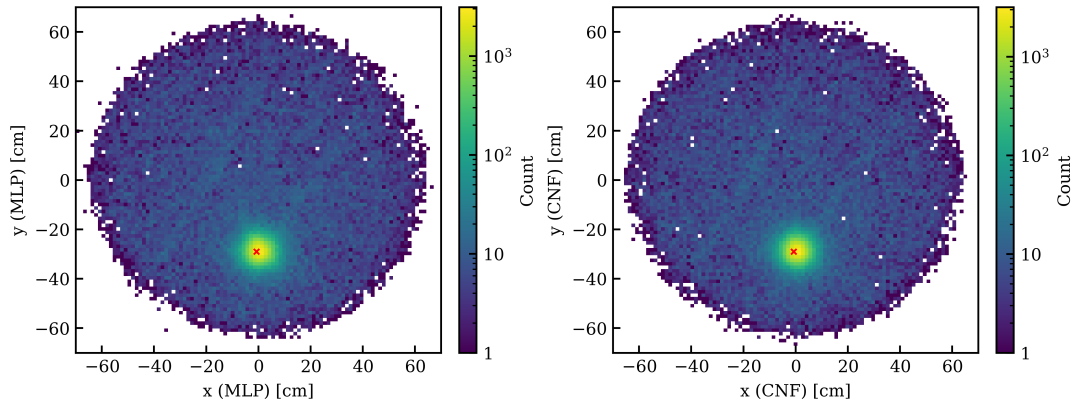


Figure 3.23: A distribution of all S2 peaks with a size between 10 PE and 40 PE for XENONnT run 065604, reconstructed by the MLP (left) and the CNF (right). In red is the maximum of the spatial distribution of the small S2s, the hotspot.

Under the assumption that the hotspots emerge from a point-like source, the horizon-

tal spread of SE events is caused by the finite position resolution. It turns out that the hotspots are not fully stationary, but slowly move over time, which smears the overall distribution of the SE positions. By splitting the data into time windows of length $\Delta t = 1$ min, the influence of the hotspot movement becomes negligible, and the remaining horizontal spread of the SE signals can be used to approximate the resolution of the position reconstruction algorithms. To construct the signal strength dependence the data is split by S2 top area into 6 bins between 10 PE and 40 PE. For each such bin, an unbinned likelihood fit is performed with the *iminuit* software [153], using a bivariate Gaussian PDF (see Equation 3.9).

The parameters σ_x and σ_y define the standard deviation in x and y direction respectively, and the parameter $\rho_{x,y} \in (-1, 1)$ quantizes the correlation between x and y . Figure 3.24 shows the SE distribution and the fit result for one example bin for both the MLP (left) and the CNF (right) position reconstruction models. The mean value of the fitted PDF is shown as red marker, and the ellipse defined by σ_x , σ_y and $\rho_{x,y}$ is shown as red line.

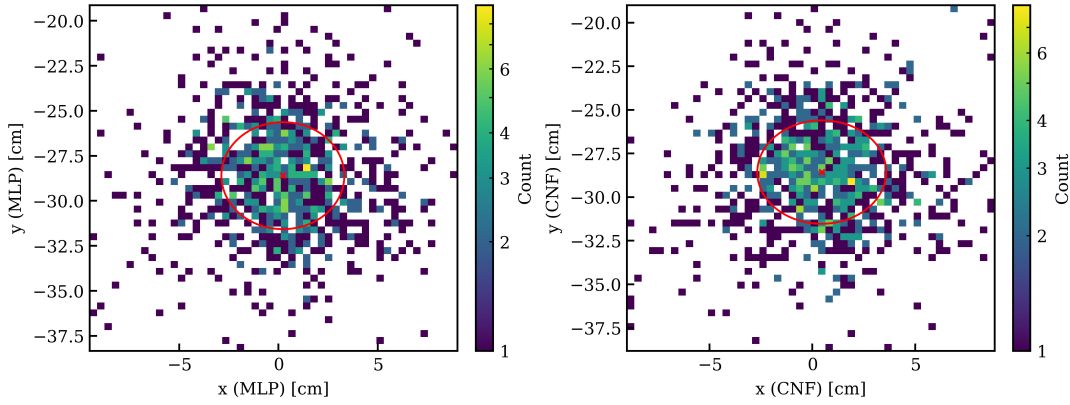


Figure 3.24: The spatial distribution of the SE signals of an example bin, reconstructed with the MLP (left) and CNF (right) model. In addition, the bivariate Gaussian mode (red cross) and $1\text{-}\sigma$ ellipse derived from the fit is shown (red).

This fit is performed for each bin, and the width w derived from one fit is defined as

$$w = \sqrt{\sigma_x \sigma_y \sqrt{1 - \rho_{x,y}^2}}. \quad (3.20)$$

However, all the values σ_x , σ_y and $\rho_{x,y}$ have their own uncertainties provided by the fit procedure, s_x , s_y and s_ρ respectively. Propagating those, the uncertainty s_w on the width w is computed per bin as

$$s_w = \frac{1}{2} \sqrt{\frac{(\sigma_y^2 s_x^2 + \sigma_x^2 s_y^2)(1 - \rho_{x,y}^2)}{\sigma_x \sigma_y \sqrt{1 - \rho_{x,y}^2}} - \frac{\sigma_x \sigma_y \rho_{x,y}^2}{(1 - \rho_{x,y}^2)^{3/2}} s_\rho^2}. \quad (3.21)$$

Finally, the resolution $\rho_{\text{hotspot}}(A_{\text{top}})$ is the average taken across all bins of the same S2 top area and the uncertainty on that resolution is the standard error on the mean, taking into account the uncertainties s_w of each fit. The result is shown in Figure 3.25. The data points show $\rho_{\text{hotspot}}(A_{\text{top}})$ and the dashed lines show the empirical functions

for ρ , fitted on the simulation derived resolutions from Figure 3.20. The data derived resolution and simulation derived resolution agree extremely well in the area range considered in this study. A secondary result of this validation is that the hotspot SEs do indeed originate from a mostly point-like source.

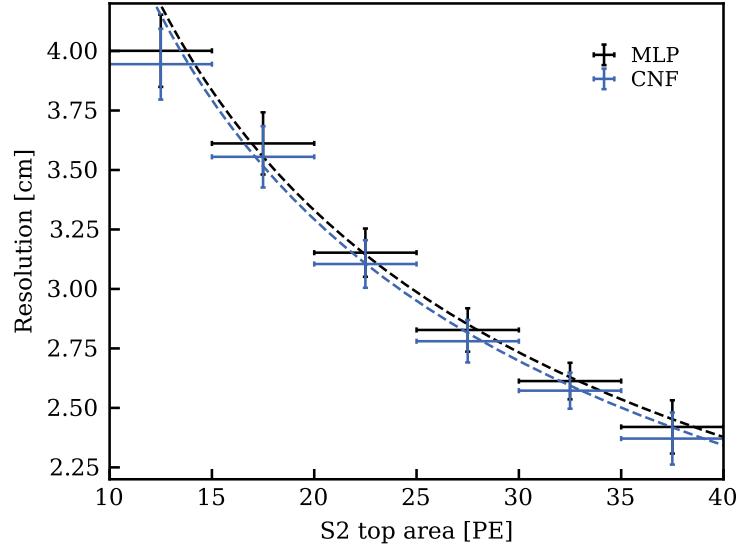


Figure 3.25: The resolution derived from the width of the hotspot. The uncertainty in S2 top area denotes the width of the area binning. The uncertainty in resolution is the error on the mean, derived as described in the text. In addition, the dashed lines show the result of the simulation driven resolution from Figure 3.20.

3.5.2 S2 Signals from beneath the Perpendicular Wires

As described in subsection 2.3.2, the XENONnT detector has perpendicular wires supporting the gate and anode. Through electric funneling, electrons at the top of the TPC are pushed aside by these perpendicular wires, creating narrow bands of higher signal density right beside them [154]. This is visualized in Figure 3.26, where the distribution of XENONnT SR2 events from 3700 background runs is shown in dependence of the S2 top area and the rotated x-coordinate x_{rot} . The rotation is chosen such that each perpendicular wire, shown as red line, has a constant position in rotated x, the gate perpendicular wire has $x_{\text{rot}} = 13.15$ cm and the anode wires have $x_{\text{rot}} = 14.15$ cm and $x_{\text{rot}} = 15.90$ cm. Three populations of higher signal density are clearly visible, and their spread is dependent on the S2 top area, because of the resolution of the position reconstruction algorithms.

The spatial extent of these bands is determined by the movement of the electrons and the position resolution. Assuming that the electrons between the gate wire at $x_{\text{rot}} = 13.15$ cm and the first anode wire at $x_{\text{rot}} = 14.15$ cm are confined by two competing electric fields and do therefore move as little as possible, an upper limit for the position resolution can be derived from the width of that signal band. The data from Figure 3.26 is divided into 11 bins equidistant in $\log_{10} A_{\text{top}}$ between $A_{\text{top}} = 500$ PE and $A_{\text{top}} = 5 \times 10^4$ PE. In each bin, a least squares fit of the count of S2 signals as a function of x_{rot} is performed with the *scipy* package [151] using a scaled normal distribution. The fit results of the first and last A_{top} bins are shown in Figure 3.27.

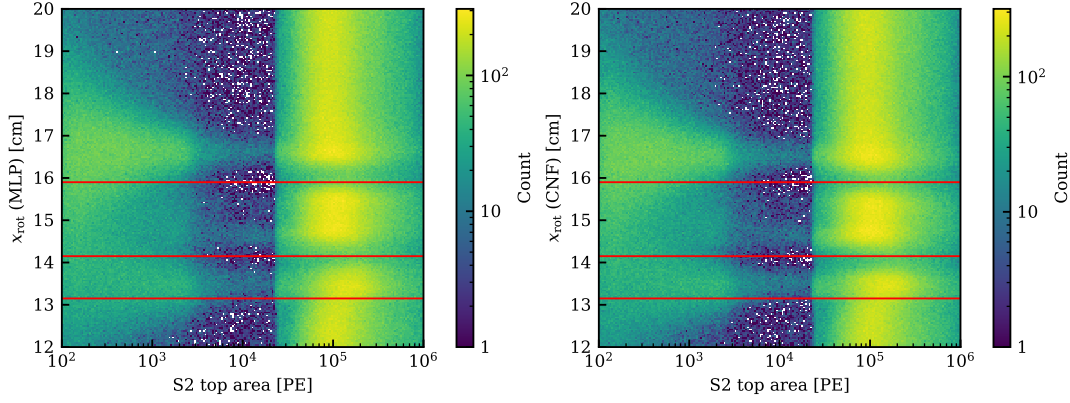


Figure 3.26: The distribution of signal density in S2 top area and rotated x coordinate, reconstructed using the MLP (left) and CNF (right). Shown in red are the positions of the perpendicular wires.

The fitted standard deviation σ serves as a data-driven estimate of the position resolution, assuming that the width of this Gaussian is dominated by the position uncertainty. [Figure 3.28](#) shows the results of the fits (data points) next to the simulation derived resolution (dashed) as a function of S2 top area. The uncertainties on the data points are the uncertainties passed on from the least squares fit and the widths of the A_{top} bins. The data matches the overall trend of the simulation derived resolution, but does not agree within uncertainty for the most part. In particular, that the Gaussian fit overestimates the resolution strongly at increasing A_{top} points towards there being another process driving the width of the signal band, in addition to the position resolution.

3.5.3 Result and Discussion

The data-driven position resolution from the hotspots in [subsection 3.5.1](#) and the perpendicular wires in [subsection 3.5.2](#) is combined in [Figure 3.29](#). For both position reconstruction models, the data points of the data derived method, a fit on these points using [Equation 3.18](#) and the fit using the resolution derived from simulations is shown. The parameters of the fit functions on the data-driven resolution are written in the figure and also listed in [Table 3.3](#). The two fit functions agree within 0.1 cm for all S2 top areas for both models. Since the simulation derived fit curve and the data derived fit curves agree so well, the simulation derived function can be used to estimate the resolution.

Having validated the simulation driven resolution, the final step is to also validate the position resolution directly predicted by the CNF model for each individual S2. The CNF computes a polygon $\vec{c} = \{(r_i, \theta_i)^T | i \in 1 \dots 16\}$ with 16 corners that contains 39.3 % (one standard deviation of the 2D normal) of the predicted CDF for each S2, as well as the mode of the PDF $(x_{\text{CNF}}, y_{\text{CNF}})$ or $(r_{\text{CNF}}, \theta_{\text{CNF}})$ in radial coordinates. If the polygon is approximated as an ellipse, the two axes serve as measures of the standard deviation in radial and angular directions. The ellipse axes are estimated as half of the maximum spread in either direction

$$\sigma_r = \frac{\max_r(\vec{c}) - \min_r(\vec{c})}{2} \quad \text{and} \quad \sigma_\theta = \frac{\max_\theta(\vec{c}) - \min_\theta(\vec{c})}{2}. \quad (3.22)$$

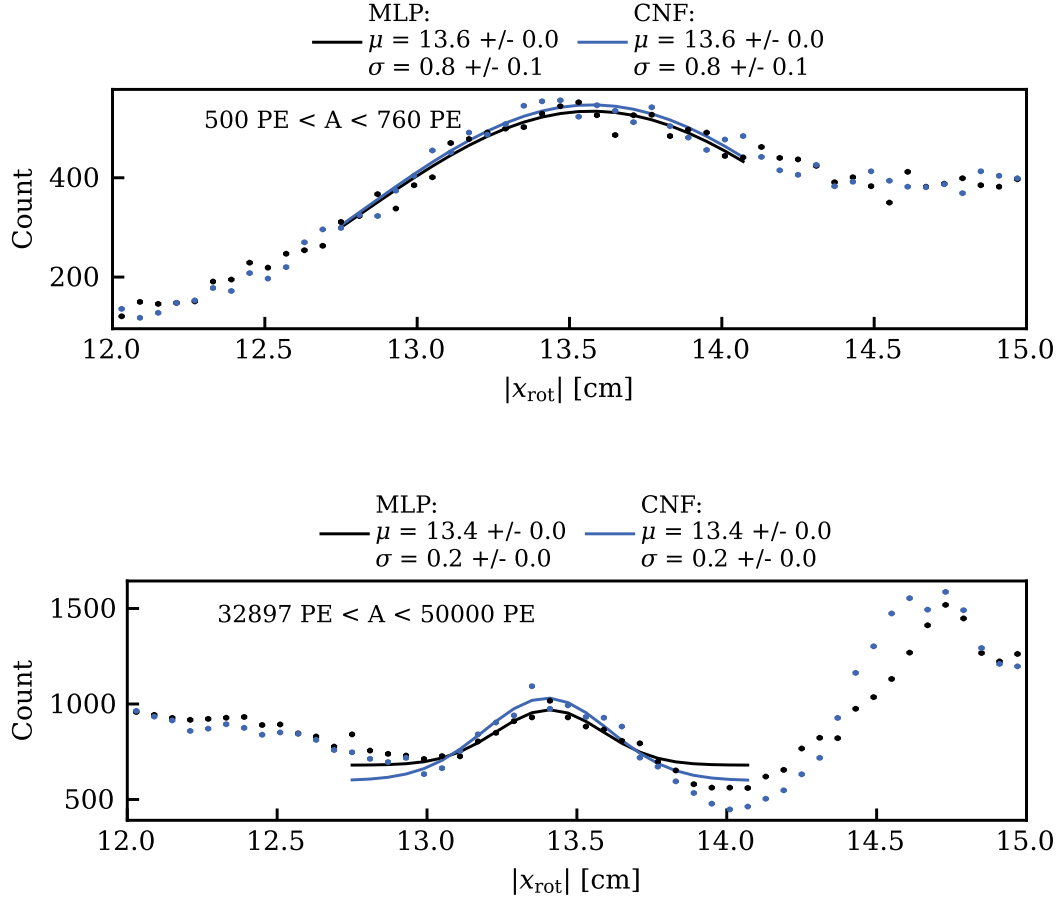


Figure 3.27: Two Gaussian fits of the signal population between gate and anode perpendicular wires, for the MLP (black) and CNF (blue) models. Top: The bin with the smallest considered A_{top} . Bottom: The bin with the highest considered A_{top} .

The average position uncertainty predicted by the CNF is

$$\rho_{\text{CNF}} = \sqrt{\sigma_r \cdot \sigma_\theta r}. \quad (3.23)$$

Figure 3.30 shows the position resolution (CNF only) derived from simulations (black, see Figure 3.20), from data (dark blue, see Figure 3.29) and finally from the immediate CNF output (light blue, see Equation 3.23). The CNF output is taken from the exact same events that were used to compute the data-driven resolution and the error bars are the standard error on the mean. The CNF predicted resolutions are also fitted with the empirical function Equation 3.18 using a least squares minimization implemented in the *scipy* package.

All three methods of computing the position resolution align within 0.1 cm across at least five orders of magnitude of the S2 signal strength. The fit parameters are collected in Table 3.3, with uncertainties listed from the fit procedure. While the fit parameters do not agree within that uncertainty, the differences amount to less than 0.1 cm for S2 signals below 100 PE, and less than 0.05 cm for stronger S2 signals, which is negligible for any actual use case in XENONnT analysis. To conclude, the

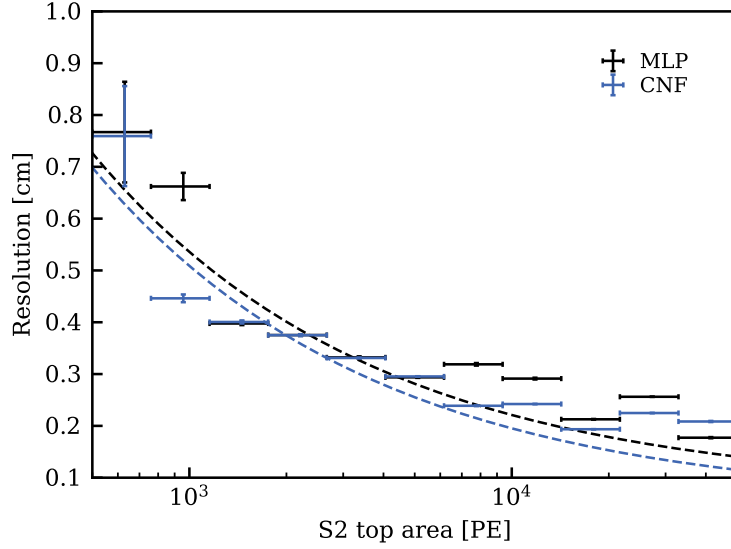


Figure 3.28: The fitted resolution from the perpendicular wire population (data points) and the resolution derived from simulations via the empirical function in [Figure 3.20](#).

Table 3.3: Best fit values for the empirical function for the resolution ρ [Equation 3.18](#)

	a [cm · PE ^{1/2}]	b [cm]
CNF (Simulation driven)	14.50 ± 0.04	0.05 ± 0.00
CNF (Data driven)	14.03 ± 0.03	0.10 ± 0.00
CNF (Direct model output)	14.24 ± 0.01	0.04 ± 0.00
MLP (Simulation driven)	14.56 ± 0.04	0.08 ± 0.00
MLP (Data driven)	14.13 ± 0.03	0.13 ± 0.00

CNF is not only the most accurate position reconstruction model for XENONnT to date, but also provides a reliable estimate of the uncertainty of its own prediction.

3.6 Applications of Probabilistic Position Reconstruction

The additional information provided by the CNF model has a wide array of applications, of which some are detailed in the following. One thing to note is that the XENONnT experiment is a collaborative effort by more than 200 scientists, and the CNF is providing all of these scientists with novel details to improve the search for DM. For this reason, two studies are summarized in the following, that were conducted by other scientists within XENON, with only minor input by me, which utilize the CNF predictions, namely the iterative Bayesian S2 merging developed by Dacheng Xu and Juehang Qin, described in [subsection 3.6.1](#) and the fuzzy definition of the perpendicular wire region made by Ivy Li, explained in [subsection 3.6.2](#). Further, three possible future applications of the CNF model are demonstrated below.

3.6.1 Iterative Bayesian S2 Merging

For SR2 of XENONnT, Dacheng Xu and Juehang Qin developed an S2 merging algorithm that takes into account the position uncertainty information provided by the

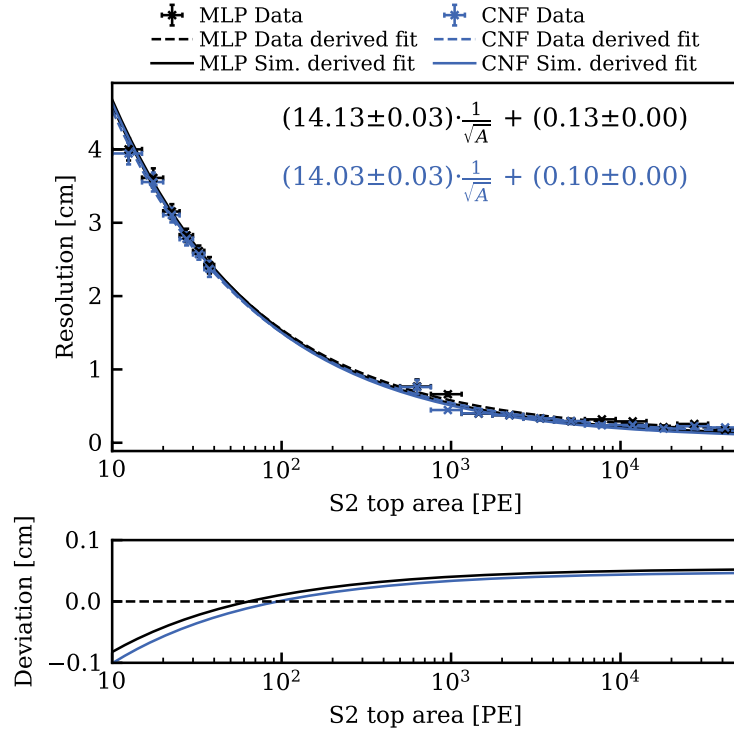


Figure 3.29: Top: The position resolution from data driven methods (data points), including a fit using Equation 3.18 (dashed), compared to the simulation derived resolution (solid). The fit function written in the figure are those derived from the data-driven resolution. Bottom: The deviation between the simulation derived fits and the data derived fits.

CNF. Due to the low drift field in XENONnT, the electron clouds that make up S2s can be spread wide in vertical direction, and for some S2s the signals of individual electrons of the S2 electron cloud do not overlap, separating what is supposed to be a single S2 into multiple peaks. In addition, XENONnT employs a peak splitting algorithm based on the natural breaks algorithm [155] that divides large S2s that have high potential of originating from multiple individual signals. The S2 merging algorithm takes care of the opposite, of combining individual sub-peaks, called peaklets, that have a high probability of belonging to a single S2.

In SR0 and SR1, the S2 merging was performed based on temporal information. The new algorithm in SR2, the iterative Bayesian S2 merging, also considers the position information. For each peaklet within a certain time window, the CNF predicts the position and the contour, the 16 cornered polygon from subsection 3.4.2, containing one standard deviation of the predicted PDF. Using these, the mean position of a group of peaklets is computed as the uncertainty weighted average. For each peaklet that potentially belongs to the main signal in this group, the uncertainty weighted distance to this mean position is calculated. The Bayesian algorithm then merges those peaks based on a threshold in the uncertainty weighted distance to the mean. This algorithm is iterative in the sense that it starts again after merging until all adjacent peaklets are either fully merged, or fully rejected from merging. The introduction of this algorithm successfully reduced the few electron backgrounds from delayed electron extraction and PI signals in SR2.

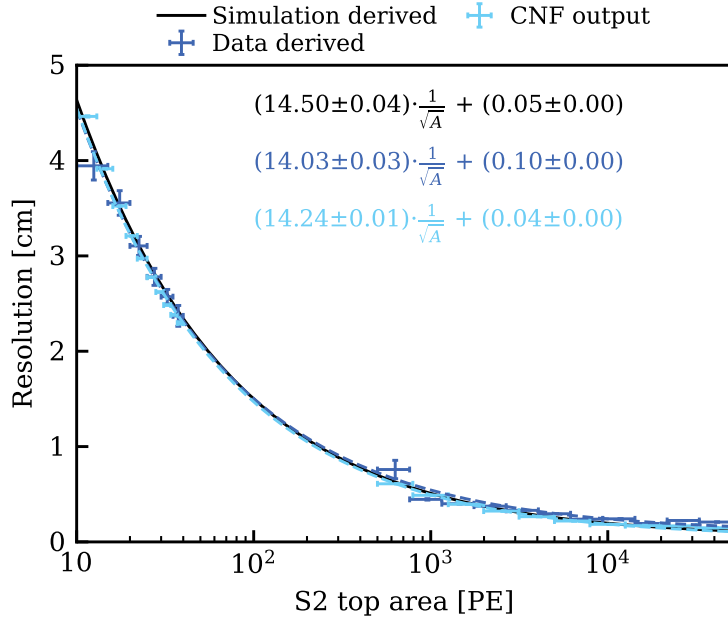


Figure 3.30: A comparison of the simulation derived (black), data derived (dark blue) and CNF predicted (light blue) position resolution as a function of the S2 top area. In addition, an empirical fit for each is given.

3.6.2 Fuzzy Selection of Perpendicular Wire Events

The perpendicular wires supporting the gate and anode described in [subsection 2.3.2](#) influence the formation of S2s in their immediate vicinity. To correctly apply signal corrections and certain cuts in XENONnT, it is necessary to know whether or not an event was formed in the *near wire* or *far wire* region. Before the introduction of the CNF, the position resolution was only available as a statistical measure, stating that on average an S2 of a certain top area will have a given uncertainty on its position. Using this, the near wire region was identified as a region around the wires with an area dependent width, according to the mean position resolution. In SR2, the position uncertainty is available for each individual S2. Ivy Li developed a near wire definition utilizing this information, where an event belongs to the near wire region if its own uncertainty contour is overlapping with a perpendicular wire.

[Figure 3.31](#) shows the selection of the far wire region using the definition based on the mean resolution (left) and the probabilistic definition based on each events individual position uncertainty (right) for 100 SR2 background runs. The near wire region is more difficult to model and is in some cases even removed completely from the analysis space, meaning a smaller near wire region is better in general. From [Figure 3.31](#) it is visible that the probabilistic selection allows more events to remain in the far wire region.

Furthermore, using the probabilistic definition of the near wire region, it is possible to identify some S2s with large area, that originate from in between the wires, but for which it is quite certain that they are not distorted by the perpendicular wires influence.

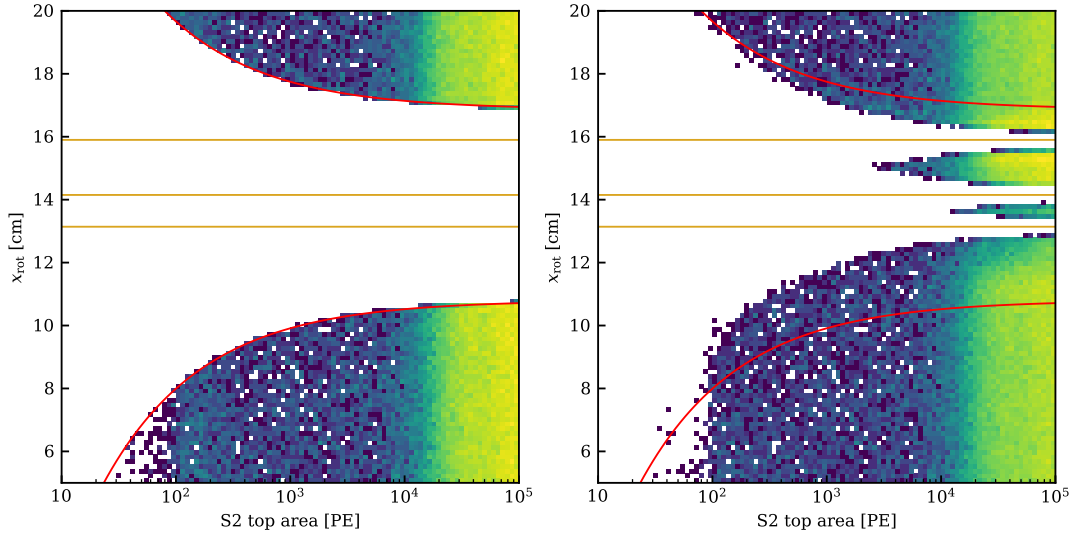


Figure 3.31: The distribution of far wire events in S2 top area and rotated x coordinate. The far wire region is selected via the average position resolution (left) and via the event individual position uncertainty (right). In addition, the boundaries defined by the mean resolution are shown in red. The positions of the perpendicular wires are shown in golden.

3.6.3 Position Uncertainty Cut

Figure 3.30 shows that the predicted position uncertainty of the CNF can be modeled well with Equation 3.18. It is possible to use the position uncertainty as a measure of signal quality, under the assumption that if an S2 has a significantly higher position uncertainty than expected from its top area, its hitpattern is hard to predict for the CNF and therefore contains some unexpected anomaly. Anomalous hitpatterns can be produced for example when a spurious uncorrelated signal gets merged into the S2, and pose a background population that should be removed from the analysis data via a cut.

Figure 3.32 shows the distribution of CNF predicted position uncertainty for simulations (left) and data from ^{220}Rn and ^{37}Ar calibration runs (right). The simulations were produced using the same settings as the CNF training set, therefore it is expected that all of those are standard hitpatterns from the perspective of the CNF model. The red solid line in both shows the CNF predicted position resolution (see Figure 3.30), and the red dashed line shows a power-law fit to the 99 % percentile of the position uncertainty of the simulated data that serves as the boundary of the position uncertainty cut.

The red lines in the right panel of Figure 3.32 correspond to the ones in the left panel. As opposed to the simulations, the calibration data shows more events with an uncertainty greater than the cut boundary. As each cut used in analysis potentially decreases the signal acceptance, it is important to evaluate new cuts on their potential to remove unwanted events that are not target of an already existing cut. For that reason, all the cuts listed in Table 2.3 were already applied to the calibration data in Figure 3.32, and in addition an ER band selection was applied, meaning only events were selected where the S1 and S2 strength match a distribution that is expected from ER events. In

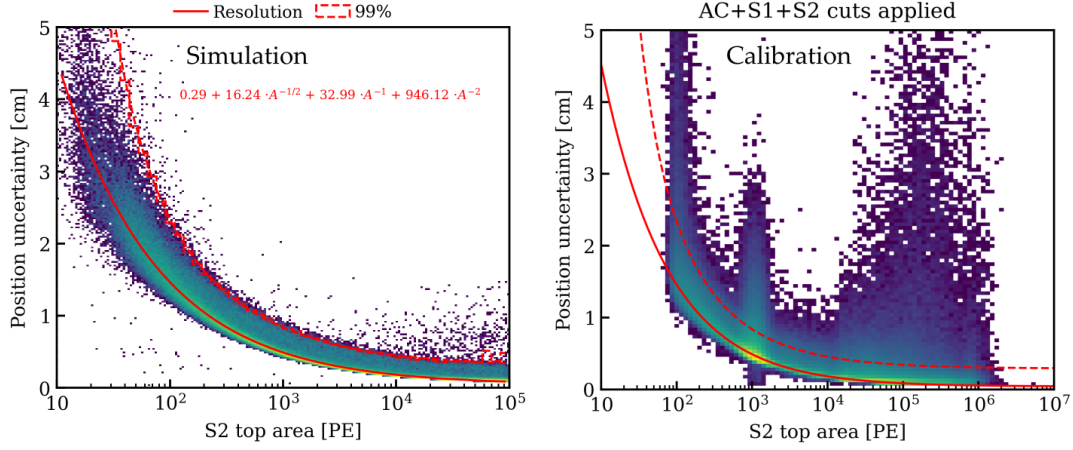


Figure 3.32: Distribution of position uncertainty and S2 top area for simulations (left) and calibration data (right). The red solid line is the CNF based empirical resolution function, shown in Figure 3.30. The red dashed line is a fit to the 99 % percentile of the position uncertainty of the simulated data.

total there are $\sim 2.3 \times 10^7$ calibration events left, of which $\sim 6.5 \times 10^4$ have a position uncertainty greater than the cut boundary, resulting in a cut acceptance of 99.7 %.

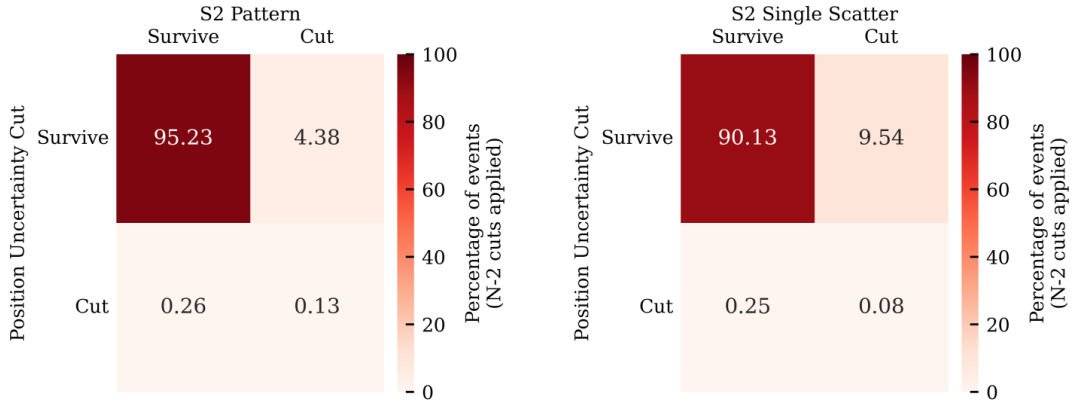


Figure 3.33: The correlation matrices between the position uncertainty cut and the S2 pattern cut (left) and the S2 single scatter cut (right). The position uncertainty cut is nearly uncorrelated to both, since the population that is cut by both at the same time, the lower right square, is the smallest of all.

The position uncertainty cut targets anomalous S2 hitpatterns, which are also potentially removed by the already existing S2 pattern cut. To maximize signal acceptance it is important that the number of cuts is kept low and specifically that no two cuts target the same background population. The left hand of Figure 3.33 shows the correlation between the S2 pattern cut and the position uncertainty cut, evaluated on the calibration data where all other cuts ($N - 2$) were already applied beforehand. Surprisingly, the population of events that is rejected by both cuts, the lower right entry of the matrix, is the smallest population. In particular, it is smaller than the populations rejected by one but accepted by the other cut, which indicates that these two cuts are not highly correlated at all. Another way of thinking about the position uncertainty cut is to investigate at the mechanisms behind anomalous S2 hitpatterns, one of which

is MS events leading to multiple energy deposits in the TPC. The right hand side of [Figure 3.33](#) shows the correlation matrix between the position uncertainty cut and the S2 single scatter cut. Once more, both cuts show nearly no correlation. However, in this case it is easier to explain, as the S2 single scatter cut targets MS events where there are two separate S2 peaks, while the position uncertainty cut would target MS where the two different S2s got merged into a single peak.

The spatial distribution of events rejected by the position uncertainty cut can be seen in [Figure 3.34](#): The left panel shows the horizontal distribution, including PMTs that are turned off (red circles) or display a higher than usual rate of afterpulsing (violet). The rejected events are distributed in clusters, but those clusters do not seem to correlate to special PMTs, but rather follow certain spots at the edge of the TPC or lines in-between the PMTs. From the distribution on the right in vertical depth and squared radius it can be inferred that most of the rejected events originate from the very top or very bottom of the TPC. These events could originate from material radioactivity, or be accidental pairings of lone S1s with electron emission from the electrode grids. Furthermore, in regions where there is a higher general activity, such as close to the detector materials, there is a higher chance of two uncorrelated S2s overlapping and being merged. Finally, since it is not yet well known which exact population would be removed by this position uncertainty cut, and because the low impact of the cut, based on only 0.3 % of all events being removed, the position uncertainty cut was not applied in XENONnT SR2.

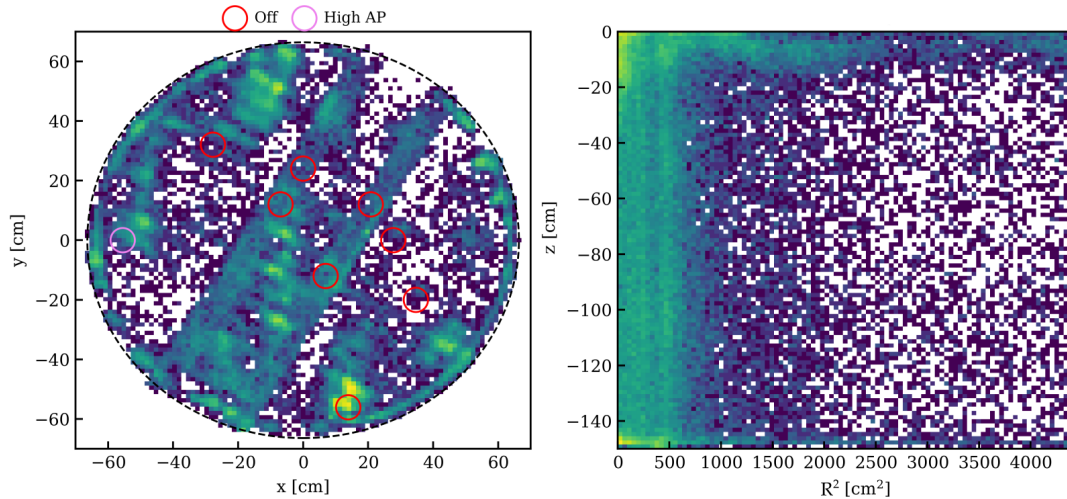


Figure 3.34: The spatial distribution of events rejected by the position uncertainty cut. The circles on the left indicate PMTs that are turned off (red) or show a high after-pulse rate (pink).

3.6.4 Multi-Scatter Detection

In the same vein as the position uncertainty cut from [subsection 3.6.3](#) it is possible to specifically try to identify hit patterns merged from two independent S2s via their predicted position uncertainty. For this, artificial merged MS events are created by summing two different S2s from ^{83m}Kr calibration runs chosen at random. Since the ^{83m}Kr decay is monoenergetic, this study focuses on events where the individual fragments making up the MS are approximately equal in size. The advantage of using ^{83m}Kr data is that those events are distributed uniformly in the TPC, making them

suitable for a demonstration of the MS detection efficiency as a function of the distance between the fragments.

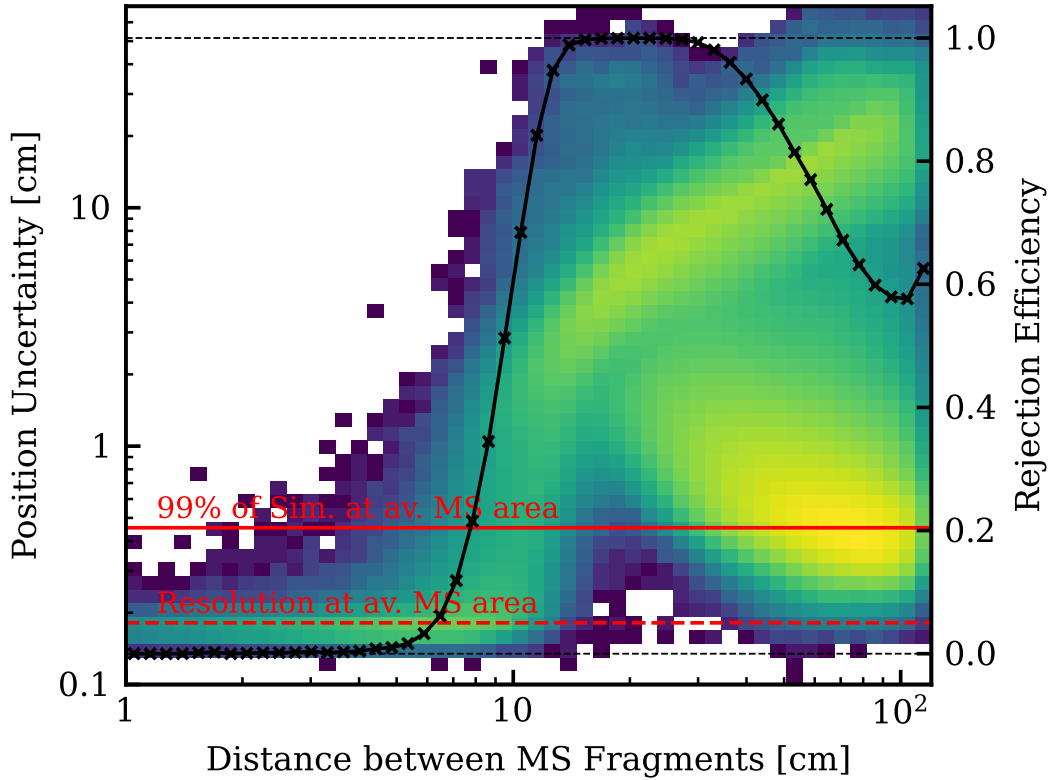


Figure 3.35: The distribution of position uncertainty and distance between individual fragments making up a fake MS event. The resolution derived from the CNF output on data and 99th percentile of position uncertainty derived from simulated hitpatterns are shown in red dashed and red solid respectively. Using the 99th percentile, a MS rejection efficiency can be computed, shown in black.

Figure 3.35 shows the distribution of the position uncertainty as a function of the horizontal distance between the individual ^{83m}Kr events that make up the fake MS, using the data from 500 ^{83m}Kr calibration runs. Below ~ 5 cm separation, there is no increase in position uncertainty and between 5 cm to 10 cm the position uncertainty rises rapidly. But, at around 20 cm the data seems to split into two populations: For some MS events, the position uncertainty increases vastly, far above the range that is typical for S2s of this area. However, for a larger fraction of the fake MS the position uncertainty decreases again, nearly down to normal position resolution levels. This behavior is potentially caused by the artificial MS hitpatterns that were injected into the CNF training set as described in subsection 3.4.2, which was done with the intention of teaching the CNF that bigger uncertainties are necessary for some events. A possible future study could investigate whether tuning of the CNF training procedure can influence the behavior observed here.

In addition Figure 3.35 shows the expected resolution at the average area of the fake MS (red dashed) and the 99th percentile boundary that was also used in subsection 3.6.3 (red solid). If events with greater uncertainty than this boundary are rejected, it is possible to compute a fragment distance dependent MS rejection power of

the position uncertainty, shown in black. From about 10 cm to approximately 30 cm of fragment separation there is an extremely high chance to detect the event as a MS, which drops afterwards again down to around 60 %.

3.6.5 Transversal Electron Diffusion

In reality, an S2 signal is not produced at a single point in the horizontal plane. The ionization electrons from the interaction site are produced in a spatially spread out cloud and were subject to diffusion effects during their drift through the TPC. The diffusion of electrons in LXe can be characterized by two constants, the longitudinal, along the applied electrical field, diffusion coefficient D_L and the transversal, perpendicular to the drift field, diffusion coefficient D_T . There are only few measurements of the transversal diffusion coefficient D_T , one of which from the EXO200 collaboration [156] is shown in Figure 3.36.

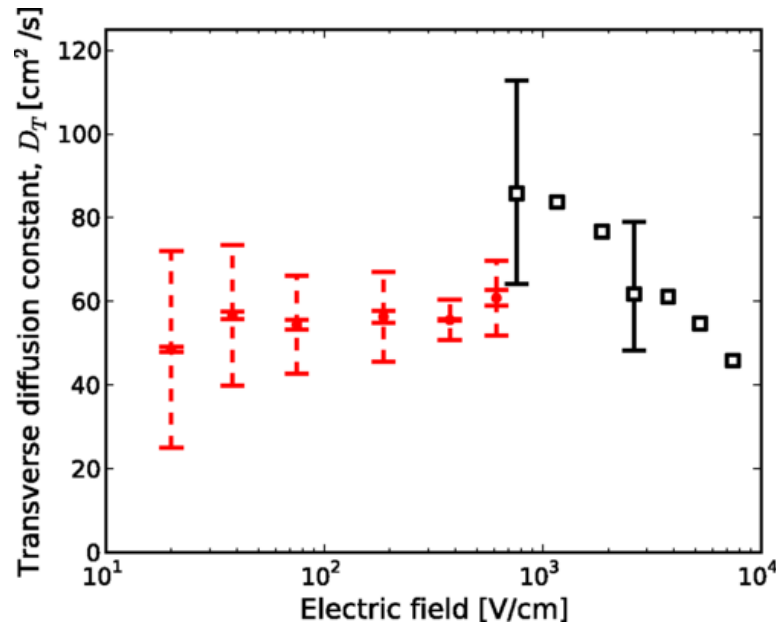


Figure 3.36: Measurements of the transversal diffusion constant, figure taken from [156]. Red data points show the measurements from the EXO200 experiment [156] and black data points show the measurements from [157].

Interpolating the EXO200 values to the drift field used in XENONnT SR2 yields $D_T(E_{\text{drift}} \approx 23 \text{ V/cm}) = 49.9 \text{ cm}^2/\text{s}$. However, the sparsity of available measurements and the large uncertainties of the available ones, shows that an independent measurement of the transversal diffusion in XENONnT would be of interest.

In [156] the transversal variance of the position caused by transversal drift is given as

$$\sigma_D^2 := \langle x^2 \rangle + \langle y^2 \rangle = 4D_T t_{\text{drift}}, \quad (3.24)$$

where t_{drift} is the drift time of the ionization electrons. This can be extended to include the variance of the predicted position of a single electron, σ_{SE} , to get the horizontal spread after reconstruction of the S2 produced by an electron cloud $\sigma_r = \sqrt{\sigma_D^2 + \sigma_{\text{SE}}^2}$. The CNF however does not predict the width of the cloud that generated the S2 signal. Rather it predicts the uncertainty on the predicted position, which can be read

as a mean position of the electron cloud containing n_e electrons. Therefore, the position uncertainty predicted by the CNF σ_{CNF} is the standard error on the mean of this predicted electron cloud spread

$$\sigma_{\text{CNF}} = \frac{\sigma_r}{\sqrt{n_e}} = \sqrt{\frac{4D_T t_{\text{drift}} + \sigma_{\text{SE}}^2}{n_e}}. \quad (3.25)$$

Using this equation it is possible to fit the CNF position uncertainty of data, leaving D_T and σ_{SE} as free parameters. However, there are three more things to note: First, Equation 3.25 takes D_T to be constant, which is only true for a constant electric field. In the XENONnT TPC there are small variations in field strength, especially near the walls and electrodes. For that reason this study only considers data from an inner sub volume, using a maximum radius of $R < 50$ cm and a minimum and maximum height of $-120 \text{ cm} < z < -20 \text{ cm}$. Second, impurities dissolved in the LXe can potentially influence the diffusion behavior of the electrons, which is why the data taken for this study comes from a single $^{83\text{m}}\text{Kr}$ calibration campaign, meaning calibration runs taken in immediate succession, with therefore similar levels of impurity concentration. Finally, the position of the light emission by the electrons in the gas phase is constrained by the geometry of the electrodes themselves. The electrons are funneled into fine bands between the gate wires, making the position perpendicular to the gate wires independent of the electron diffusion [126, 154]. The position parallel to the gate wires however is not constrained in this way. For this reason, σ_{CNF} is taken to be the predicted position uncertainty in x_{rot} direction.

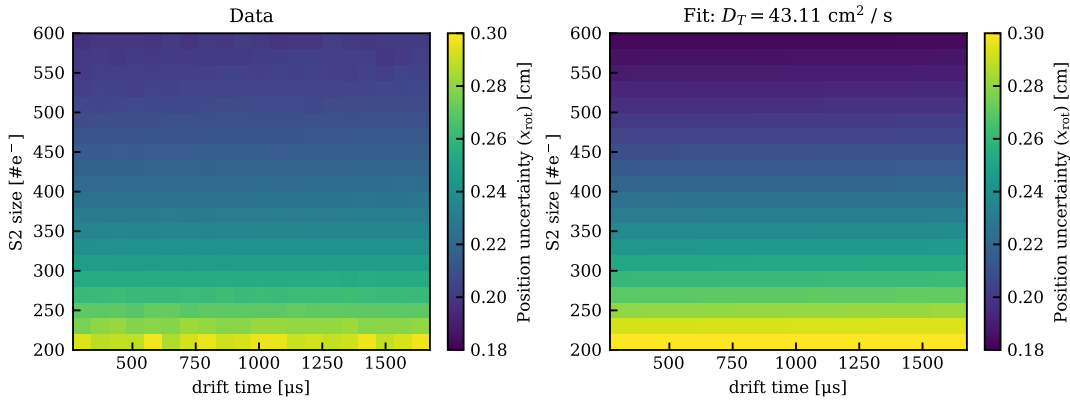


Figure 3.37: The average position uncertainty as a function of drift time and S2 size for $^{83\text{m}}\text{Kr}$ calibration data (left) and fitted using Equation 3.25 (right).

Figure 3.37 shows the distribution of the $^{83\text{m}}\text{Kr}$ calibration data on the left and the fit result on the right. The best fit values are $\sigma_{\text{SE}} = 4.42 \text{ cm}$ and $D_T = 43.11 \text{ cm}^2/\text{s}$. Computing the expected resolution of a single electron using the fit function of the CNF output from Figure 3.30, with an S2 top area approximated from a preliminary g_2 value and average S2 AFT value as $A_{\text{top, SE}} = 10.6 \text{ PE}$, results in $\sigma_{\text{SE}} = 4.51 \text{ cm}$. The best fit value for σ_{SE} can therefore be taken as quite accurate. However, Figure 3.37 clearly shows a much stricter dependence on n_e than on t_{drift} . To get a reliable measurement of D_T , it is recommended to thoroughly investigate all potential sources of systematic uncertainty, while the study presented here serves more as a proof of prin-

ciple. Further, Equation 3.25 can be expanded to take into account additional effects, like the coulomb repulsion among the electrons during the drift [158].

3.7 Conclusion and Outlook

This chapter outlines the position reconstruction process of XENONnT, for which various ML models are employed that take the hitpattern an S2 signal leaves in the top PMT array and provide a prediction for the horizontal position (x, y) . It is shown that these models can be extended in two ways to change their point-like output into a PDF that spans the full horizontal detector plane, by using a One Hot (OH) encoding, or by exchanging the loss function with a predefined analytical likelihood function. It is to note that the purpose of turning the position reconstruction probabilistic is *not* the improvement of the accuracy, as that is already near the optimum dictated by geometry, but rather to supply additional uncertainty information without degrading the performance.

The OH extension separates the horizontal plane into bins and treats them as distinct classes. The predicted chance of an S2 belonging to a class can be interpreted as the chance of the S2 having originated at the position of that bin. The parameterized extension replaces the point output of the ML model with the parameters of an analytical PDF. To correctly learn the parameters, the loss function of the model must incorporate the likelihood function of the chosen PDF.

Although the OH model has the great advantage of being agnostic to the actual underlying distribution of the position uncertainty, it increases the necessary parameters of the ML model by a vast amount, which poses a problem due to overtraining. A combination of both approaches is presented: A high resolution OH model with a bin width of 2.5 mm is trained and used to predict the position probability distribution of a set of simulated S2 hitpatterns. These probability distributions are fitted with different analytical PDFs, with the aim of determining that analytical function which best approximates the true distribution. The best matching PDF is a bivariate Skew Gaussian distribution.

The Skew Gauss (SG) model achieves slight improvements in resolution while providing not only the position but also the full probability distribution of each S2. The predicted position uncertainty is evaluated in a coverage study and matches the true uncertainty with a maximum of $\sim 5\%$ overcoverage.

In addition to the NN models of SR0 and SR1, a novel Conditional Normalizing Flow (CNF) model is introduced for SR2 of XENONnT. The CNF works by mapping a two dimensional standard normal to the unknown conditional probability distribution of the S2 positions given observed hitpatterns, via bivariate mappings, called flows. This CNF model manages to outperform the conventional NNs, in particular at the edge of the TPC. In addition, the predicted uncertainty of the CNF achieves nearly perfect coverage.

The simulation trained position reconstruction models are validated on measured de-

tector data. For this purpose two signal populations are identified: SE S2s from the so-called hotspot serve to validate the position resolution at low S2 top areas, and a fine band of S2s originating from in between the perpendicular wires for validation at higher S2 top areas. Using these data it is shown that the simulation based resolution and the uncertainty output of the CNF are accurate within any degree necessary for the analysis in XENONnT.

The CNF model has become the default position reconstruction algorithm for SR2 and beyond. The additional position uncertainty it provides is already employed in the S2 merging procedure. Further, it can be used to transform spatial selections from hard boundaries into selection with fuzzy edges that work on an event to event basis. This is demonstrated on the near and far wire distinction, but can in principle also be applied for example to the FV selection.

In the future, the position uncertainty provided by the CNF model can serve as a measure of S2 signal quality and as a way of detecting MS events. Proof of principle studies are shown for these applications. Furthermore, position uncertainty specific problems can be investigated, such as the spatial spread of electron clouds that make up the S2 signals.

Quality Assurance of the XENONnT Upgrade Cathode

This chapter presents the evaluation of the new cathode of the XENONnT detector in regard to the repair campaign performed in 2025, and the development of electrodes for even bigger experiments like XLZD. The work presented here is was published in [2]. The general importance of electrodes in LXe TPCs and the challenges associated with the the XENONnT cathode are detailed in [section 4.1](#). An introduction to basic computer vision (CV) concepts is found in [section 4.2](#), as well as a description on how to use CV and ML methods to extract millimeter scale images from larger images of the cathode. In [section 4.3](#) the concept of anomaly search using Variational Autoencoders (VAEs) is introduced and multiple VAEs are evaluated and compared. Finally, [section 4.4](#) describes high voltage (HV) tests of the XENONnT upgrade cathode performed at KIT.

4.1 The Upgrade Cathode of XENONnT

The electrodes cathode, gate, and anode, play an important role in the signal formation of the XENONnT detector. The anode and gate define the extraction field E_{ext} , the electric field at the top of the TPC that serves the purpose of extracting electrons into the gas phase and accelerating them, which leads to the secondary scintillation that makes up the S2 signal. To ensure 100 % electron extraction, the minimum requirement for this field is $E_{\text{ext}} \approx 10 \text{ kV/cm}$ [159]. The ability to register S2 signals at lowest energies is directly dependent on the ability to extract electrons into the gas phase. Further, to ensure signal uniformity, the extraction field should be as homogeneous as possible. As mentioned in [subsection 3.5.2](#) and [subsection 3.6.2](#) features, such as the perpendicular wires in XENONnT, influence the S2 signal shape and pose challenges for signal modeling.

The drift field E_{drift} , formed by the gate and cathode, is also a vital part of a LXe TPC. As presented in [subsection 2.3.3](#), it influences the probability of recombination of electron ion pairs after an energy deposit, and the drift of the ionization electrons

to the top of the TPC. Electric field lines that end in a TPC wall lead to absorption of the S2 electrons, effectively blinding a region of the detector by forming a charge insensitive volume (CIV). A too low drift field in particular can lead to problems like sparsity of the S2 forming electron cloud, a challenge also mentioned in [subsection 3.6.1](#). The designed and realized drift fields for a selection of LXe dual-phase TPC experiments are shown in [Figure 4.1](#). None of the listed experiments managed to achieve their nominal field strength, and had to tune the field down by a wide margin, which exemplifies apparent difficulties in the realization of electric fields, and the design of the electrodes generating them. An in-depth description of the requirements and accompanying challenges that have to be overcome in electrode design is detailed in [\[2\]](#) and [\[160\]](#).

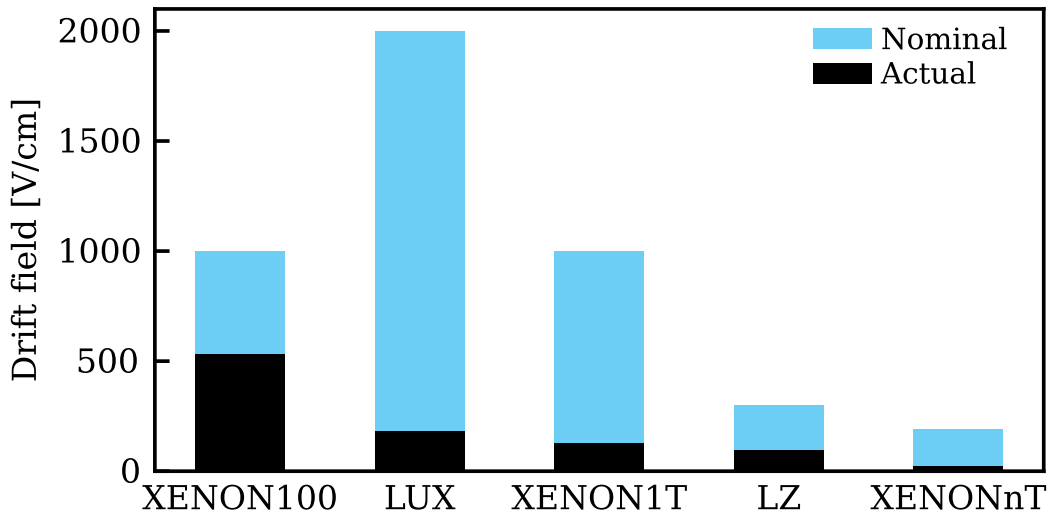


Figure 4.1: The designed (blue) and achieved (black) fields of different LXe TPC experiments. Inspired from [\[160\]](#). Data from: XENON100: [\[161\]](#), LUX: [\[162, 163\]](#), XENON1T: [\[164, 165\]](#), LZ: [\[64, 80\]](#), XENONnT: [\[79\]](#)

As shown in [Figure 4.1](#), XENONnT had a design field of 191 V/cm, but in the end had to operate in SR0, SR1 and SR2 with $E_{\text{drift}} = 23$ V/cm, due to a short circuit between the cathode and the bottom screening electrode. In 2025, the XENONnT detector underwent an upgrade, whose main goal was the replacement of its electrodes. The characterization of the replacement electrodes is detailed in [\[2\]](#), and for the replacement cathode in particular in the following sections. In contrast to the previously installed parallel wire cathode, the replacement cathode is a hexagonal mesh, as is also used in other TPCs [\[161, 166\]](#), manufactured by PCM Products, Inc. [\[167\]](#) via photochemical etching. A sketch of the mesh is shown in [Figure 4.2](#). The mesh was produced as two separate half-meshes that were laser welded together along a welding seam of width 4 mm. The mesh is encompassed by two stretching rings that are used to tension the mesh during its installation, the outer of which is mechanically cut off after its tensioning and fixation of the electrode frame, indicated by the red line. Both rings consist of small segments connected via omega springs to counteract deformation during stretching, circled in black in [Figure 4.2](#). The full mesh consists of approximately 26200 hexagonal *cells*, given as approximate because cells at the welding seam and the stretching ring are incomplete, and it is not clear whether or not

they should contribute to the count. Each cell has an inner diameter, called pitch, of $p_{\text{hex}} = 7.5 \text{ mm}$, and each edge of a hexagon, called *leg*, has a width of $w_{\text{hex}} = 0.3 \text{ mm}$.

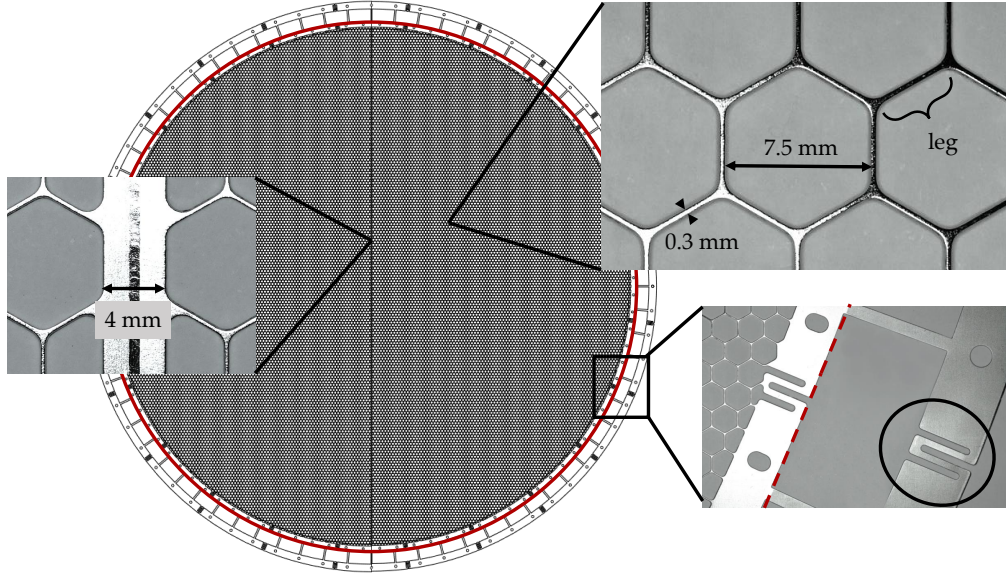


Figure 4.2: The XENONnT replacement cathode mesh from [2]. The mesh is made from two half-meshes, laser welded together on a 4 mm wide welding seam (left inlet). The individual hexagonal cells have a pitch of $p_{\text{hex}} = 7.5 \text{ mm}$ and a leg width of $w_{\text{hex}} = 0.3 \text{ mm}$ (top right inlet). In addition, the mesh is enclosed in an inner and an outer stretching ring (bottom right inlet), the outer of which is supposed to be cut off after installation (red line).

As seen in Figure 4.1, many other experiments that did not experience a short circuit also had to lower the electrical field, the reason for which is an increase in observed few electron signals that comes with increasing voltages [64, 115, 161, 168]. This increase can be explained by spurious electron and photon emission from the electrodes themselves. One potential source of spurious electrons and photons is a local amplification of the electrical field caused by sharp surfaces on the cathode. The mesh itself is electropolished to reduce the potential for electron emission [169], but as every physical component it is susceptible to damage during handling or fabrication, or may have innate anomalous features. Such *defects* can limit the performance of the electrode and therefore should be identified and where possible repaired.

The replacement cathode consists of approximately $n_{\text{cells}} \approx 26000$ individual hexagonal cells, which corresponds to $n_{\text{legs}} \approx 80000$ independent legs, and $n_{\text{surf}} = 160000$ leg surfaces. Each such surface could be damaged, or come produced featuring a defect, and potentially deteriorate the performance of the cathode mesh. Examples of observed defects are shown in Figure 4.3, where a leg with unwanted extra material, which often comes in the form of pointy spikes (a), a leg weakened by surface damage (b) and a fully broken leg with a sharp edge can be seen.

The individual legs of the cathode have a thickness of 0.3 mm, making the inspection by eye extremely tedious and prone to human error. Future detectors are planned to be even larger, making a manual inspection impossible [84]. For this reason, a set of 336 images of the cathode, including front and back side, was taken using a NIKON

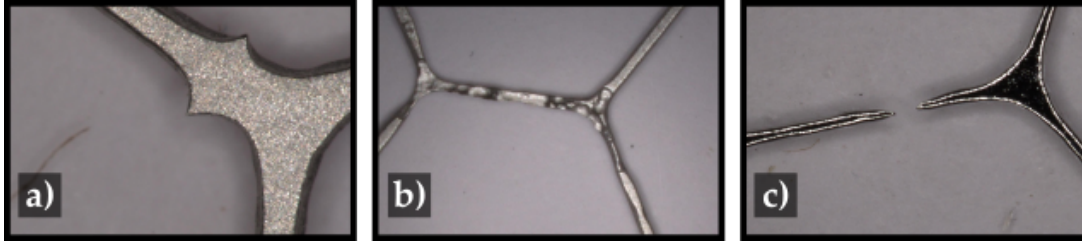


Figure 4.3: Examples of observed defects of the cathode mesh: Extra material, or spikes (a), surface damage (b) and a broken leg (c).

Z50 digital camera, with the goal of identifying defects using computer vision (CV) and ML methods. An example of such an image is shown in [Figure 4.4](#), which also showcases the printed rectangular grid that was used for orientation.



Figure 4.4: An example of one of the 336 base images taken of the cathode mesh to ease the inspection.

Under the reasonable assumption that defects are rare, one can aim to construct some statistical measure q , that represents the normal-ness of an image. If constructed correctly, q will have a somewhat tractable distribution for all non defect images, while a defect will appear as an outlier, making it possible to define every image with q greater than some threshold as an *anomaly*. Anomaly search is a common application of CV in different industry and research settings. With the inflation of data availability and computing power in recent years, this is a task that is largely dominated by Machine Learning (ML) methods [[170–175](#)].

4.2 Leg Segmentation

The *base images*, such as [Figure 4.4](#), are far too large in size and dissimilar to each other to perform the anomaly search directly on them. In addition, 336 samples are not enough to properly train an ML model capable of defect detection. For this reason, smaller *leg images*, each featuring only one single hexagon leg, were extracted from the large base images, the procedure of which is described in this section. The advantages using smaller images are:

1. **Size reduction:** The base images have a size of $(5568 \times 3712) > 2 \times 10^7$ pixels. Any ML model with an input this large will have an excessive number of parameters, which is problematic as explained in [section 3.2](#), especially when the amount of training data is limited by the number of available meshes.
2. **Background removal:** The majority of pixels in [Figure 4.4](#) feature the printed rectangular grid background instead of the actual mesh. Those pixels do not carry information¹ about the defect-ness of a leg, and should therefore, if removed, not deteriorate the ML models performance. On the contrary, removing background pixels should increase the models ability to learn the distribution of mesh pixels, and make it so that a potential defect covers a greater fraction of one image.
3. **Adjustability:** As seen in [Figure 4.4](#), the lighting, focus and background differs in different regions of a single base image. With one large image, this is more difficult to counteract than it is for individual small images.
4. **Standard normal-ness:** The base images differ quite a lot among themselves, especially regarding images containing the welding seam or stretching rings, which constitute much larger differences than small defects do.

Therefore, the base images were cut down into smaller leg images. However, this task was not trivial, since the base images had different lighting conditions, different positioning of the mesh, and via parallax effects, the mesh pixels do not form perfectly regular hexagons.

4.2.1 A short Introduction to Computer Vision

A digital image can be represented by a tensor of 3rd order with size $w \times h \times c$, where w and h are the width and height of the image in pixels, and c is the number of *channels*. It is common to refer to images as having a size $w \times h$, omitting the information about the channels, as this is the size of the image in *pixels*. Each pixel is a c -dimensional vector. A standard digital image has either $c = 4$ channels in red green blue α (RGBA) representation, $c = 3$ channels in red green blue (RGB) representation or $c = 1$ in grayscale representation. The α in RGBA is the opacity value, a decimal number $\alpha \in [0, 1]$ where $\alpha = 0$ corresponds to a transparent image. An RGB image

¹A background pixel in a position where one would expect a mesh pixel does indicate for example a broken leg. Them carrying no information is therefore not fully correct.

is the same as an RGBA image with $\alpha = 1$. The red, green and blue values of a pixel are unsigned integers with values between 0 and 255, inclusive. A grayscale image has only one value y per pixel, which is a decimal number between 0 and 1, where 1 represents white and 0 represents black. To convert RGB values r, g, b to grayscale, this work uses the weighted average method, also called the luminance method [176],

$$y(r, g, b) = 0.299 \cdot r + 0.587 \cdot g + 0.114 \cdot b, \quad (4.1)$$

where the weights of each color come from the human perception of color. The images considered in this work are either RGB or grayscale, as the α channel offers no use to the task at hand. A *binary mask* of an image is a matrix of dimension $w \times h$, where each entry is either 0 or 1. Binary masks are used to select subsets of all the pixels of an image.

Image Preprocessing Methods

The digital camera base images are subject to imperfections, for example from varying lighting conditions or thermic noise. A standard way to handle such nuisances, that can potentially hinder the performance of an ML model, is via image preprocessing. An in-depth description of image processing and CV can be found for example in [177]. Basic examples of image preprocessing are shown in Figure 4.5, where a leg image is shown in each top panel to visualize the effect on the perceived image, while the distributions of RGB values are shown in the bottom panels to visualize the effect on the actual numbers that represent the image. The original leg image was extracted from the base image using the method described in subsection 4.2.2.

The most basic form of preprocessing in image recognition tasks is *scaling* of the pixel values by some factor, which can be a constant factor for all images, an image-dependent factor, or even a variable factor depending on the individual pixel and its environment inside the image. Two common scalings in ML are zero-one scaling and standard scaling. In zero-one scaling each pixel of an image is scaled such, that the maximum value of the image is 1 and the minimum value of the image is zero, while retaining relative distances between the pixel values. In standard scaling the pixel values are shifted by the mean and divided by the standard deviation, which leads to the new mean being zero and the new standard deviation being one. Such scalings make it easier for ML models to distinguish differences in pixel values, which becomes clear after looking at the typical domain of common activation functions, which lies close to 0 (see section 3.2). Figure 4.5 shows no examples of scaling, as scaling does not influence the appearance of an image.

Other preprocessing methods common in CV tasks are adjustment of the total color scale, brightening or darkening the image, or adjustment of the gamma factor. Brightening or darkening adds or subtracts a constant to (parts of) the image, while keeping the minimum and maximum values capped at 0 and 255. In general, this should be done before scaling the image. The gamma factor is a factor that captures the relation between pixel values as seen by computers and cameras, and perceived brightness values as seen by human eyes. Examples of brightening, darkening and gamma adjustment are shown in Figure 4.5b, Figure 4.5c and Figure 4.5d respectively.

A basic concept in image processing is kernel convolution, also called *filter* or *mask* operation. A kernel is a small matrix H that is convolved with the image F . Since the pixels of an image pose discrete coordinates (x, y) , the convolution is also a discrete operation

$$G_{x,y} = H * F_{x,y} = \sum_{i=-a/2}^{a/2} \sum_{j=-b/2}^{b/2} H_{i,j} \cdot F_{x-i,y-j}, \quad (4.2)$$

where $G_{x,y}$ is the resulting pixel at (x, y) and a and b are the kernels width and height respectively. For more information see for example [177, 178].

A big factor in CV is is dealing with the noisiness of the acquired images and a standard way of reducing general white noise is a Gaussian filter, also called a *blur* or *smoothing* filter. A Gaussian filter is a kernel whose entries follow a bivariate normal distribution centered on the respective pixel, characterized by the standard deviation σ . This operation smooths out the values of the image, scaling each pixel by an average weighted on its immediate environment, blurring the image, as can be seen in Figure 4.5e. While this Gaussian filter smooths out white noise very well, applying too much of it can potentially hide or smear out important features of an image.

A reverse of this process, the exaggerating of differences in neighboring pixels, can be achieved with a sharpening filter, that involves a matrix of the general form

$$M_{\text{sharp}} = \begin{pmatrix} 0 & -a & 0 \\ -b & c & -b \\ 0 & -a & 0 \end{pmatrix}, \quad (4.3)$$

where $a, b, c > 0$ are constants chosen according to the desired strength of the sharpening effect. The values of a and b can be different if the effect is desired to differ in vertical and horizontal directions. The effect of a sharpening kernel with $a = b = 1$ and $c = 5$ is showcased in Figure 4.5f. Many more filters and their implementations in python can be found in the scikit-image library [179].

Morphological operations are special kernel operations performed on binary masks. The kernels for morphological operations, called the *structuring elements*, are binary masks themselves and govern the nature of the transformation performed by the operation. An in-depth formulation of morphological operations can be found in [180], while here only the effects of specific operations on binary masks are described. An *erosion* can be understood as the intersection of the binary mask and the structuring element, where resulting pixels are only equal 1 if they are one in both the mask and in the structuring element. In effect, this erodes the width of structures in the binary mask. A *dilation* is the opposite, similar to a union of the binary mask and the structuring element, that effectively widens all 1-like structures of a binary mask. An *opening* is erosion and consecutive dilation of a mask, which has the effect of removing 1-like outliers not surrounded by a greater structure. A *closing* is the inverse order, dilation followed by erosion, that removes 0-like outliers in bigger structures of a binary mask. In general, the extent of a morphological operation can be influenced by the size and shape of the structuring element. Implementations of morphological operations can be found for example in OpenCV [181].

Data Augmentation

Data augmentation is a concept that is similar, but separate from data processing. In ML, data augmentation refers to the practice of generating new data from existing data, which is done for example when there is a lack of data of one specific category. In the use case here, there are by far fewer (known) legs with defects, than there are non-defect, clean, legs.

Additional images generated via image augmentation techniques should have the same general properties as the original images. More specifically this means if the original images went through some preprocessing routine, the augmented images should go through the same, and in particular, a processed image can not be used to increase the pool of non processed images. If we train the defect detection on 160000 images of non processed legs, and 1000 additional images of legs convolved with a Gaussian filter, the ML model has a high chance of detecting the processed images as anomalous. Data augmentation methods must therefore preserve the features that are used to generate the statistical anomaly measure q .

Example procedures that preserve the pixel value domain of the original images are mirroring or translating of an image, as they maintain the numerical values of the original but assign them new horizontal and vertical positions. The symmetry of individual leg images allows us to easily increase the pool image samples by a factor of 4 via mirroring (original, horizontal mirror, vertical mirror, both directions at once). Translating of images that need to keep their dimensions is generally done through rolling, which is the practice of stitching pixels removed at one edge of the image to the opposite edge of the image. Since this can create visuals that look like broken mesh legs, a feature explicitly searched for, there will be no data augmentation via translation here.

The low number of known defects makes it hard to evaluate the performance of defect detection models in a statistically significant manner. To alleviate this, there is also the possibility of manually augmenting images, by introducing either defects from other images into them, or by making anomalous by drawing on them or distorting them.

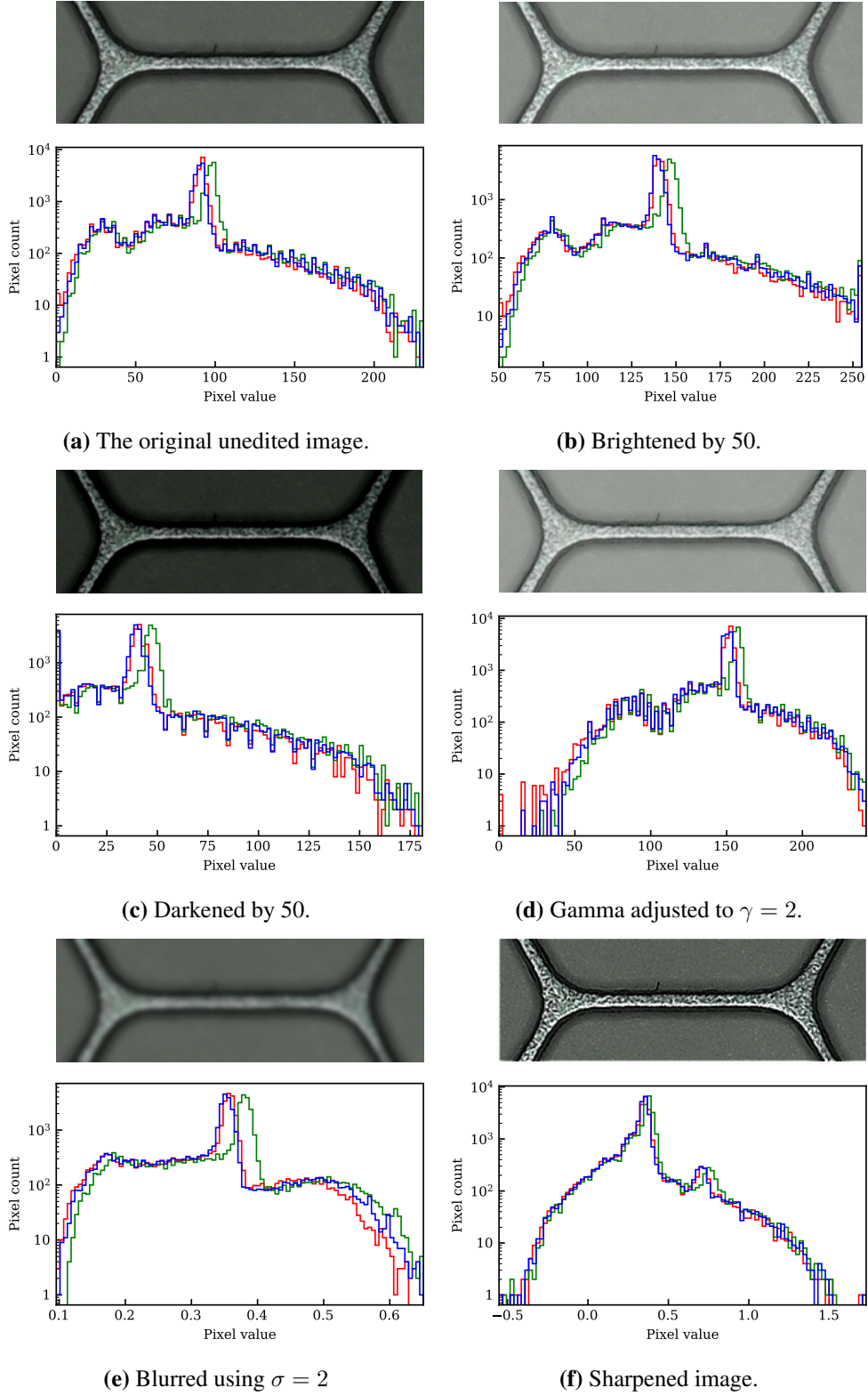


Figure 4.5: Examples of common image processing methods. The top panel shows processed picture, while the bottom panel shows the distribution of RGB color intensity in their respective color.

4.2.2 Machine Learning based Pixel Segmentation

To cut the leg images out of the base images it is necessary to know whether a given pixel is part of the imaged electrode mesh, a *mesh pixel*, or part of the background, a *background pixel*. The act of partitioning an image into distinct regions is called *segmentation* [178, 182]. To segment an image using ML a training data set is required. The training samples are pixels of one sample image, labeled as either mesh pixel or background pixel using the LabelMe software [183], shown in Figure 4.6.

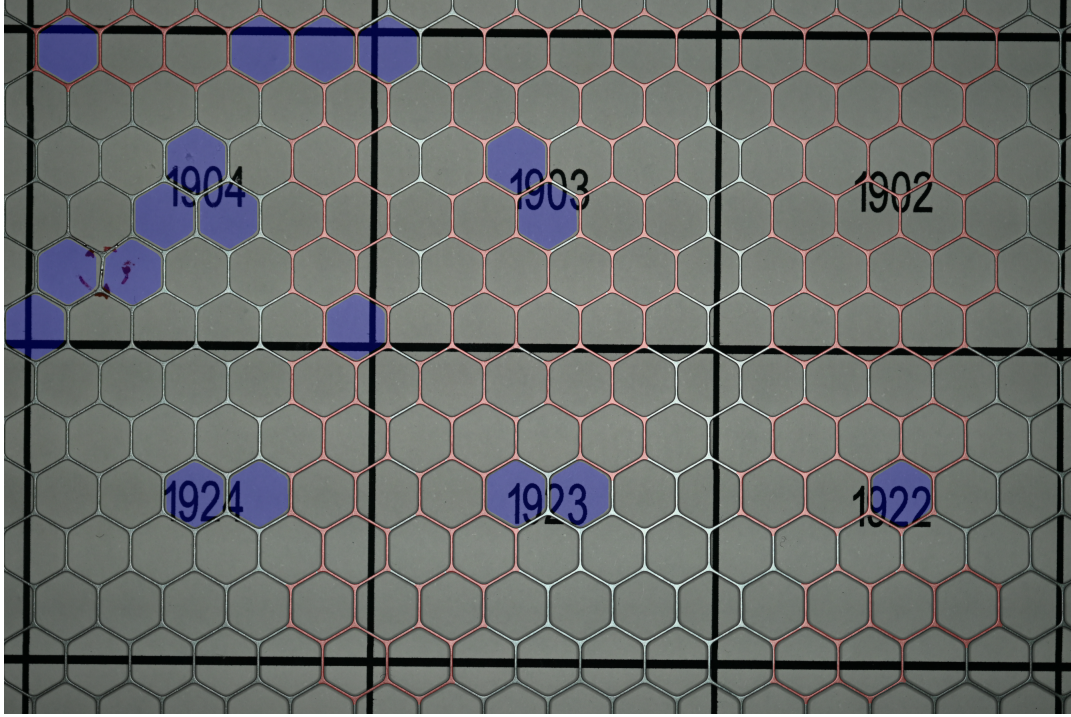


Figure 4.6: An overlay of the training data for the segmentation models. The blue shaded pixels are labeled as background pixels and the red shaded pixels are labeled as mesh pixels.

For each pixel, 600 intensity, gradient and texture features are computed using the scikit-image python package [179]. The features are defined as weighted averages over neighboring pixels at multiple scales, where the neighborhoods and weights are defined by 50 Gaussian kernels with sizes σ between $\sigma = 1$ and $\sigma = 64$ pixels. Intensity features are the RGB values, or rather their Gaussian average per kernel size σ per pixel. Gradient features are, as the name suggests, the gradients across the neighborhood defined by the Gaussian kernel, corresponding to the weighted mean Jacobian matrix for $\sigma \times \sigma$ pixels. The texture features, also referred to as local structure features, correspond to the second derivative, the Hessian matrix across some distance σ from the pixel [182].

While the full image consists of $(5568 \times 3712) = 20668416 \approx 2 \times 10^7$ pixels, only a subset of those was labeled, including 1.4×10^6 background pixels and 1.1×10^6 mesh pixels, as shown in Figure 4.6. Of this training set, 10000 pixels were left out of the training set for validation of the prediction. Two different ML models were trained for this task, a random forest classifier (RFC) and an MLP.

Random Forest Classifier

A random forest classifier (RFC) is an example of ensemble learning, where many weak models are combined to create a strong model. The individual weak models in the case of an RFC are decision trees (DTs), that show a high performance variance which makes them very suited for ensemble learning [131, 184, 185]. An illustration of the working principle of a DT is shown in Figure 4.7, taken from [185]. In a DT the input sample is tested for multiple conditions in a hierarchical order, where the conditions are generally thresholds of the samples features. In the illustration, the sample consists of three variables, x_1 , x_2 and x_3 , and the sample is classified into one of four classes, indicated by the colored circles, according to the size of these variables. The thresholds of the variables are the parameters that the model tunes during training.

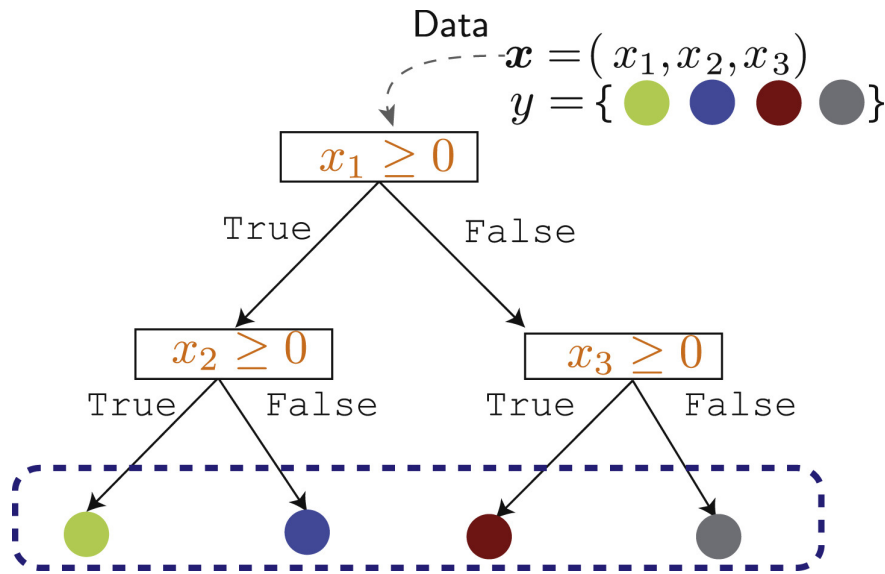


Figure 4.7: The working principle of a DT classifier, from [185]. A sample $\mathbf{x}$ with features (x_1, x_2, x_3) is evaluated against a hierarchy of conditions. The structure is that of a tree-like graph, and the leaves of the tree are the target classes (colored circles).

A RFC is the ensemble of many different DTs, where each DT sees only a subset of the features of all samples. This so-called feature bagging prevents high correlation among the individual DTs of the ensemble.

The RFC used here consists of 100 individual trees, where each tree has 5 hierarchical levels, and sees 5 % of all samples and only 24 features per sample, which is the (rounded) square root of the total number of features $\lfloor \sqrt{600} \rfloor$. The out put of the RFC is either 0, which corresponds to background pixels, or 1, which corresponds to mesh pixels. Figure 4.8 shows the predicted values of the RFC for the full training image. It can be seen that the RFC is very well able to distinguish the mesh pixels (red) from the background pixels (blue). There are isolated clusters of background pixels misclassified as mesh pixels in the center of the image, and around the marker in cell 1904.

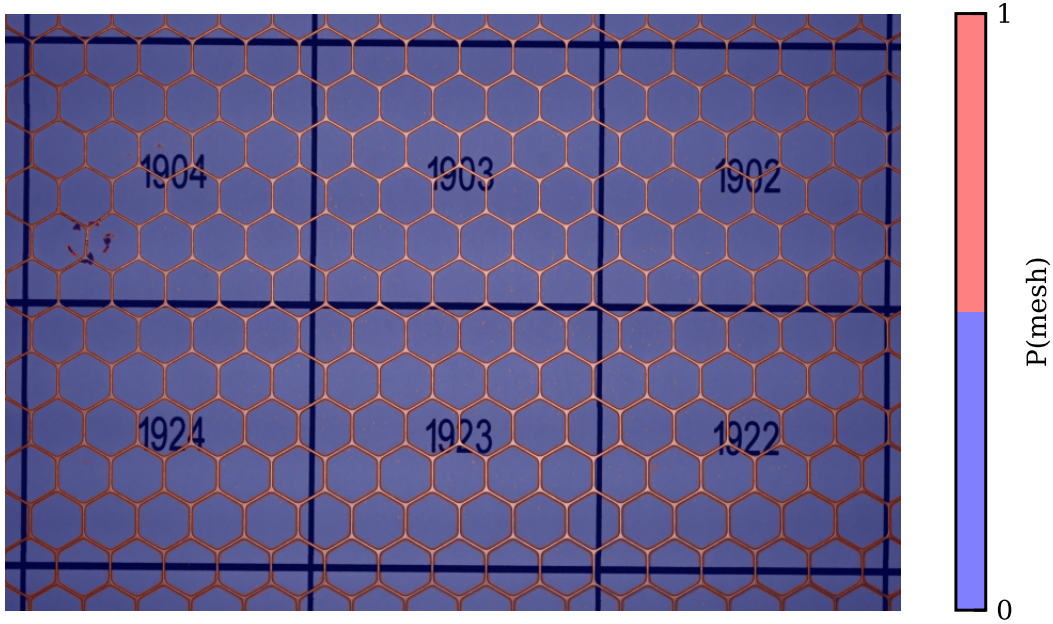


Figure 4.8: The RFC classification of all pixels of the training image as either 0, which corresponds to background (blue) or 1, corresponding to mesh (red).

Multi-Layer Perceptron

The general working principle of an MLP is described in [section 3.2](#). The model used here has 6 hidden layers with 500, 250, 100, 50, 20 and 10 neurons respectively, creating an MLP with $n_{\text{par}} \approx 4.5 \times 10^5$ parameters. Since the model is intended to work as a classifier, its output is constrained to be within $(0, 1)$, by using the sigmoid function as an activation function of the output layer. The model is trained on the binary CE loss function for 100 epochs using the Adam optimizer. [Figure 4.9](#) shows the pixel classifications of the training image predicted by the MLP. Unlike the RFC, which predicts either background or mesh, the MLP predicts a number between 0, for background-like, and 1, for mesh-like pixels, that can be interpreted as a pseudo probability of any pixel being a mesh pixel. Qualitatively, the performance seems to be better than that of the RFC. In particular, there are no clusters of background pixels misclassified as mesh, and the marker in cell 1904 is correctly classified as background.

Validation

While [Figure 4.8](#) and [Figure 4.9](#) seem promising, they show the predictions of pixels of the training image and therefore their excellent results can not necessarily be transferred to the remaining 335 base images. To compare the performance, the models are tasked to predict the annotated pixels of the validation image shown in [Figure 4.10](#). This validation image has not previously been seen by either model, and contains 3.2×10^6 labeled background 1.6×10^6 labeled mesh pixels. The validation data specifically includes a section of the stretching ring and uninterrupted black lines of the printed background grid.

The predicted classifications of the pixels of the validation image are shown in [Fig-](#)

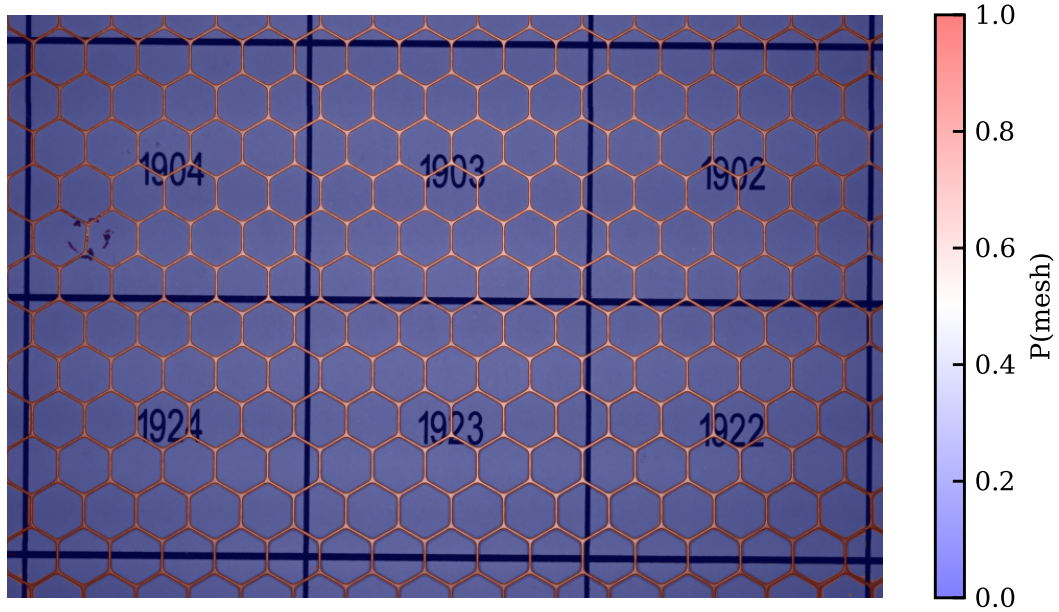


Figure 4.9: The MLP classification of all pixels of the training image, where red corresponds to mesh-like and blue corresponds to background-like.

ure 4.11. It can immediately be seen that the MLP struggles to correctly predict the stretching ring structure. In addition, the MLP has trouble predicting the less well illuminated areas in the corners of the image. The RFC on the other hand appears to showcase a very good result, correctly discriminating between background and mesh pixels.

For a classifier with a continuous output in $[0, 1]$ such as the MLP, the model performance depends on the classification threshold, the number above which a pixel is classified as a mesh pixel. Choosing a higher threshold will reduce the model efficiency, the fraction of mesh pixels correctly predicted as mesh pixels, also called true positive (TP) rate, but simultaneously reduce the number of false positive (FP) classifications, meaning background pixels mislabeled as mesh. As an illustrative example, a classification threshold of 0 means that every single pixel will be classified as a mesh, leading to an efficiency of 100 %, since all existing mesh pixels are correctly labeled as mesh by the model, however, every remaining pixel is also classified as mesh, which equals a FP rate of also 100 %. The reverse is true for a threshold of 1. The classification threshold needs to be optimized to maximize the efficiency, while keeping the FP rate low enough for the model to be suitable to its task. A useful tool to decide for a classification threshold is the so-called receiver operating characteristic (ROC), which describes the efficiency as a function of the FP rate [185]. Figure 4.12 shows the ROC curve of the MLP model (black solid) determined on the annotated pixels of the validation image Figure 4.10. In addition, the ROC of a random guessing classifier (black dashed) and the efficiency and FP rate of the RFC (blue cross) are shown.

Assessed on the validation data, the RFC has a TP rate, the efficiency of correctly identifying mesh pixels, of 96.6 % and a FP rate, the chance of misidentifying background pixels as mesh, of 1.3 %. To achieve the same TP rate with the MLP, one

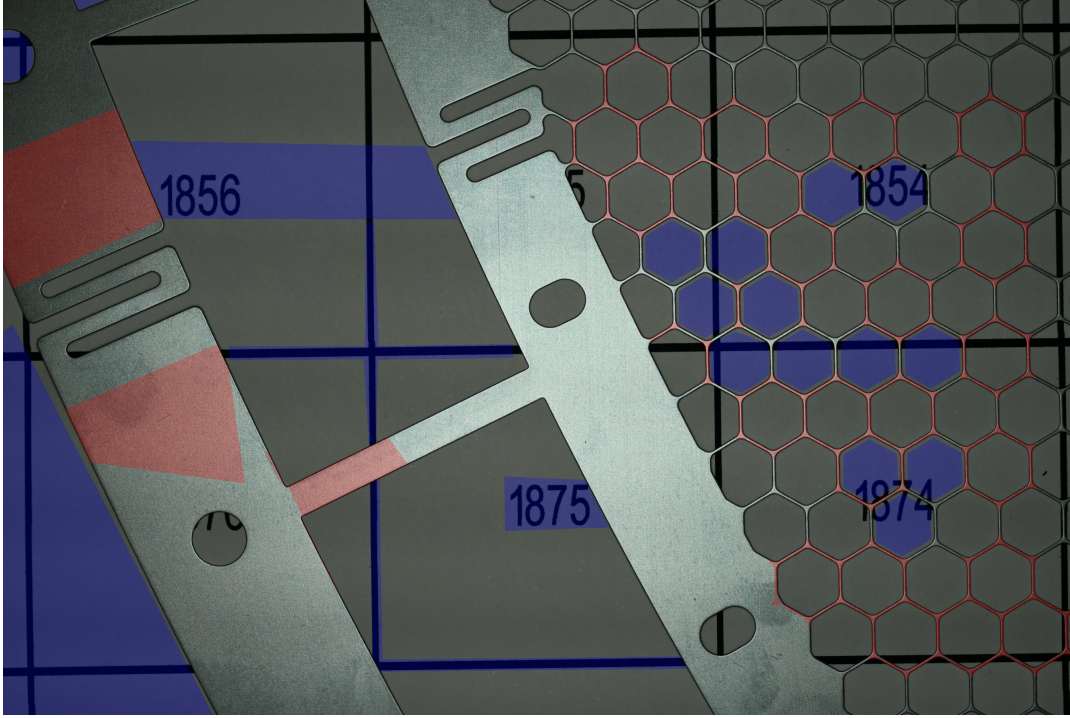
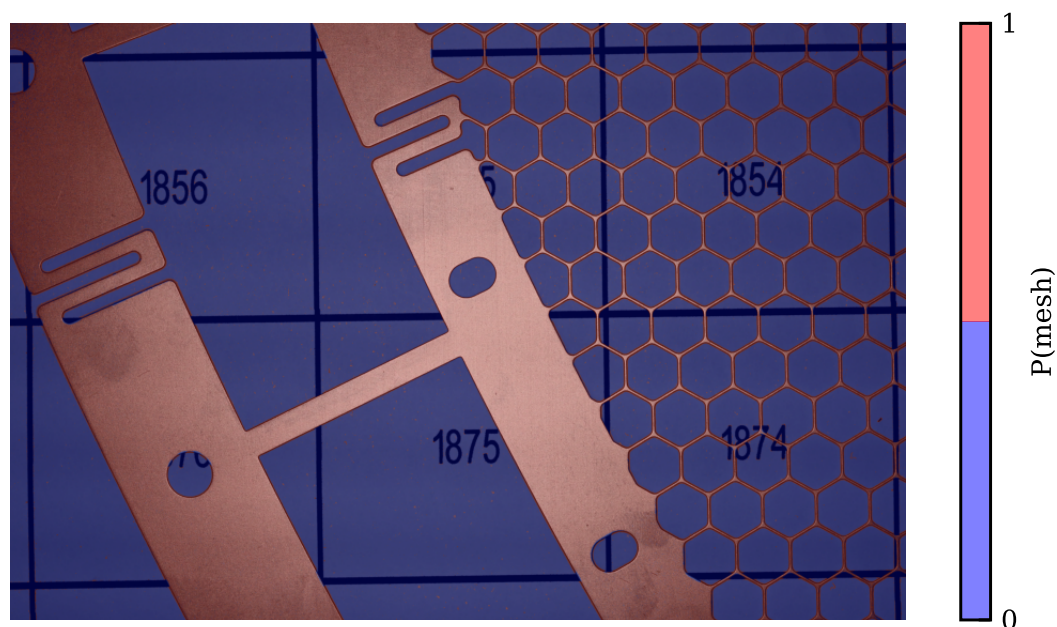


Figure 4.10: An overlay of the pixel labels used for validation of the segmentation models. The blue shaded pixels are labeled as background pixels and the red shaded pixels are labeled as mesh pixels. No pixels of the validation image have been seen by either of the models.

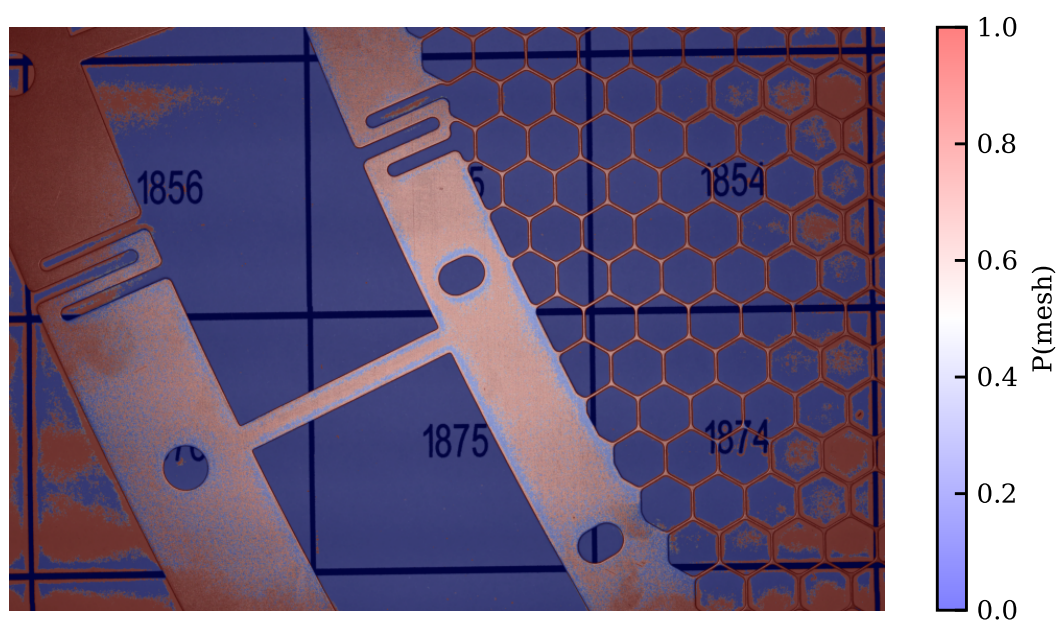
would have to accept a FP rate of 29.0 %, which is visualized by the red dashed lines. For this reason, the RFC model is chosen for the pixel classification task in the following work. A more detailed comparison of DT based models and NNs can be read for example in [186].

Using the RFC segmentation model, a binary mask of the mesh pixels is created for each base image. Subsequently, with the use of the algorithm described in [Appendix B](#), a digital hexagonal mesh is fitted to the binary mask. The digital mesh being is by *anchor points*, the points where three hexagonal edges of neighboring cells meet. An example is shown in [Figure 4.13](#), with the original image overlayed with the digital mesh and the deduced anchor points highlighted in red.

Using the derived anchor points, approximately 600 leg images of size 300×100 pixels are extracted from each base image, for a total number of 204043 images. Note that the algorithm in [Appendix B](#) also produces sub images of sections of the stretching ring that are included in the base images. Removing those leaves us with $\sim 1.6 \times 10^5$ leg images. These leg images constitute the full dataset for the detection of defects on the XENONnT replacement cathode.



(a) The classification into mesh (red) and background (blue) by the RFC.



(b) The classification into mesh (red) and background (blue) by the MLP.

Figure 4.11: An overlay of the prediction values of the RFC (top) and MLP (bottom) over the validation image.

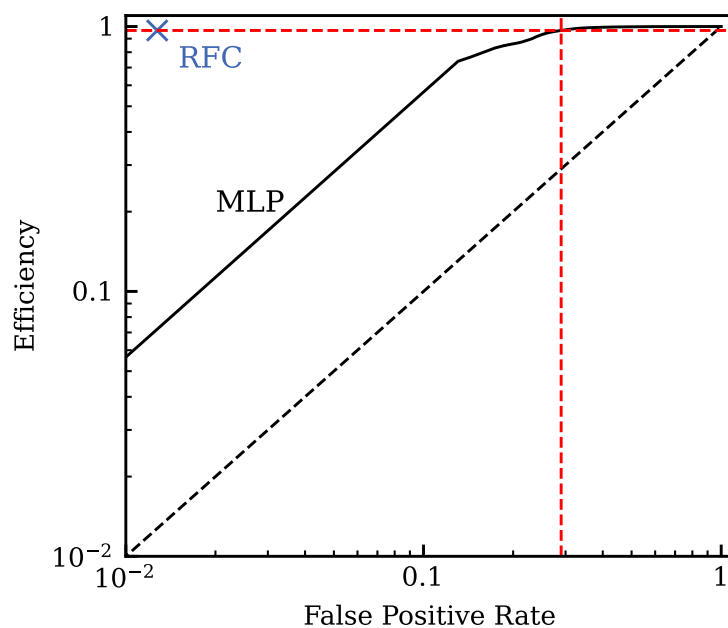


Figure 4.12: The ROC curve of the MLP model (solid black) compared to the efficiency and false positive rate of the RFC (blue cross). The black dashed line shows the hypothetical performance of a random guessing classifier. The red dashed lines visualize the efficiency of the RFC, and the corresponding false positive rate the MLP would show for the same required efficiency.

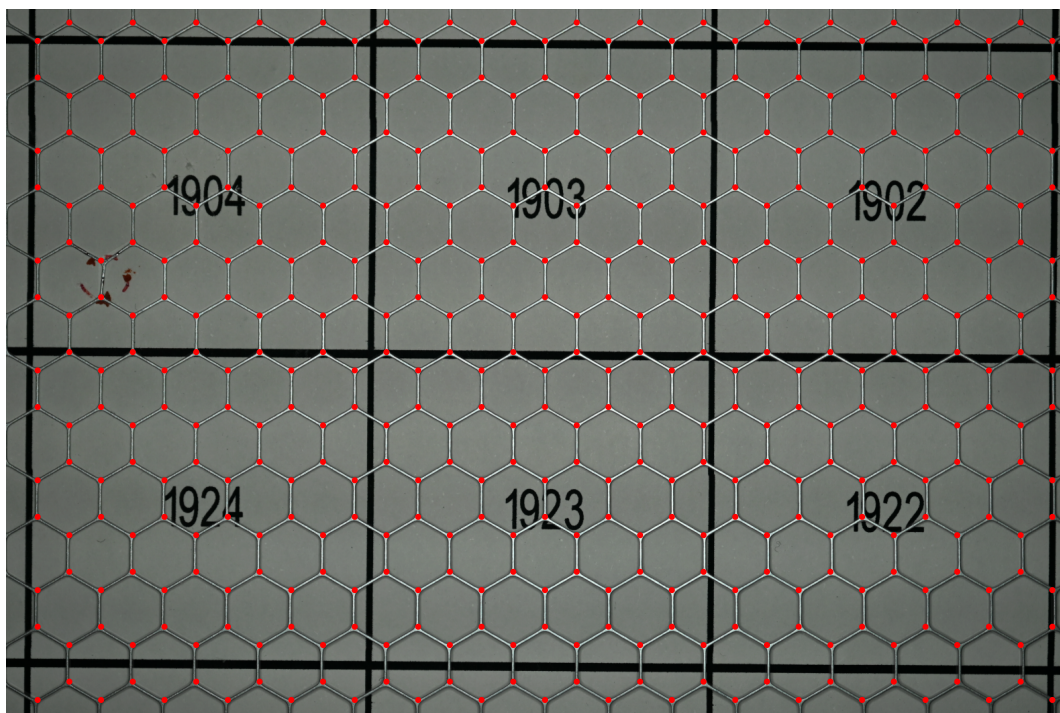


Figure 4.13: An overlay of the original picture, the digital mesh, and the anchor points (red) derived with the hexagonal mesh algorithm described in [Appendix B](#).

4.3 Machine Learning based Detection of Defects for the new XENONnT Cathode

The procedure detailed in [section 4.2](#) provided $\sim 1.6 \times 10^5$ leg images, such as the examples shown in [Figure 4.14](#). The majority of the legs look similar to (a), with some shifted within their image as shown in (b). A significant subset of the leg images has black writing or lines in the background from the printed rectangular grid that was underneath the mesh, see (c). Finally, there is a minor amount of images that showcase a defect, such as (d). This dataset was produced with the goal of training a dedicated ML model aimed at the identification of defects on the legs.

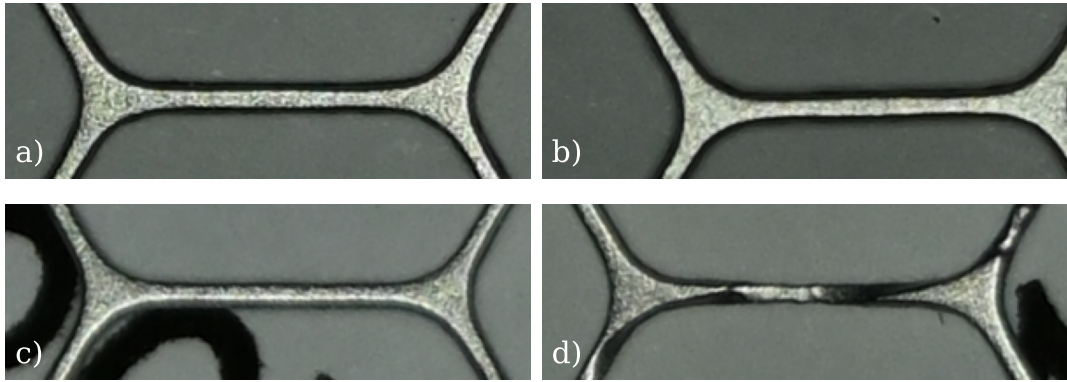


Figure 4.14: Examples of leg images. A standard image, representative of most of the data is shown in (a). Some legs are shifted slightly within their image boundaries, as that in (b). A significant number of images has background text from the printed grid, as seen in (c). Finally, there is a minor subset of images that contain defects, like the one shown in (d).

In ML, there are two types of training: *supervised* and *unsupervised* training. Supervision in ML refers to the availability of labeled data. A classification or regression model is supervised, if it is supplied with training data x and a corresponding target label y , so that it can learn the mapping $x \rightarrow y$. In unsupervised learning, there is no such label y . All applications of ML seen so far in this work have been cases of supervised learning, for example, the position reconstruction of [chapter 3](#), where x were the hitpatterns and the labels y were the positions. For the defect detection, supervised training requires access to a set of leg images labeled as "defect", and a set of images labeled as "non-defect", also called "clean" in the following. However, manual annotation of data samples is an extremely tedious task, and prone to error as even popular publicly available data sets contain mislabeled data [187]. Furthermore, there are simply not that many defects available. To train an ML model, a large quantity of data is necessary, and as stated in [section 3.2](#), an approximately even representation across different classes should be given. For those reasons, an unsupervised anomaly detection was performed, similar to the one outlined for the XENON1T electrodes in [188]. This anomaly search was done using a Variational Autoencoder (VAE).

4.3.1 A short Introduction to (Variational) Autoencoders

Autoencoder Theory

The basic idea behind an Autoencoder (AE) is to have two NNs working in accord: One *encoder* network q_ϕ , which encodes some high dimensional input $x \in \mathbb{R}^n$ through deterministic operations into a lower dimension, the so called latent variable $q_\phi(x) = z \in \mathbb{R}^{d_{\text{lat}}}$, with $d_{\text{lat}} \ll n$. The second network, the *decoder* p_θ , is provided with the latent variable and inflates the dimensionality through deterministic operations, to the output $q_\theta(z) = \tilde{x} \in \mathbb{R}^n$ [189–192]. Here, ϕ and θ denote the tunable parameters of the encoder and decoder respectively. A sketch of the working principle of an AE is shown in Figure 4.15.

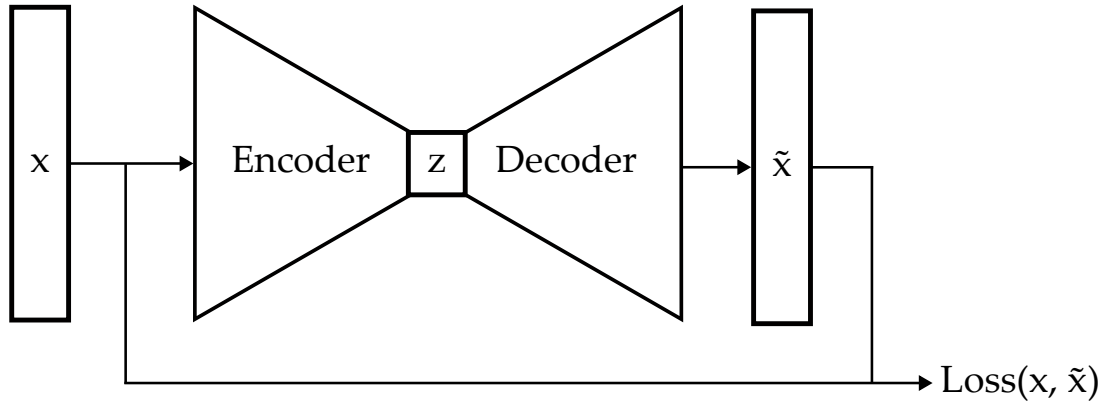


Figure 4.15: A sketch of the working principle of an autoencoder. It can be trained on the input x alone without an additional label vector, making the AE an example of unsupervised ML.

The goal of an AE is to have the output $\tilde{x}$ represent the input x as well as possible. In particular, this means that an AE does not need labels to train, as it can be trained on the input x alone, making the AE a case of unsupervised learning. An AE is typically trained on a measure of distance between the input x and the output $\tilde{x}$ as the loss function, such as the MSE. AE models are constructed to learn the underlying topology of a data set. They are used as generative models, producing new samples of the learned topology [193], to reduce information dimensionality to fight the curse of dimensionality [194] or as anomaly detectors [170].

The AE is suited for anomaly detection for two reasons: For one, it is unsupervised and does not require prior annotation of data, and class imbalance is not of concern. Second, because the latent dimension d_{lat} is much smaller than the input dimensionality n , the AE will learn to encode only the most common features of the training set. Since we assume that defects on the mesh are rare, they will not be encoded in the latent space and therefore can not be reconstructed by the decoder network. An anomaly will in this case show a very high difference between input and output $|x - \tilde{x}|$.

Variational Autoencoder

The probabilistic extension of the AE is the Variational Autoencoder (VAE), first introduced in [195]. In a VAE, the latent variable $z \in \mathbb{R}^{d_{\text{lat}}}$ is replaced by d_{lat} independent probability distributions $p(z)$. This so-called prior distribution is a simple

and known distribution, for example a d_{lat} -dimensional normal distribution $p(z) = \mathcal{N}(z; \mu, \text{diag}(\sigma))$. The encoder and decoder are then rephrased as conditional probability distributions. The encoder network is the distribution $q_\phi(z|x)$, the probability of encoding the latent variable z , given the observation x , with ϕ being the tunable parameters of the NN. The decoder $p_\theta(x|z)$ is not provided with the output of the encoder, but with a sample from the distributions $z_i \sim \mathcal{N}(\mu_i, \sigma_i)$.

A VAE is trained on the evidence lower bound (ELBO) [195]:

$$\mathcal{L}_{\theta,\phi}(x) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x, z) - \log q_\phi(z|x)], \quad (4.4)$$

where $p_\theta(x, z) = p(z)p_\theta(x|z)$ is the joint distribution of the latent space $p(z)$ and the decoder $p_\theta(x|z)$. Equation 4.4 can be rewritten as

$$\mathcal{L}_{\theta,\phi}(x) = \log p_\theta(x) - D_{\text{KL}}(q_\phi(z|x), p_\theta(z|x)), \quad (4.5)$$

where $p_\theta(x)$ is the true distribution of the data and D_{KL} is the Kullback-Leibler divergence (KL) [196]. Since the KL is always greater than 0, or 0 for $D_{\text{KL}}(p, p)$, the ELBO is a strictly lower bound of the logarithm of the true data probability, hence the name evidence lower bound.

To train a VAE, both encoder and decoder must be differentiable with respect to ϕ and θ , which is not given when the latent variable z is sampled from $q_\phi(z|x)$. This problem is circumvented with a change of variables, called the *reparameterization trick* [197], where the latent variable is transformed to the deterministic result of a differentiable function, that depends on a new and independent random variable. An in-depth explanation of VAEs, a derivation of the ELBO and more details on the reparameterization can be found in [198] or [190].

Common Neural Network Layers for VAEs

The encoder and decoder of a VAE are neural networks. In section 3.2 the introduction to NNs explained *dense* layers, hidden layers of a NN where each neuron is connected to each neuron of the preceding layer via matrix multiplication. There is a multitude of other layers, some of which are specifically suited for the use in VAEs or CV. In CV in particular, VAE encoder and decoder networks are generally Convolutional Neural Networks (CNNs) [199].

The basic building block of a CNN is a *convolutional layer*, characterized by a kernel size K and the number of kernels, also called filters, n_K . Each kernel is convolved with the input using Equation 4.2, similar to kernel processing in subsection 4.2.1, and the entries of the kernels are the learnable parameters. In general, the convolution from Equation 4.2 is only defined in regions where both input and kernel matrices overlap, which makes the convolution a shape changing operation. Convolutional layers have the option to add P pixels padding on each side of the input, to manipulate the output size. In addition, one can define a stride S , that defines the step length of the discrete convolution. A stride of $S = 1$ means convolving the kernel with every pixel of the input, $S = 2$ means every second pixel and so on. The output size W_{i+1} of a convolutional layer with n_K kernels of size K , P padding and stride S is

$$W_{i+1} = \frac{W_i - K + 2P}{S} + 1, \quad (4.6)$$

where W_i is the input size of the layer. For an image with a width w , height h and channels c , the size changes according to $w \times h \times c \rightarrow w_{i+1} \times h_{i+1} \times n_k$, where w_{i+1} and h_{i+1} are computed according to Equation 4.6. As an image is two-dimensional, possessing a width and a height, the variables in Equation 4.6 can be different in each direction.

A *pooling layer* has the purpose of drastically reducing the height and width of the input, characterized by a kernel size K and a stride S . The pooling layer takes a region of size K and reduces it to a single number for its output, most often computing the maximum or the average. If $S = K$, then the input is separated into distinct regions, if $S < K$ the regions overlap and if $S > K$ some pixels of the input are skipped entirely. The pooling layer has no learnable parameters.

These two layer types, being able to reduce dimensionality, make up the main components of an encoder. A decoder on the other hand needs to increase the size per layer, which is done by doing the opposite operation of the convolutional and pooling layers. The opposite of a pooling layer is an *upsampling layer*, that replicates certain pixels. The inverse of a convolutional layer can be achieved with a transposed convolutional layer, also called a *deconvolution layer* [200].

Using these, it is possible to construct the encoder and decoder as mirror symmetric models of each other, ensuring that they have the same number of trainable parameters and the same dimension after each corresponding layer. Having the same size allows for so-called *skip connections*, where part of the encoding and decoding step is skipped. During training, a fraction of the data is transferred from the encoder directly to the corresponding layer of the decoder, skipping all layers in between, in particular the latent space.

4.3.2 Performance of the VAE Defect Detection

The principle of defect detection using VAEs is visualized in Figure 4.16. The base assumption of the unsupervised anomaly detection is that the defects are in general rare, and therefore the VAE learns the representation of the *clean*, meaning non-defect, legs. The VAE has not learned how to reconstruct a defect, and therefore returns an intact leg if provided with a defect leg. The defect appears as a high deviation in the squared difference between input and output of the model.

Comparison of different VAE Architectures

There are a multitude of tunable hyperparameters in the VAE architecture that can be adjusted to try and improve the model, including but not limited to the layout of convolutional layers that defines the model parameters n_{par} , the size of the latent dimensions d_{lat} , the preprocessing pipeline that can include any combination of processes introduced in subsection 4.2.1 and the loss function.

To decide which VAE architecture to use, there is the need for some figure of merit to compare different models, which also represents the efficiency to detect known anomalies. However, as mentioned above, defects on the mesh are a rare occurrence.

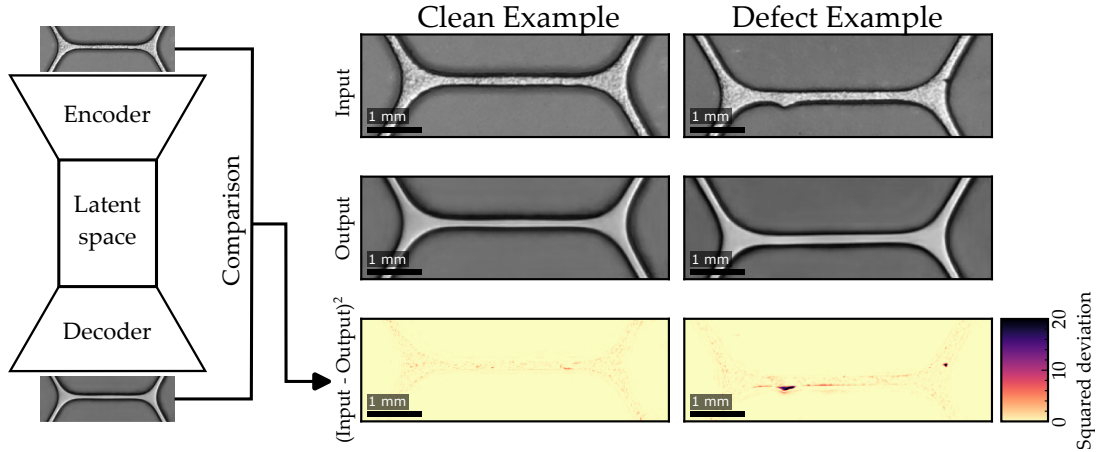


Figure 4.16: A sketch of the VAE anomaly detection. An input image is encoded into a small-dimensional latent space by an encoder NN. The information of the latent space is decoded by a decoder NN, that tries to replicate the input. Since defects are rare, the VAE has not learned to represent them in the latent space and they are therefore not reconstructed by the decoder. A defect appears as a region of increase squared differences between input and output. Adapted from [2].

While we utilize this fact to our advantage during training, it poses a challenge when judging the performance of different defect detection models. There are 18 legs with known defects out of $\sim 1.6 \times 10^5$ leg images, which is a low number to make statistically significant statements about model accuracy. The 18 defects were partially known beforehand from manual inspection or discovered over time during the testing of different VAE models. However, at the time the base images were taken one round of repairs had already taken place (see stage 1 in [section 4.4](#)). During these repairs the worst of the defects were repaired via laser welding replacement legs in place of the defect ones. As the laser welding spots are definitely anomalous, they are included in the list of defects.

To gauge and compare the efficiency of the models for defect identification, ROC curves were derived, using the 18 aforementioned defects and 429 manually selected "clean" leg samples. A ROC curve is the efficiency of a model to detect a defect as a function of its FP rate, which depends on the threshold chosen to label an image as defective. A very general figure of merit is the area under the curve (AUC) of the ROC. The higher the AUC, the better the model. [Table 4.1](#) shows a selection of hyperparameter combinations and the respective AUC of VAEs trained for 200 epochs on the full set of 1.6×10^5 leg images. For each of these settings, that are explained in more detail below, multiple variations with different layer combinations, and therefore n_{par} , as well as different latent sizes d_{lat} , were trained and tested. Note that the total number of trainable parameters includes both encoder and decoder and is mainly driven by the fully connected layers that transfer from the 3-dimensional convolutional layer to the 1-dimensional latent space and back. The latent space is a

Table 4.1: A selection of different defect detection VAE models and the architecture hyperparameters that were used. In addition, the last column shows the AUC of the ROC curve.

Model	n_{par}	d_{lat}	Preprocessing	Loss	AUC
Baseline	3.5×10^5	200	Standard scaling	ELBO	0.87
Gauss kernel	2.2×10^5	200	Standard scaling Gaussian blurring	ELBO	0.85
Remove writing	5.4×10^5	100	Standard scaling Print removal	ELBO	0.82
KL annealing	2.1×10^5	100	Standard scaling	KL annealed ELBO	0.82
Skip connections	1.1×10^5	50	Standard scaling	ELBO	0.51
Fake defects	0.8×10^5	10	Standard scaling	Defect informed MSE	0.78

designed bottleneck that automatically counteracts overtraining. The total number of parameters n_{par} is therefore not as important as it is for other ML models, as long as there are enough parameters for the encoder and decoder to properly learn the image distribution. Additionally, an insight gleaned from this study is that all models require some sort of scaling to actually learn to encode and decode the images, and standard scaling is suited very well for this purpose.

The **Baseline** model is a standard VAE with symmetric encoder and decoder models, consisting of alternating convolutional and maximum pooling layers, or transpose convolutional and upsampling layers in the case of the decoder. Between the last convolutional layer of the encoder and the latent space the data is flattened and linked via fully connected layers, which provide the majority of trainable parameters. For the baseline model, only standard scaling is applied to the images and the model is trained on the ELBO as described in [Equation 4.4](#). All the following models are based on this model, with slight variations in the configurations of the convolutional layers.

The **Gauss kernel** model is similar to the baseline, but with Gaussian blurring with a width of $\sigma = 2$ px applied to the images before training and inference. Multiple combinations of different processing methods listed in [subsection 4.2.1](#) were experimented with, but they did not change the performance of the VAE in any significant way. While some, namely Gaussian blurring, improved the general reconstructive capabilities of the VAE, it did so for both defects and clean legs, in particular not improving the anomaly detection. Therefore the idea of applying basics CV preprocessing functions to the data, with the exception of scaling, was discarded again.

A major challenge for all models was posed by the black prints on the background grid present in many of the leg images. The **Remove writing** preprocessing identifies the prints via histogram matching and replaces them with a patch of pixels, the values of which are sampled from a Gaussian distribution. This distribution is centered on the average gray value of the image, and has a width corresponding to the standard deviation of the grayscale values of the image. An example is shown in [Figure 4.17a](#). This process worked quite well, but outlines of the prints remained, and on rare occasions a part of a leg got misidentified as writing and removed too. The latter part is

also a problem for blinded models, that utilize the RFC pixel classification described in subsection 4.2.2 to set background pixels to zero before training, visualized in Figure 4.17b. In addition, the blinded models showcased serrated edges around the legs, that made it impossible for the models to correctly identify extra material on legs. Based on this, both the writing removal and the blinding via RFC pixel segmentation were discarded again.

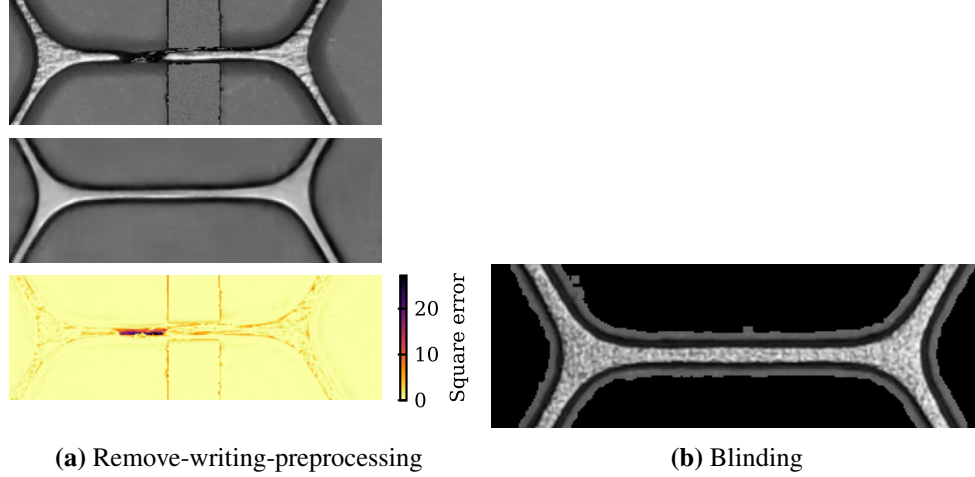


Figure 4.17: Examples of ways to counteract model confusion created by the black writing on the background grid. Figure 4.17a shows the removal of the writing based on color histogram matching in the top panel. The central panel depicts the reconstructed image, and the bottom panel the squared difference between original and reconstruction. Figure 4.17b portrays an example of a leg image with all background pixels, as identified by the RFC model described in subsection 4.2.2 set to zero.

KL annealing - the process of *annealing* in ML refers to the adjustment of a parameter of the loss function during training. What is done for the training of VAEs, is KL annealing [201], where, the ELBO loss function Equation 4.5 is modified with a parameter β to become

$$\mathcal{L}_{\theta,\phi}(x) = \log p_{\theta}(x) - \beta \cdot D_{\text{KL}}(q_{\phi}(z|x), p_{\theta}(z|x)). \quad (4.7)$$

At first β is set to 0, to train the model only on the distribution of the output data, disregarding the probability distribution in the latent space and with increasing training epochs, β is ramped continuously up to $\beta = 1$. Different schedules were tested, from high frequency circular annealing, where β rose from 0 to 1 during each epoch, to long winded annealing where β only started increasing after the model had already been trained for 50 epochs. Adjusting the ELBO loss via **KL annealing**, as expressed in Equation 4.7, did not yield any improvements in model performance.

The usage of **Skip connections** makes it possible to bypass the latent bottleneck, and is done for example in the widely used U-net model [202]. However, here it was observed that the VAE became too skilled at reconstructing the original image, in the sense that it learned to perfectly reconstruct also defect legs, as highlighted in Figure 4.18a. This was observed even for models where the only skip connection was between the penultimate convolutions around the latent dimension. This suggests that the latent encoding is a large factor in the suitability of VAEs for anomaly detection.

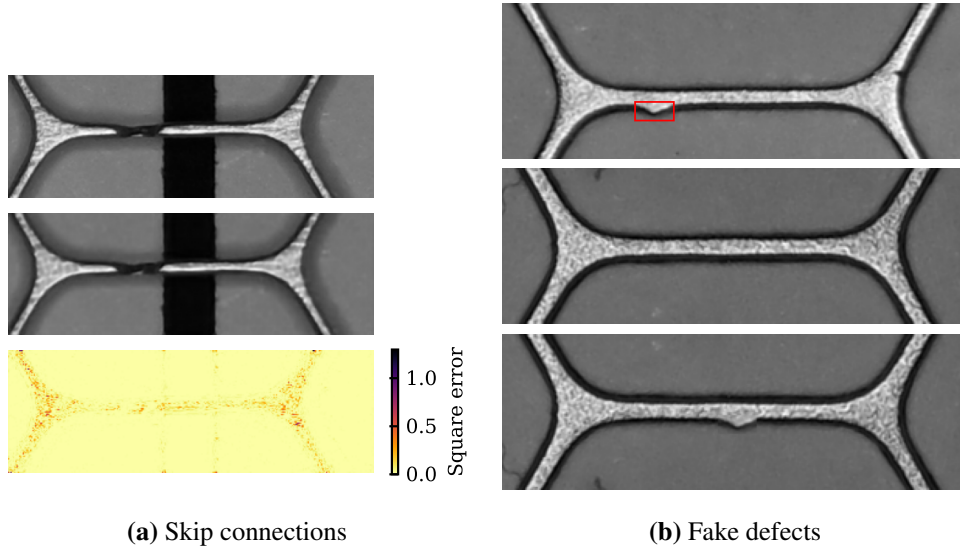


Figure 4.18: Figure 4.18a shows an example reconstruction performed with the skip connection VAE. The top row shows the original image, the center row the reconstruction and the bottom row the squared difference between both. Figure 4.18b visualizes the procedure to generate fake defects. The original image containing the defect is shown in the top panel, the original clean image in the center and the artificial defect image at the bottom. The red square highlights the section that is extracted from the defect image.

The final method listed in Table 4.1 is labeled **Fake defects**. The possibility of augmenting leg images to generate artificial defects is briefly mentioned in subsection 4.2.1, which, if successful, can alleviate the challenge of only having limited defect leg images available. Figure 4.18b shows an example of an artificial defect: The top row shows an original image of a defect leg, the defect highlighted with a red rectangle. The central row shows the unedited clean leg image that is chosen to become an artificial defect and finally, the bottom row shows the combination of the defect, extracted from the defect image, and the clean image with the extracted defect copied on top of it. The artificial defect can be placed anywhere, and on any side, of the chosen clean leg. The idea behind the artificial defect model is to adjust the loss function to train the model to explicitly reconstruct clean images if supplied with a defect image. Given some fraction k , that can also be annealed during training, for k images of the training set, the original image X is modified by adding an artificial defect to create X' . This modified image is propagated through the encoder and decoder resulting in a reconstruction $\tilde{X}$. However the loss function is not the ELBO, but rather the MSE between reconstruction and *original unmodified image* $\mathcal{L} = \langle (X - \tilde{X})^2 \rangle$. The idea behind this model was to teach the VAE to reconstruct a non-defect image even if a defect image is supplied as input. While a default VAE anomaly detection model is doing so implicitly, by not learning to properly reconstruct defects, the fake defect training procedure explicitly demands the model to learn a "clean" reconstruction of defect samples. However, as glanced from Table 4.1, this fake defect training procedure did not improve the models capabilities beyond the baseline model.

The ROC curves of the models listed in Table 4.1 are shown in Figure 4.19 all together. The Skip connection model is reconstructing even defects too well and does not perform better than a random classifier. The remaining model adjustments do not

seem to have a large impact on the model performance. For this reason, the baseline VAE was chosen as the final anomaly detection model.

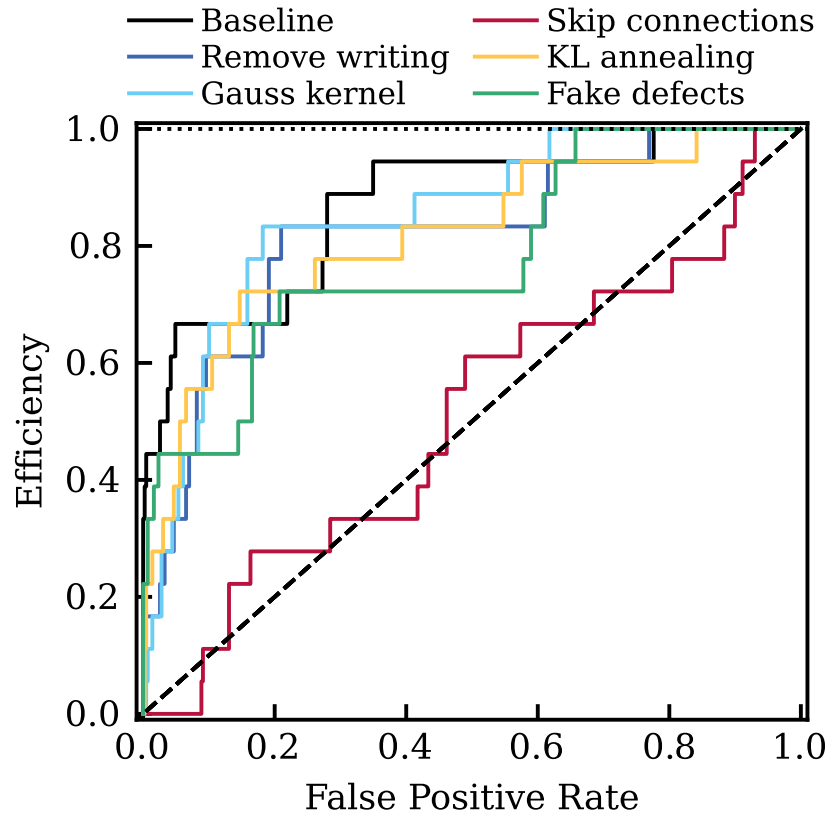


Figure 4.19: The ROC curves of various VAE models. The black dashed line indicates the performance of a randomly guessing classifier. The black dotted line indicates an efficiency of 100 %.

VAE Performance on Artificial Defects and Outlook for XLZD-scale Electrodes

To evaluate the feasibility of the developed VAE model for defect detection, in particular in regard to future experiments like XLZD, an ROC curve with more statistics was constructed. However, since there are only 18 known defects available, the higher statistics were achieved by using the artificial defect generation, detailed in [subsubsection 4.3.2](#). [Figure 4.20](#) shows the ROC curve deduced with 40,000 clean images and 40,000 artificial defects. The clean images were taken from the general sample of the leg images, with the 18 known defects removed, meaning in particular that those 40,000 can by chance also include so far undiscovered defects. Using artificial defects, the baseline VAE achieves an AUC value of 0.98. However, this value is to be accepted carefully, since there is no distinction between the model recognizing defects, and the model recognizing *digitally altered images*, such as the artificial defects.

In addition, [Figure 4.20](#) also shows the corresponding FP rates if one wants to achieve 90 %, 95 % or 99 % defect detection efficiency. If an efficiency of 90 % is sufficient, the model comes with a FP rate of 4 %. With a total pool of 1.6×10^5 images, this leaves 6400 images that get flagged as defect, and that still need to be analyzed by eye. With a higher required efficiency, this number only increases drastically, which

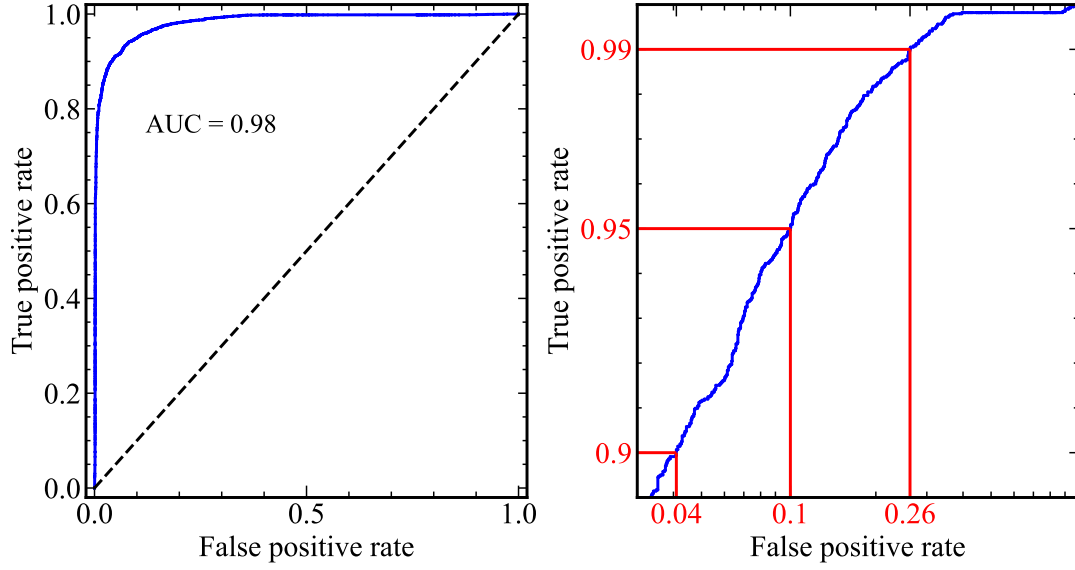


Figure 4.20: The ROC curve of the VAE anomaly detection using artificial defects, from [2]. The right panel is a logarithmic zoom-in of the full curve on the left. Shown in red are the FP rates one has to accept, given a required detection efficiency of 90 %, 95 % or 99 %.

underlines a general issue with using anomaly detection to find potentially detrimental spots on the electrode. Let ϵ be models efficiency to detect a defect and n_d be the unknown number of actually present defects. The chance to miss at least one defect is $p = 1 - \epsilon^{n_d}$. Assuming there are $n_d = 20$ defects, if one wants less than 1 % chance of missing a defect, the required defect detection efficiency equals $\epsilon = 0.9995$. This becomes a bigger challenge for electrodes of increased size, such as planned for the XLZD experiment. Lastly, the defect detection model detailed here can only find potential defect, it does not provide any information about whether or not an identified spot is actually detrimental to the performance of the electrode. Hence, the electrode underwent a dedicated HV test, which is detailed in the next section.

4.4 High Voltage Test of the new XENONnT Cathode

The process outlined in [section 4.3](#) is suited for the identification of defects, features that can potentially harm the performance of the replacement cathode, however, it does not provide information about how the detected defects act under HV conditions. For this reason, a series of HV tests were performed in a gaseous argon environment, detailed in [2]. The goal of the tests was to put the cathode under HV and observe spurious electron and photon emission. This was done via the current that was read out via the used HV supply, or by light emission from defects, that is commonly associated with effects such as corona discharge, electroluminescence or electrical breakdown.

4.4.1 Experimental Setup

The HV test setup was built inside a Faraday cage that was covered with curtains to darken the test environment and be more sensitive to light emission from the cathode. Inside the Faraday cage, a PVC box with side of length 160 cm and a height of 20 cm

was placed, covered by a sealed polycarbonate lid to maintain a gaseous argon (GAr) atmosphere inside. The GAr was introduced into the setup via a tube inlet on the side and controlled via a flow meter connected to the supply bottle. A circular aluminum plate with rounded edges with a thickness of 0.5 cm and a diameter of 120 cm serves as a ground plate. The cathode mesh was mounted on its ring and suspended around 4.97 cm above the ground plate with the use of dedicated insulating spacers. In the first iterations, insulators made from PVC, POM or 3D printed ASA plastic were tested and ultimately discarded, as breakdowns were frequently observed on or in close proximity to them. The final design used PEEK pillars suspending the mesh from the box lid. It should be noted that the mesh was mounted in non cleanroom conditions, and that the stretching structure was not removed for this test. The gap of 4.97 cm between mesh and ground plate was used to define the electrical field, but the ground plate had an unevenness of 6 mm, leading to a relative uncertainty of $\sim 15\%$ for the derived electrical field. The mesh was connected to a FuG Elektronik HCP 140 power supply outside the box that can provide up to 200 kV of voltage.

Approximately 140 cm above the cathode, two cameras were mounted to the ceiling, which had the full mesh in their field of view. The cameras, that were used alternately, were a Nikon Z 50, a Canon EOS 90D, and a Canon EOS 700D, equipped with the NIKKOR DX 18-140mm f/3.5-6.3, Canon EF-S 18-135mm f/3.5-5.6 IS USM, and Canon EF-S 18-135mm f/3.5-5.6 IS STM lenses correspondingly. During the tests, the cameras took pictures with an exposure of 5 s or 10 s, to increase the sensitivity to detect dim light from potential photon emission. A schematic of the test setup is shown in [Figure 4.21](#).

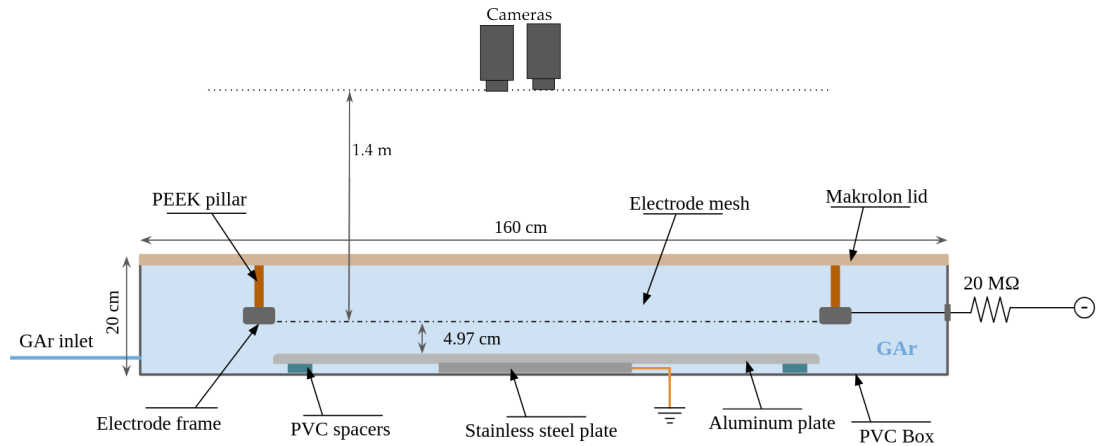


Figure 4.21: A schematic of the HV test setup, not to scale, adapted from [2], inspired by [160].

The HV testing procedure consisted of multiple test runs, each composed of multiple HV ramps, during which the cathode voltage was stepwise increased, while current and voltage were monitored, read out and stored, using a Labview based HV power supply software. For each test run, the box was closed and continuously flushed with GAr. The test was ended prematurely when electrical breakdown or great fluctuations in the current were observed. Before each test run, multiple calibration images were taken with each camera: One position calibration image, with an illuminated Faraday cage, allowed the correlation of observed light emission to the position on the mesh. In addition, multiple images were taken with the lens cap attached to the camera.

Pixels that were consistently showing a signal despite the lens cap were deemed noisy and discarded from the analysis of the corresponding test. Finally, the internal clocks of the cameras and that of the HV supply were synchronized, to properly correlate recorded images with the voltage and current profiles.

4.4.2 High voltage Test Results

An example ramp, performed on a spare mesh, is shown in [Figure 4.22](#). This spare mesh was produced identically to the XENONnT upgrade cathode, but to study whether the test setup can detect defects, one leg was cut and bent. In addition to the two top-view cameras, the third camera was placed next to the mesh and took a video of the region around the bent leg. The top panel of [Figure 4.22](#) shows the voltage (black) and current (green) during the ramp. At around 450 s after the start of the ramp, a sharp increase in the current was observed, indicative of electrical breakdown. Afterwards, the voltage was ramped down again. The lower left panel shows the current zoomed in around the breakdown, indicated in red. In addition, the pixel brightness of the video camera is shown. In gray is the average pixel brightness of each video frame, and in blue is the average brightness of pixels in a 50×25 pixel region around the bent leg. In the seconds before the electrical breakdown, the current increases, as well as the brightness around the bent leg. The two images show frames of the video some seconds before the breakdown (a) and right at the breakdown (b). In the time building up to the breakdown, a faint glow around the bent leg could be observed. The ability to detect this faint glow and correlate it with both the increase of current prior to breakdown and the location of the defect, demonstrated a method for detection of detrimental defects on the mesh electrode without the need to induce breakdown. It is to be noted, that image (a) had its brightness and contrast increased for visualization purposes in the figure shown, however, the pixel brightness profile was obtained without the use of any image enhancement techniques. The faint glow shown in (a) was also detected by the two cameras suspended above the HV box.

The assessment period of the XENONnT upgrade cathode was split into two stages, as some repairs were conducted in between HV tests. Prior to the stage 1 HV test runs, 6 defects had been identified manually, 2 of which were polished to remove minor material excess, while the remaining 4 were repaired by laser welding intact replacement hexagons from the spare mesh into the upgrade cathode. At this point, after the repair measures, but before the first HV tests, the images utilized in the ML-based defect detection were taken. Including those defects detected with the VAE model, there were 15 points of interest present in stage 1 of the HV tests. The number of total defects differs between here and [section 4.3](#), because two neighboring defect legs are considered as a single feature for the HV tests, while [section 4.3](#) considered every leg individually. Prior to stage 2 of the HV testing, four features were treated by grinding them down, and 7 additional ones were replaced via laser welding, leaving in total 5 untreated features on the mesh. Further, one hexagonal cell got slightly deformed during handling of the mesh between stages one and two, leading to one additional, untreated, feature that is present during the stage two tests but was not present during stage one. During stage 1, which consisted of 17 HV ramps, 19 breakdowns were observed, of which 16 occurred on the stretching structure and 3 on the mesh itself, where a breakdown is defined as a drop in voltage of more than 50 V. Stage 2 had 5

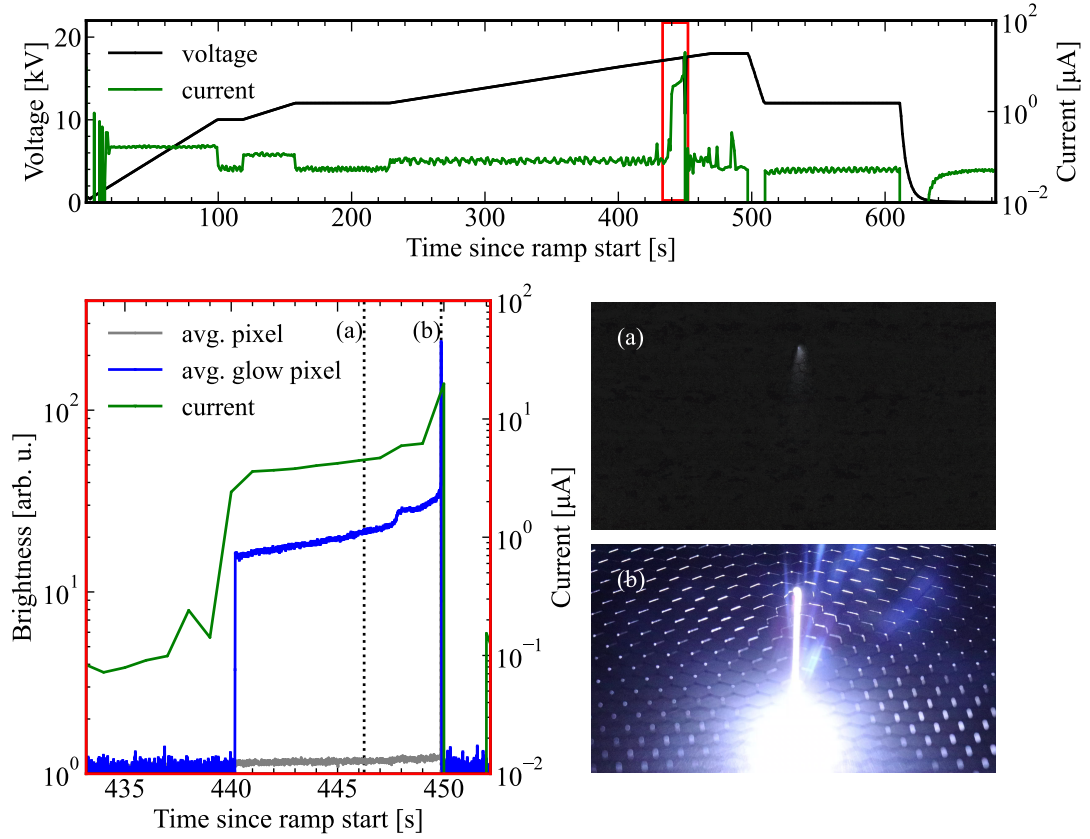


Figure 4.22: The correlation of electrical current and image brightness during a breakdown for an example ramp, taken from [2]. The top panel shows the current and voltage during the ramp. The lower left panel depicts a zoom in into the time window around the breakdown, showing the current (green) and pixel brightness (gray and blue). Around the bent leg, there is a visible increase in brightness (a), building up to the breakdown (b), that correlates with the behavior of the current.

HV ramps and 3 observed breakdowns, all along the cathode frame. The positions of known features and the observed breakdowns that happened on the mesh are shown in Figure 4.23. The stage 1 defects shown are the ones identified with the anomaly detection detailed in section 4.3.

The identified breakdown locations were of no concern, as during the final installation inside the detector, the stretching ring was removed from the cathode and the electrode frame was covered. Breakdowns on the mesh itself during stage 1 can be explained by dust particles, that got burned off during the first few tests, since they were not observed during stage 2 of the HV tests. Finally, the locations of the observed breakdowns did not correlate with the position of known features, indicating that neither the repaired positions, nor the remaining defects pose any problems to the HV performance of the cathode.

Considering all breakdowns, no matter their position on the electrode, and the electric fields they were happening at, a survival probability was computed, shown in Figure 4.24. The bulk field was derived using the last stable voltage before breakdown and the distance of 4.97 cm between ground plate and mesh. In cases where multi-

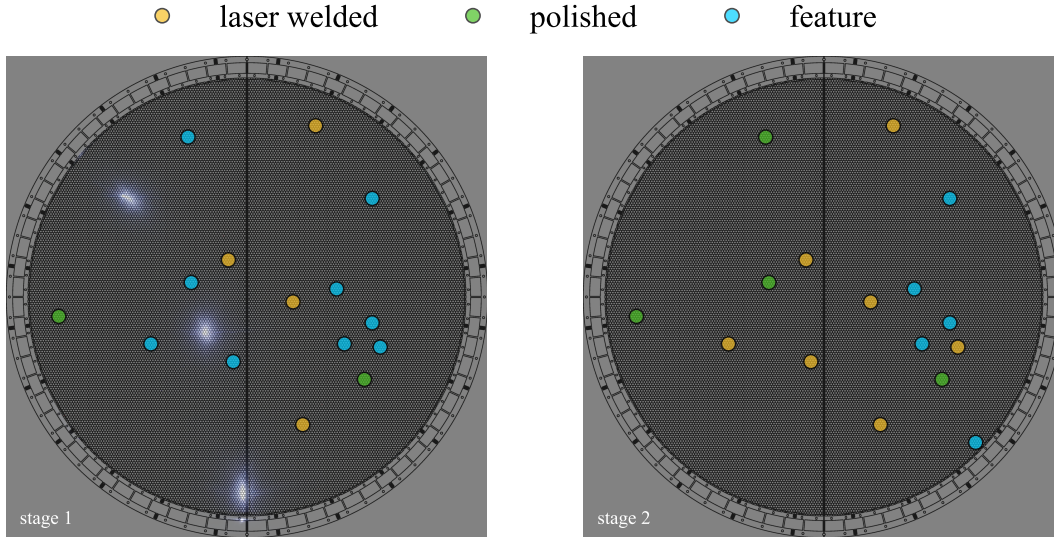


Figure 4.23: The breakdowns observed on the cathode mesh during stage 1 (left) and stage 2 (right), from [2]. Yellow and green indicate features repaired before the respective tests. Only breakdowns on the mesh itself are shown, not those that occurred on the stretching ring or on the cathode frame. The markers are bigger than the spatial resolution of the image analysis procedure used to identify glow.

ple breakdowns happened during one ramp, only the first one was considered, and in cases where no breakdown was observed, the maximum achieved voltage was taken as a right-censored value. The achieved fields were fitted with a Weibull distribution using the Reliability package [203].

The fit shown in Figure 4.24 indicates that in GAr the mesh can withstand an electric field of 3.1 kV/cm at 95 % survival probability, with an upper and lower limit of 2.8 kV/cm and 3.4 kV/cm respectively. This corresponds to a field strength in LXe of 5.9 kV/cm [2].

4.5 Conclusion and Outlook

The XENONnT experiment underwent an upgrade conducted during 2025, which included a replacement of the cathode electrode. The replacement cathode underwent extensive quality assurance, detailed in this chapter and in [2].

The cathode consists of two laser welded half meshes with a hexagonal grid. The hexagon edges, called mesh legs, have a width of only 0.3 mm and are therefore susceptible to damage during handling and production. To identify potential weak points, and sharp features that can lead to locally increased electric fields, an ML based anomaly detection model was developed. By using CV concepts and a RFC segmentation model, pictures of individual legs are extracted from base images taken with a commercial digital camera. These leg images are then used to train a VAE, a model consisting of an encoder and decoder NN working in accord. Under the assumption that defects are rare, the VAE does not learn to properly represent them and can identify them via deviations in the reconstruction. Multiple different architec-

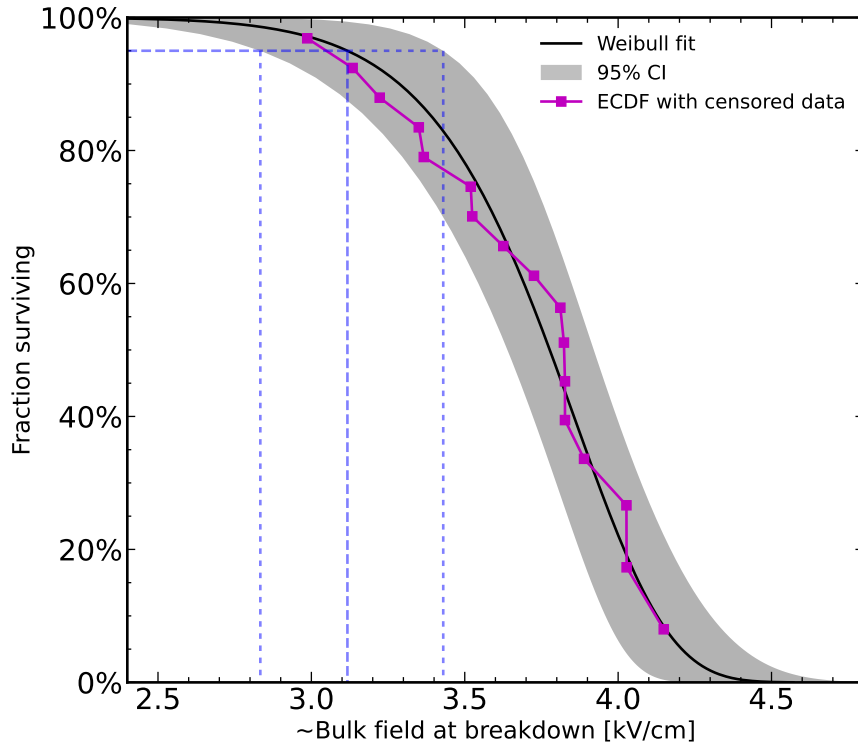


Figure 4.24: The survival probability of the XENONnT upgrade cathode, from [2]. The data points (violet) show the CDF of the achieved electric fields before breakdown. The black line is a fitted Weibull distribution.

tures and preprocessing pipelines were tested and compared, settling in the end on a basic convolutional VAE without additional image processing or modification of the loss function. In total, 18 anomalous legs were identified and, if deemed necessary, repaired using laser welding or polishing.

Optical identification methods only allow the detection of anomalous features, but not the assessment of their performance under HV conditions, or their danger to the electrode performance as a whole. For this reason, additional tests under realistic usage conditions such as noble gas environment and HV were performed, and yielded that in its final stage, the cathode mesh can withstand an electric field of 5.9 kV/cm at 95 % survival probability [2]. In particular, none of the laser welding repaired features showed any anomalous behavior in HV conditions.

In regard to even bigger detectors, like the XLZD experiment, inspection of the electrode by eye becomes even less feasible, making an automatic system more necessary. The defect detection using VAEs is a promising approach, and can be improved for future detectors for example by using an automatic imaging system that does not need to rely on a background grid for spatial orientation. The laser welding approach proved to be very suitable to repair identified defects. Setups to further investigate the electrode development for future TPC experiments include the GRANITE experiment at the JGU Mainz [204], or the HiCUTIE test setup at KIT [205].

The Search for Bosonic Dark Matter with XENONnT

As described in [section 2.2](#), bosonic DM, such as axion-like particles (ALPs) or dark photons (DPs), can interact with SM matter via the axio-electric effect ([Equation 2.9](#)) and absorption ([Equation 2.11](#)) respectively. The parameters that characterize these interactions are the axio-electric coupling strength g_{ae} and the kinetic mixing strength κ . Both interactions mimic the photoelectric effect in terms of signal generation in LXe and as such are expected to produce a monoenergetic electronic recoil (ER) in the TPC, with energy equal to the mass of the DM particle. The work presented in this chapter focuses on such photoeffect-like absorption, neglecting that bosonic DM particles with mass greater than approximately 100 keV can also interact Compton-like in LXe. The implications from taking into account Compton-like scattering are detailed for example in [\[206\]](#).

As seen in [Figure 2.7](#), the leading experimental exclusion limits for these interactions in the energy range from 1 keV to 140 keV for g_{ae} or 1 keV to 200 keV for κ come from the XENONnT experiment, using SR0 data [\[54\]](#). However, the PandaX collaboration, another LXe direct DM detection experiment, recently published a paper extending their search region for bosonic DM signals up to 1050 keV [\[72\]](#). This chapter presents the extension of the energy region of interest (ROI) for the search for bosonic DM signals with the XENONnT experiment up to 1000 keV, the projected sensitivity using data from SR0 and SR1 and the development of the full framework required to perform a fit of the below presented model to the data. Some of the work presented in this chapter was developed in parallel to the XENONnT search for solar neutrinos from proton-proton fusion, currently being prepared for submission as [\[3\]](#), and used some of the same intermediate results, such as data selection cuts and studies to constrain background contributions.

At first, [section 5.1](#) describes the detector response model, consisting of the energy reconstruction, the peak detection thresholds and the event building efficiency. Following that, [section 5.2](#) lists the data selection cuts utilized in the search for bosonic

DM and the computation of the signal acceptance. In [section 5.3](#) an in-depth description of the material-originating background radiation is presented, in addition to the optimization of the FV. The background contributions by the decays of ^{212}Pb and ^{214}Pb are detailed in [section 5.4](#), and the impact of various neutron-activated isotopes can be found in [section 5.5](#), while the remaining parts of the background model are summarized in [section 5.6](#). Finally, [section 5.7](#) describes the statistical model and the sensitivity projection XENONnT towards bosonic DM absorption.

5.1 Detector Response Model

To search for monoenergetic energy deposits, such as potential signals from the interaction of bosonic DM, it is necessary to know how they appear in the recorded data, which is determined by the detector response model, detailed in the following section.

5.1.1 Energy Reconstruction

The energy reconstruction is described by two functions, the energy bias, $\Delta E(E)$, that is the difference between reconstructed energy and true energy, and the energy resolution, $\rho(E)$, that quantifies the width of a monoenergetic signal. The energy resolution is taken from [\[112\]](#) and the energy bias is the one derived in [\[207\]](#). Both, bias and resolution, are determined by fitting the energy spectrum of multiple monoenergetic lines with a Skew-Gaussian function

$$f(x) = \epsilon_{S1} \cdot \frac{N}{w\sqrt{2\pi}} \exp\left(-\frac{(x - \xi)^2}{2w^2}\right) \cdot \left(1 + \operatorname{erf}\left(\alpha \cdot \frac{x - \xi}{\sqrt{2}w}\right)\right). \quad (5.1)$$

Here, $\epsilon(S1)$ is the S1 detection threshold, detailed in [subsection 5.1.2](#), the variables w , ξ and α are the parameters that characterize the Skew Gaussian and N is the amplitude of the spectrum. The mean energy, μ , and the width of the Skew Gaussian, σ , can be determined from the above parameters via

$$\mu = \xi + \sqrt{\frac{2}{\pi}} \frac{\alpha w}{\sqrt{1 + \alpha^2}} \quad \text{and} \quad \sigma^2 = w^2 \left(1 - \frac{2}{\pi} \frac{\alpha^2}{1 + \alpha^2}\right). \quad (5.2)$$

The sources of monoenergetic signals used for energy calibration, and their respective energies, are listed in [Table 5.1](#). ^{37}Ar and ^{83m}Kr are specific sources introduced into the detector for calibration purposes. In SR0 and SR1 AmBe calibrations exposed the LXe to high rate neutron radiation, during which many isotopes present in the TPC underwent neutron capture and produced short lived or metastable neutron-activated isotopes. The gamma rays emitted during neutron capture or the consequent decay also serve as sources of monoenergetic events. The neutron activated isotopes include ^{129m}Xe , ^{131m}Xe and ^2H , and are discussed again in [section 5.5](#). The remaining lines originate either from material radioactivity or from intrinsic isotopes naturally found in xenon.

After fitting the energy spectra for SR0 and SR1 data, the resulting energy resolution ρ and bias ΔE are characterized by two empirical functions:

$$\rho(E) = \frac{\sigma}{E} = \frac{a}{\sqrt{E}} + b \quad (5.3)$$

Table 5.1: The monoenergetic peaks used for energy calibration.

Source	Description	Energy [keV]
^{37}Ar	Calibration source	2.82
^{129}Xe	Excited state after inelastic scattering. Not used in SR1	39.6
^{83m}Kr	Calibration source	41.5
^{124}Xe	Naturally occurring isotope. Not used in SR1.	64.3
^{131m}Xe	From neutron activation	163.9
^{129m}Xe	From neutron activation	236.2
^{60}Co	From materials	1173.2
^{60}Co	From materials	1332.5
^{40}K	From materials	1460.8
^2H	Neutron capture of H contamination. Not used in SR1	2224.5
^{208}Tl	From materials and ^{232}Th calibration. Not used in SR0	2614.5

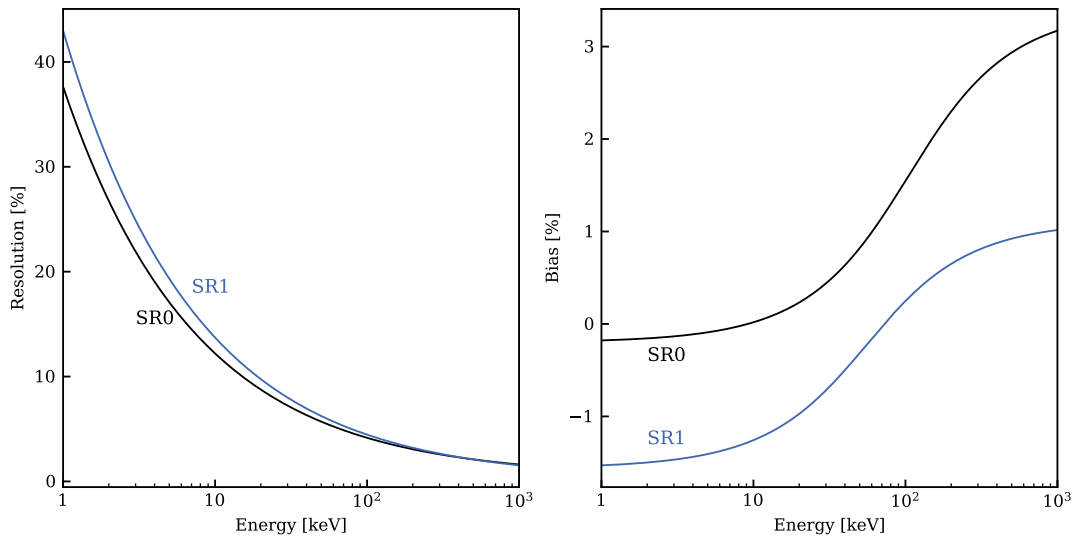
Table 5.2: The parameters of the empirical functions that characterize the energy resolution of reconstruction bias.

	a [$\sqrt{\text{keV}}$]	b	c	d	k [keV^{-1}]
SR0	0.372	0.00436	0.0230	-0.0020	0.010
SR1	0.428	0.00178	0.0170	-0.0156	0.018

and

$$\frac{\Delta E(E)}{E} = \frac{E_{\text{rec}} - E_{\text{true}}}{E} = c \cdot \arctan(k \cdot E) + d. \quad (5.4)$$

The parameters of these two empirical functions are listed in Table 5.2 and the functions can be seen in Figure 5.1. More details on the energy reconstruction in XENONnT can also be found in [208].

**Figure 5.1:** The energy resolution (left) and bias (right) for the SR0 and SR1 analysis.

5.1.2 Energy Threshold

The energy threshold is the lowest recoverable energy and is determined by the ability to measure S1 and S2 signals, which is characterized by the S1 and S2 detection

efficiencies ϵ_{S1} and ϵ_{S2} . The S1 detection efficiency is the probability that a physical signal of a certain number of photons is reconstructed as an S1 signal. In general, this is a function of the S1 area. Conventionally the efficiency is modeled with simulated data, since the information about the initially produced photons, as opposed to the measured ones, is cannot be obtained from data. Instead, single and few photon signals are simulated and propagated through the full chain of event reconstruction. The efficiency per number of photons is the fraction of signals recovered to total simulated signals. However, this method relies on the correctness of simulations.

In XENONnT a data-driven method was introduced to validate the simulation derived S1 efficiency. For small S1 signals, it can be assumed that each PMT hit is the result of a single photon. These single photon signals are used to build a data driven distribution of photon signals that includes the statistics about non-recorded photons, from where individual single photon signals can be sampled. The sampled single photons are propagated through the signal reconstruction process, and the fraction of recovered single photons equals the S1 detection efficiency. The data driven study supported the results of the simulation based study [112].

Validating the simulation derived S2 detection efficiency, ϵ_{S2} , is much simpler. Since each S2 electron leads to a multitude of detected photons, there is a well defined population of recorded SE S2 signals, that can be used to compute a fraction of recovered S2s.

The final challenge is the conversion from S1 and S2 area space to energy. For the S1 and S2 efficiencies this is also done via simulations, by sampling S1 and S2 areas of ER interactions and reconstructing their corresponding energies. The resulting efficiencies are shown in [Figure 5.2](#). Both efficiencies reach 100 % for energies greater than ~ 3 keV, in particular the S2 efficiency will be effectively equal one for energies where the detector can record S1 signals. This is the same reason why only ϵ_{S1} , and not ϵ_{S2} , appears in [Equation 5.1](#).

5.1.3 Event Building Efficiency

The event building efficiency (EBE) quantifies the ability to correctly pair recorded S1 and S2 signals into an event. An event can be misreconstructed when the corresponding S1 or S2 signal is overshadowed by a bigger S1, or S2 respectively, signal that happens in the event time frame by coincidence. In XENONnT the event building is initiated by S2 signals with an area greater than 120 PE, called a triggering S2. The event building algorithm searches in a time window of t_{right} after the triggering S2, and $t_{\text{left}} + t_{\text{max}}$ before the triggering S2 for S1 signals, where $t_{\text{right}} = t_{\text{left}} = 250 \mu\text{s}$ are extension windows and $t_{\text{max}} \approx 2.2 \text{ ms}$ is the maximum possible drift time of electrons in the TPC. The largest S1 and S2 signal inside this window are labeled the main signals and used to reconstruct an event. An event is misreconstructed, if the main S1 and S2 do not have the same physical origin, which happens if one of these signals is overshadowed by a bigger signal. Therefore, the EBE is the inverse of the probability for a bigger S1 or S2 to happen inside the time window $\Delta t = t_{\text{left}} + t_{\text{max}} + t_{\text{right}}$. An S2 is overshadowed if there is a larger S2 within Δt of its *corresponding* S1, and similarly, an S1 is overshadowed if there is a larger S1 within Δt of its corresponding

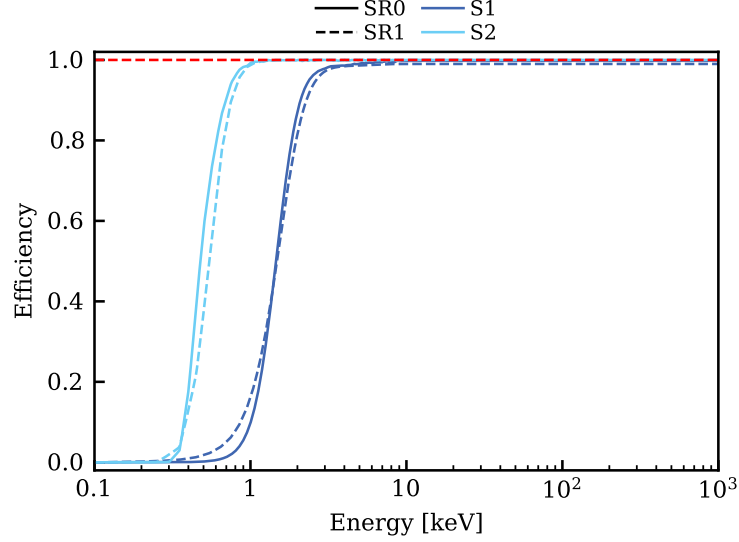


Figure 5.2: The S1 and S2 detection efficiencies for SR0 and SR1, derived via simulations and projected onto energy scale.

S2 signal. A sketch of the correct event building, and the case of a larger S2 and S1 happening is shown in [Figure 5.3](#).

The EBE is heavily dependent on the rate of spurious S1 and S2 signals, called the detector *ambiance*. The *ambiance* and shadow cuts mentioned in [subsection 2.3.5](#) are specific cuts targeting high rate environments in the detector, making them extremely correlated with the EBE. For this reason, other analyses in XENONnT combine the EBE and the *ambiance* and shadow cut into the so-called anti-correlation efficiency. In this study however, the *ambiance* and shadow cuts are not applied, therefore only the EBE is of concern.

A high granularity salting study was performed to compute the EBE in the context of [\[3\]](#). Simulated events, meaning S1 and S2 peaks of the same physical origin, were salted into recorded detector data. These augmented data were propagated through the standard XENONnT processing pipeline, and the fraction of the salted events that was correctly reconstructed as an event was determined. Since this method uses the exact detector *ambiance*, it is a very precise tool to obtain the EBE. However, this salting study is only available for the energy interval of $[0, 200]$ keV. For this reason, the EBE for the full energy range of $[0, 1000]$ keV used this study is estimated using the peak shadow method, described in the following.

Any S1 (S2) peak has a so-called *shadow*, which covers a time window of $t_{\text{blind}} = t_{\text{left}} + t_{\text{max}} + t_{\text{right}} \approx 2.7$ ms following (preceding) it, where peaks with an area smaller than itself will not be considered as the main S1 (S2) of the event. Assuming all peaks of all areas are distributed uniformly in time inside one run, the probability f_{shadow} for an S1 (S2) peak to be overshadowed is therefore the number of S1 (S2) peaks with larger area $N_{\text{S1}>} (N_{\text{S2}>})$, multiplied with t_{blind} . The estimated EBE ϵ_{EBE} can be

computed as

$$\epsilon_{\text{EBE}}(S1, S2) = 1 - f_{\text{shadow}} = 1 - (N_{S1>} + N_{S2>}) \cdot \frac{t_{\text{blind}}}{t_{\text{tot}}}, \quad (5.5)$$

where t_{tot} is the total time considered when counting $N_{S1>}$ and $N_{S2>}$.

The distribution of $\epsilon_{\text{EBE}}(S1, S2)$ for various values of S1 and S2 is mapped in [Figure 5.4](#). Note, that the value was only computed for specific combinations of S1 and S2 from the so-called ER band, because only events within this band are relevant to the study in this chapter. For a more detailed description and the derivation of this band refer to [section 5.2](#). In addition, [Figure 5.4](#) shows the equienergy lines of selected energies, computed via [Equation 2.22](#). In SR0, the EBE is mainly dependent on the S1 area, which intuitively makes sense, since ER events have a large charge yield, compared to their light yield. If an event with a large S1 happens, chances are that its S2 is also large enough to decrease the probability of it being overshadowed. In SR1 at low energies in particular, the opposite is visible. The EBE is suppressed in large because of the low chance of detecting the correct S2. This can be explained by the high amount of PI contamination in SR1.

The projection from S1 and S2 to energy is again done via [Equation 2.22](#). However it should be noted that the EBE estimation presented here works on peak-level, not on event-level, which means that the S1 and S2 areas are uncorrected areas¹. The resulting EBEs for SR0 and SR1 as functions of energy $\epsilon_{\text{EBE}}(E)$ can be seen in [Figure 5.5](#) in blue, and in addition, the results of the aforementioned salting study are shown in black.

The uncertainty of the peak shadow EBE includes the systematic uncertainty of 1 % that was determined in the salting study, and is inflated by the maximum difference between the two calculated EBEs in the energy range where both are defined. Using this conservative estimation of uncertainty ensures that the salting derived EBE is always within the uncertainty band of the peak shadow derived EBE. For SR0, the salting and peak shadow derived EBEs lie slightly apart, especially at lowest energies. However, in SR1 the two curves match astonishingly well.

¹Since corrections depend among others on the z coordinate, which is basically the time difference between paired S1s and S2s, the corrections as they are defined are not applicable to peak areas.

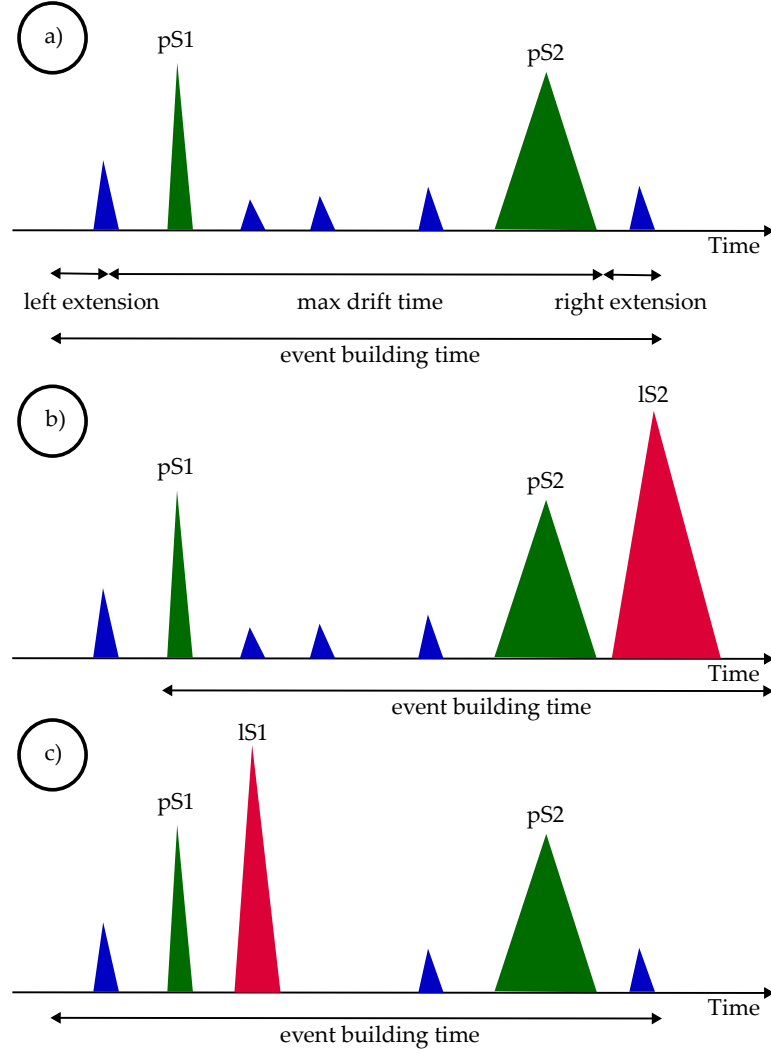


Figure 5.3: A sketch of XENONnT event building. In case (a), the primary signals $pS1$ and $pS2$ (green) are the biggest signals within Δt of the $pS2$. Spurious other signals (blue) are smaller and ignored by the event building. Case (b) shows the $pS2$ being overshadowed by a larger $S2$, labeled $IS2$ (red). Since the $IS2$ is within Δt of the $pS1$, it will be taken as the main $S2$ of the reconstructed event. Case (c) shows overshadowing by a large $S1$, $IS1$ (red). The $pS1$ is not the biggest $S1$ within Δt of the $pS2$, and therefore not taken as the main $S1$ of the event.

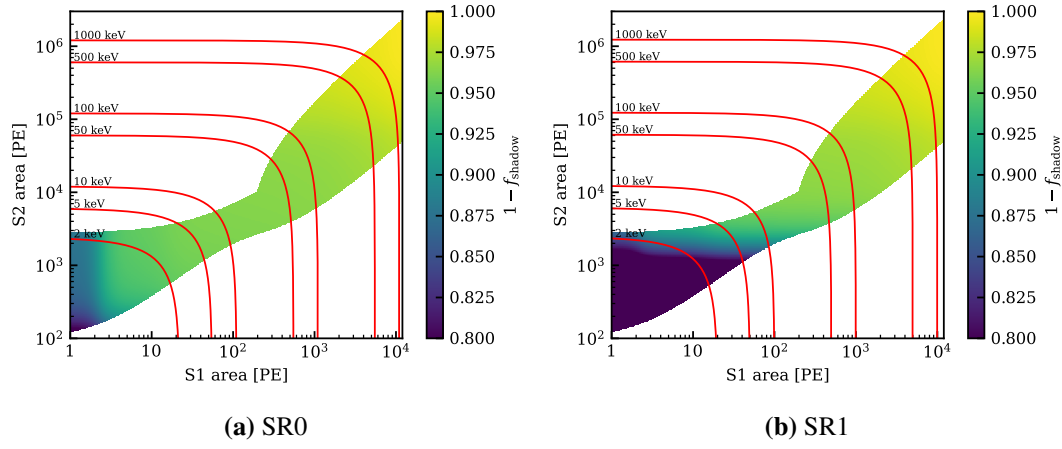


Figure 5.4: The EBE estimated via the peak shadow method for SR0 (left) and SR1 (right) as a function of the S1 and S2 size. In addition, the red lines visualize equi-energy contours for selected energies.

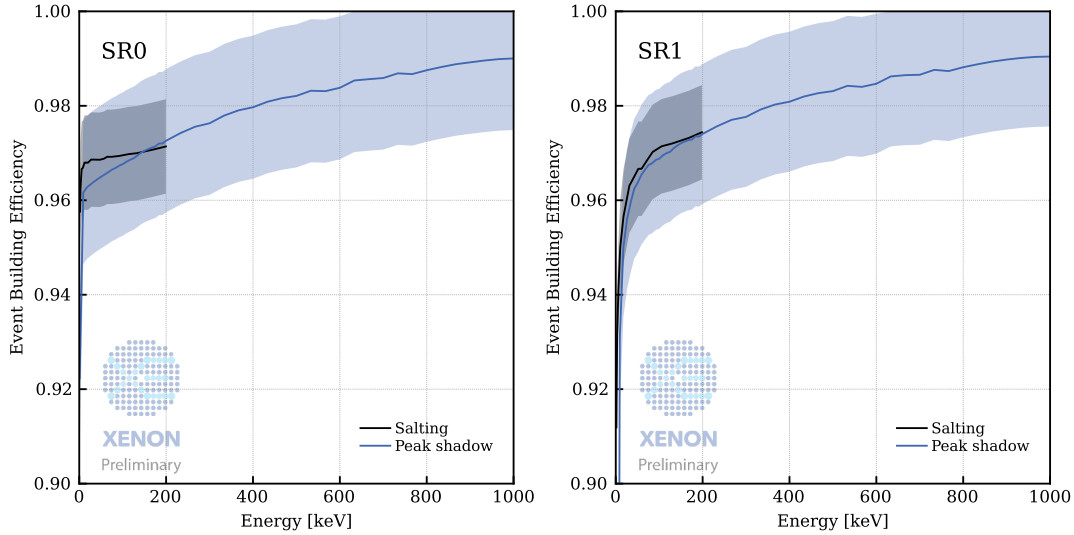


Figure 5.5: The EBE estimated via the peak shadow method as a function of recoil energy for SR0 (left) and SR1 (right). In addition, the low energy salting results for the EBE are shown in black.

5.2 Signal Acceptance

In XENONnT analysis there are many different data quality *cuts*, as described in [subsection 2.3.5](#). Their purpose is the removal of background events, however, they also inevitably remove some portion of signal-like events, and therefore have a certain *signal acceptance*. The acceptance is the probability for a signal-like event to survive, meaning not be removed, by any of the applied cuts. The study presented in this chapter adapted S1 and S2 quality cuts that were developed in for the solar neutrino analysis detailed in [3] and in addition introduces individual ER band, SS and FV selections. The optimization of the FV is explained later in [section 5.3](#) jointly with backgrounds originating from detector materials. This material background provides the vast majority of events in the vicinity of the TPC wall and PMT arrays. The purpose of the cut acceptance study, if performed on data as opposed to simulations, is to determine the acceptance of signal-like events from calibration sources. To not have the calibration events overshadowed by material background events, in the following cut acceptance studies, a preliminary FV selection was applied to the data, in the form of a cylinder defined by $-134.2 \text{ cm} < z < -13.6 \text{ cm}$ and $R < 60.7 \text{ cm}$.

There are two datasets used to compute the cut acceptance: A signal-like simulation and calibration data. The signal-like simulation consists of monoenergetic ER-like energy depositions, uniform in energy up to 1000 keV and uniform in space, simulated with the *fuse* software. The calibration data is the combined set of events from ^{220}Rn calibrations below 200 keV, and ^{232}Th calibrations above 200 keV, to avoid the γ peak of the ^{212}Pb daughter in the ^{220}Rn data. The spectra of the available data are shown in [Figure 5.6](#) for SR0 (left) and SR1 (right). In total, after FV, ER band and SS selections, there are approximately 3.2×10^6 (2.9×10^6) events available in the SR0 (SR1) simulation dataset and 4.6×10^4 (1.5×10^5) in the calibration dataset. Note that not all of the available ^{220}Rn calibration data is used, because the binning to compute the signal acceptance is optimized to minimize the uncertainties per bin, and if the imbalance between the numbers in the ^{220}Rn and ^{232}Th datasets is too large, the bin size will be determined mostly by the ^{220}Rn data. To counteract this, the amount of ^{220}Rn events used to compute the final cut acceptance is capped to be at maximum two times the amount of ^{232}Th events.

5.2.1 ER Band Selection

As explained in [subsection 2.3.3](#), there is a difference in relative S1 to S2 strength for events caused by an NR compared to those caused by an ER. In addition, some ACs can be identified via non-matching S1 to S2 ratios. [Figure 5.7](#) shows the (corrected) S1 and S2 strength distributions for ER events from ^{220}Rn and ^{232}Th calibrations (blue) and NR events from AmBe calibrations (red). The separation between ER and NR bands is visible especially at low energies. The ER band selection in this study is not meant to reduce NR background, but rather reject AC events with unphysical combinations of S1 and S2 strength, while keeping a signal acceptance of 100 %. The red lines in [Figure 5.7](#) achieve this. The selection is a combination of the ER band selection from [54] for $cS1 < 200 \text{ PE}$ and the band selection used in the search for

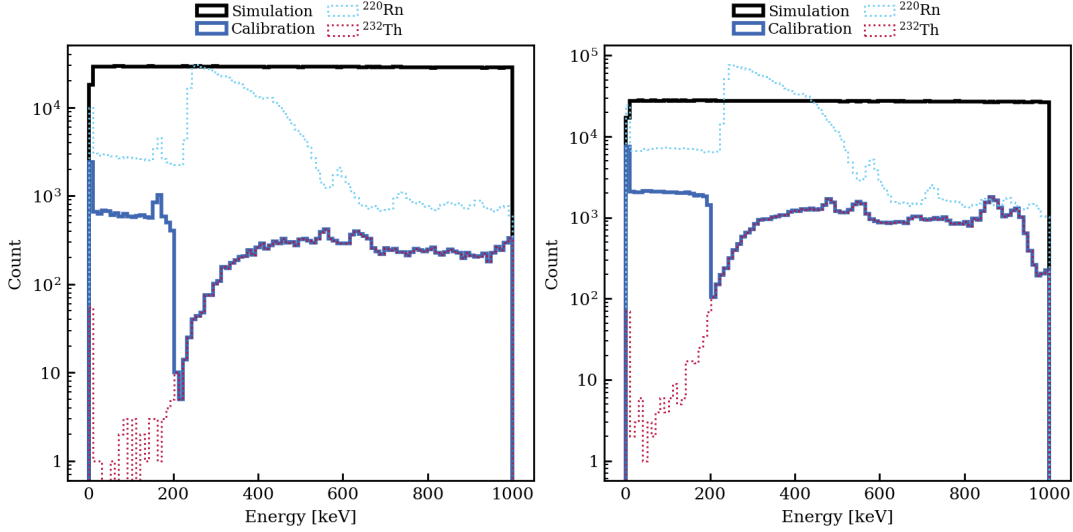


Figure 5.6: The spectrum of the SR0 (left) and SR1 (right) data available for the cut acceptance study, with FV, ER band and SS selections applied. The calibration data is a combination of the ^{220}Rn events below 200 keV and ^{232}Th events above 200 keV.

$0\nu\beta\beta$ for bigger signals. The band boundaries are

$$f_{\text{bound}}(cS1) = \begin{cases} p_1 + p_2 \cdot cS1 - p_3 \cdot cS1^2 + p_4 \cdot cS1^3 & \text{if } cS1 \leq 200 \\ m \cdot (cS1 - 200) + c & \text{if } cS1 > 200 \end{cases}, \quad (5.6)$$

with parameters $p_1 = 92.27, p_2 = 30.18, p_3 = -0.1409, p_4 = 0.0003, m = 4, c = 2718$ for the lower and $p_1 = 2806, p_2 = 52.00, p_3 = -0.1163, p_4 = 0.0002, m = 200, c = 10398$ for the upper boundary. The units of these parameters were omitted for readability. All following cut acceptance studies have the ER band selection applied as a prior selection, and the signal acceptance of the ER band can be assumed to be 100 %.

5.2.2 Single Scatter Selection

The signals from bosonic DM are expected to be SS events, as opposed to MS events. In XENONnT the distinction between SS and MS events is based on the ratio between main and alternative S2 size. Figure 5.8a shows the main and alternative S2s for the signal-like simulations (left), as well as events that were explicitly simulated as MS (right). It is clearly visible that the vast majority of MS events has an alternative S2 of significant size, justifying the choice of drawing the SS boundary in that parameter space.

The distribution of main and alternative S2 strength for ^{220}Rn calibration data is shown in Figure 5.8b. It is important to note that the ^{220}Rn calibration events contain mostly ^{212}Pb β - γ cascades, that pose a very particular event topology, especially in regard to SS MS discrimination. More information on the ^{212}Pb and ^{214}Pb backgrounds can be found in section 5.4. Figure 5.8b is meant to visualize two things: First, the SS boundary in alternative S2 space is indeed separating MS also in measured data, not only simulation. And second, that a small amount of events above the boundary can

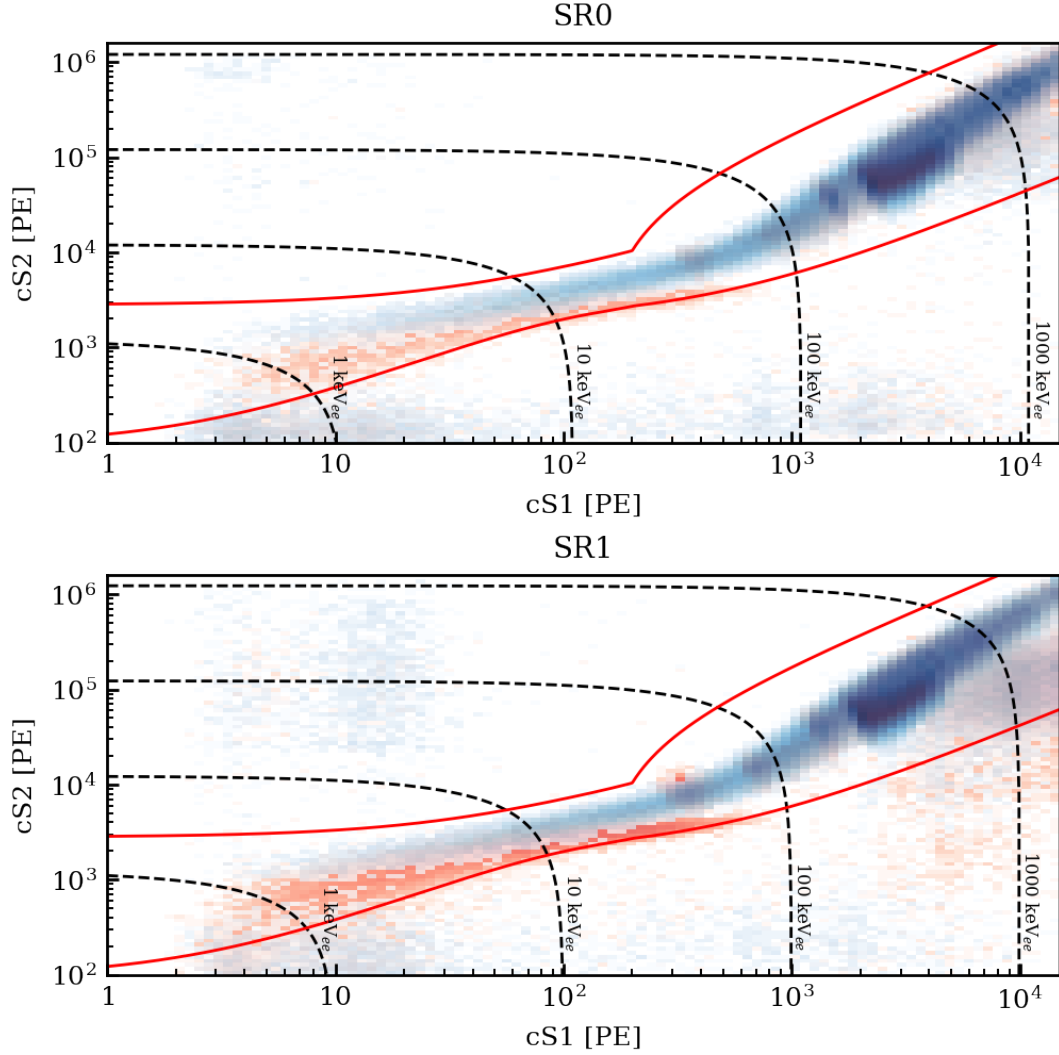


Figure 5.7: The cS1 cS2 distribution of NR (red) and ER (blue) calibration events for SR0 (top) and SR1 (bottom). In addition, the ER band selection used in this study is shown in red.

still be classified as SS, provided they pass additional criteria placed on their width. If an alternative S2 is too wide in time for its area, it is probable that it does not originate from a single interaction, like a MS, but rather from the merging of multiple signals. Events with such alternative S2s are again classified as SS.

Figure 5.9 shows the acceptance, the fraction of surviving signal-like simulated events, computed as a function of energy and spatial position for SR0 (SR1 can be found in Figure C.2). No spatial bias is observed in the SS classification. There is a very slight decrease in acceptance for increasing energies, however, this effect is so minor as to be negligible. In general, for all further cut acceptance studies, the SS cut is applied and assumed to have an acceptance of 100 %.

5.2.3 Derivation of the Cut Acceptance

The FV, ER and SS selections are somewhat special, in that they are assumed to have a signal acceptance of 100 %. They are always applied apriori in the following sections

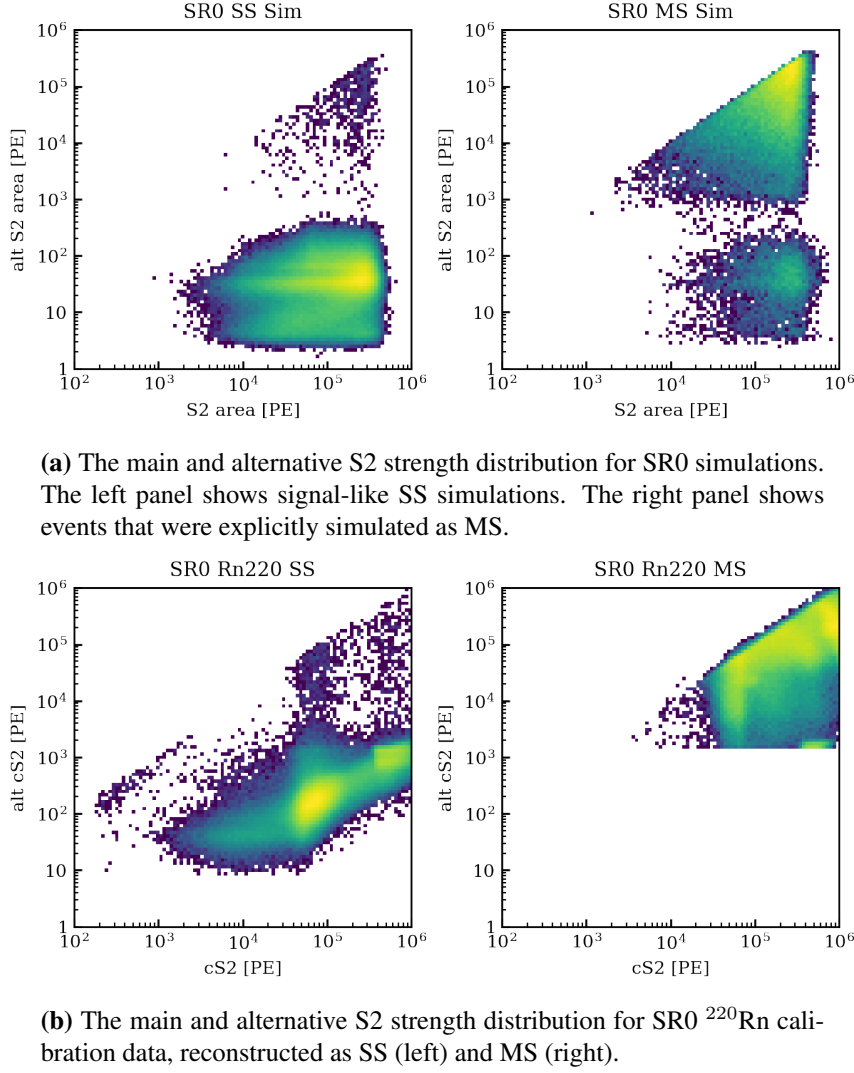


Figure 5.8: The main and alternative S2 strength distributions of different datasets: ^{220}Rn calibration data *reconstructed* as SS or MS (top), and simulation data *simulated* as SS or MS (bottom). The same for SR1 can be found in [Figure C.1](#).

about the remaining data quality cuts.

To maximize the signal acceptance it is important to keep the over all number of applied cuts low, and specifically to avoid having cuts that target the same population of background events. For this reason, the acceptance of a data quality cut is computed as the $N - 1$ acceptance, meaning all other quality cuts are already applied to the data, and a cut does not appear improved if it removes a population already targeted by another cut. The $N - 1$ acceptance is defined as

$$A = \frac{\#(\text{events after } N \text{ cuts})}{\#(\text{events after } N - 1 \text{ cuts})}. \quad (5.7)$$

We require the signal acceptance to be a function of the event energy. To express the acceptance as a function of a parameter, the first step is to construct a histogram of all events according to the chosen parameter and then compute [Equation 5.7](#) inside

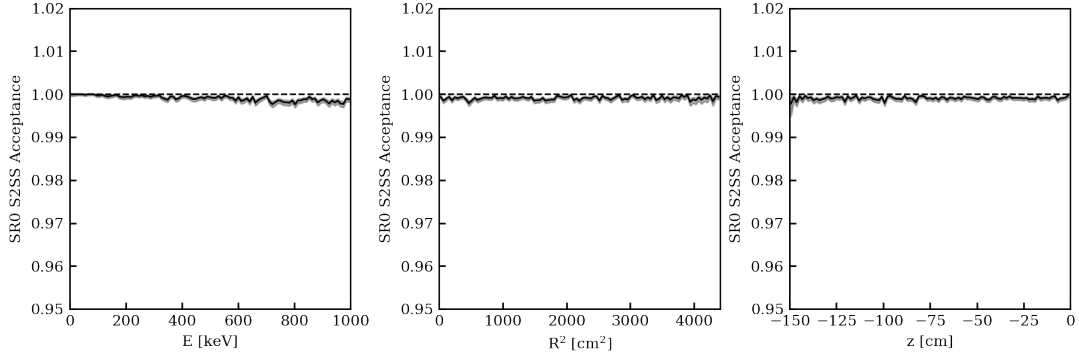


Figure 5.9: The acceptance of the SS selection in SR0, determined on signal-like simulations of SS events, as a function of energy (left), radius squared (center) and vertical position (right).

Table 5.3: The datasets used to compute the cut acceptance.

Cut	Primary dataset	Supporting dataset
S1 width	Calibration	-
S1 AFT	Calibration	Simulation
S1 max PMT	Simulation	Calibration
cS2 AFT	Calibration	-
S2 pattern	Calibration	-
S2 width	Simulation	-

of each bin. The resulting acceptance, as a function of the parameter, depends on the chosen binning for the histogram: a too wide binning will not capture dependencies of the acceptance on the chosen parameter and a too fine binning will be dominated by noise. The binning in this study is chosen by the method introduced in [209], that is based on minimizing the mean integrated square error (MISE) between the histogram and the underlying data distribution via kernel density estimation (KDE).

For the uncertainty of the acceptance, the Clopper-Pearson confidence interval is used, as that is a conservative choice in the case of low statistics [210]. Finally, when the cut acceptance can be computed on both datasets, simulation and calibration, the uncertainty is combined via the following procedure: One dataset is defined as the primary dataset, the other as the supporting, and the acceptances A_{prim} , A_{sup} and acceptance uncertainties σ_{prim} , σ_{sup} are computed. We define a measure $\chi^2 = (A_{\text{prim}} - A_{\text{sup}})^2 / (\sigma_{\text{prim}}^2 + \sigma_{\text{sup}}^2)$. If $\chi^2 < 1$, the uncertainties are already large enough and the acceptance computed on the primary dataset is used. Otherwise, a systematic uncertainty σ_{sys} is defined as

$$\sigma_{\text{sys}}^2 = (A_{\text{prim}} - A_{\text{sup}})^2 - \sigma_{\text{prim}}^2 - \sigma_{\text{sup}}^2, \quad (5.8)$$

which leads to $\chi^2 = 1$ if σ_{sys}^2 is added to the denominator. If $A_{\text{prim}} > A_{\text{sup}}$, then σ_{sys} is added to the upper uncertainty, and if $A_{\text{prim}} < A_{\text{sup}}$, then σ_{sys} is subtracted from the lower uncertainty, meaning the Clopper-Pearson intervals are inflated by this additional systematic uncertainty. The primary and secondary datasets per cut are listed in Table 5.3.

5.2.4 S1 Width Cut

The S1 width cut is defined as a boundary on the 50 % and 90 % percentiles of the temporal width of an S1 signal, dependent on its respective area. At low energies, this cut is meant to remove SEs that were misclassified as S1s and at higher energies this cut targets S1s with large afterpulses.

The simulation of S1 waveform parameters, in particular the width, does not match the data. A comparison of simulation data and calibration data can be found in [section D.2](#). For this reason, the cut parameters were derived from ^{220}Rn calibration data, and the acceptance for this study will also be calculated from calibration data. In [Figure 5.10](#) the 50 % and 90 % width of SR0 ^{220}Rn calibration data is shown, including the cut boundaries in red. The 50 % boundary ends at S1 strength of 2000 PE and the 90 % boundary ends at S1 strength of 100 PE. Events at higher S1 areas pass that boundary, meaning all events with an S1 area above 2000 PE automatically pass this cut.

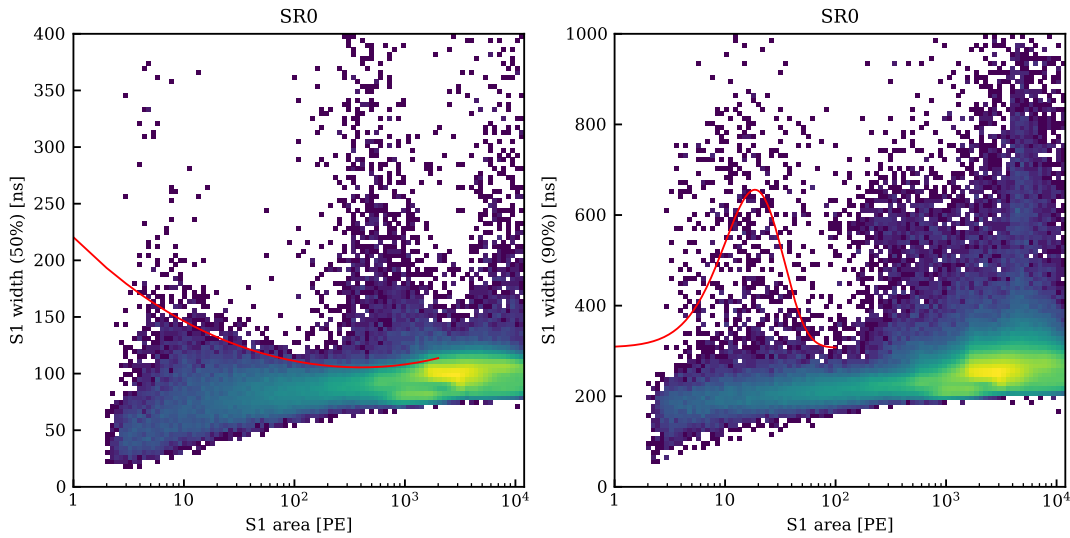


Figure 5.10: The boundaries of the S1 width cut, shown over SR0 ^{220}Rn calibration data. The same figure for SR1 is shown in [Figure C.3](#)

At low areas, the cut removes multiple high width outliers, but not all of them, particularly visible in the 90 % case. Since the cut boundaries were derived on data that used the ambiance cut, explained in [subsection 2.3.5](#), and the ambiance cut was shown to remove SEs misclassified as S1, it can be assumed that the additional outliers removed by this cut, are also misclassified SEs. This is especially important for the low energy region, since the ambiance cut is not applied in this study. At higher areas there is an overabundance of high width S1s, which are mostly normal S1s that have some additional small signals merged onto them. Since these do not negatively impact the energy reconstruction, there is no immediate reason to remove them. The S1 width cut has an average acceptance of 99.4 % in SR0 and 99.9 % in SR1, and the acceptance as a function of energy, radius squared and vertical position can be seen in [Figure C.4](#).

5.2.5 S1 Area Fraction Top Cut

The S1 AFT is the fraction of the total signal that is measured by the top PMTs, which can be modeled with a binomial PDF that is dependent on the position of the event. The cut is based on the comparison between the measured AFT, and the expected one which can either be derived from simulations, or from the statistical majority of the measured data. An in-depth description of this cut can be found in [108]. This cut targets events where a significant amount of the S1 signal got lost, or ACs, where the mispairing of the S1 and S2 leads to a wrong inferred z position, and therefore a mismatch between expected and observed AFT. Since the data derived AFT maps serve as input for the S1 AFT distribution of the *fuse* simulations, this cut should work the same on simulations and calibration data, enabling the usage of both datasets. The S1 AFT cut has an average acceptance of 98.2 % in SR0 and 98.1 % in SR1, and the acceptance as a function of energy, radius squared and vertical position can be seen in Figure C.5.

5.2.6 S1 Max PMT Cut

A minor amount of S1s signal show an uncharacteristically large contribution from a single PMT. The cause of this behavior is not yet known with certainty, but one possible explanation is for example Cherenkov radiation inside the PMT quartz window. The cut boundary of this cut depends on the vertical z -position, the area of the PMT with the largest signal and the total S1 area. Since such behavior can also be caught by the S1 pattern and S1 AFT cuts, the boundaries of this cut are very loose to ensure a high signal acceptance. The S1 max PMT cut has an average acceptance of > 99.9 % in SR0 and > 99.9 % in SR1, and the acceptance as a function of energy, radius squared and vertical position can be seen in Figure C.6.

5.2.7 Corrected S2 Area Fraction Top Cut

The vertical height where the S2 photons are emitted is confined by the liquid interface and the anode, which leads to a small range of typical S2 AFT values, roughly around 76 %. The GXe between anode and top PMT array can also interact with incoming particles, leading to so-called *gas events*. Since the light from gas events has to cross the anode to reach the bottom PMTs, their S2 AFT is in general greater than that of normal events, which the S2 AFT cut tries to utilize to filter them out.

However, in particular in SR1 there is a time dependent concentration of PI contaminants inside the TPC, that absorb photons traversing the LXe, as explained in [subsection 2.3.4](#). This PI contamination leads to fewer photons reaching the bottom PMTs, which in turn leads to a higher AFT. This time dependent change in S2 AFT is corrected for, and this cut operates on the corrected S2 area fraction top.

The modeling of the S2 AFT does not match between data and simulations, which is why this cut needs to be derived and validated on calibration data. A comparison of simulation data and calibration data can be found in [section D.3](#). The cut boundaries are defined on the 5 % and 95 % percentiles of the corrected S2 AFT as a function of the corrected S2 area disregarding the electron lifetime correction. Since the average S2 AFT depends on detector and extraction field geometry, the S2 area should

be corrected for this effect. But the vertical depth of the event, and therefore potential electron loss during drift, does not impact the parameter space of this cut. The boundaries are derived using ^{220}Rn data an AmBe data, and the acceptance here is computed using the calibration dataset. The S2 AFT cut has an average acceptance of 98.7 % in SR0 and 99.4 % in SR1, and the acceptance as a function of energy, radius squared and vertical position can be seen in [Figure C.7](#).

5.2.8 S2 Pattern Cut

The S2 pattern cut works similar to the S1 AFT cut, in the sense that it is supposed to reject events that showcase an unlikely distribution of detected light signals in the PMTs. The distribution of S2 photons detected by the PMTs can be modeled with a likelihood resembling [section 3.1](#). Events consisting of mergers of physically unrelated signals, like unresolved MS or pile-up of multiple small signals, will have a lower pattern likelihood than normal events. The actual likelihood differs from [section 3.1](#), since the per-PMT probabilities are modeled with a Gaussian smeared Poissonian function, instead of a pure Poissonian, and also consider double photo electron emission of the PMTs.

The lightcurves required for the likelihood are not determined via simulations, but by training a CNN on the *reconstructed positions* and observed hitpatterns. In particular, this means that the S2 position is an input for this network, signifying that the such derived data driven lightcurves can not be used to train the position reconstruction models described in [chapter 3](#). However, as long as the position reconstruction is in itself consistent, the resulting data driven S2 pattern likelihood can be derived via CNN for all position and hitpattern combinations, given the position was derived by the exact same position reconstruction model. Sadly, there is still some inconsistency observed between the data driven lightcurves, and the ones modeled in the *fuse* simulations, hinting to (minor) not yet modeled optical effects. The important point here is that the S2 hitpattern distributions do not match between data and simulations, which means only calibration data can be used to evaluate the S2 pattern cut. The S2 pattern cut has an average acceptance of 99.9 % in SR0 and 99.7 % in SR1, and the acceptance as a function of energy, radius squared and vertical position can be seen in [Figure C.8](#).

5.2.9 S2 Width Cut

In [subsection 2.3.3](#) it is described how the electron clouds that produce the S2 signals drift through the TPC, guided by the electric drift field. During this drift, the electrons are subject to vertical diffusion, characterized by the vertical diffusion constant D_V , which broadens the cloud in space with increasing drift time, and therefore increases the temporal width of the observed S2 signal. Using this relation, it is possible to model the S2 width as [\[112\]](#)

$$r_{50}^{\text{mod}} = \sigma_{50} \sqrt{\frac{2D_V(t_{\text{drift}} - t_{\text{gate}})}{v_{\text{drift}}}}, \quad (5.9)$$

where t_{drift} is the drift time, t_{gate} is the time the electrons travel from the gate to the liquid interface, v_{drift} is the drift velocity and $\sigma_{50} \approx 1.349$ is the conversion from a

Gaussian standard deviation to the range covering 50 % of the total wave-form. The values D_V , t_{gate} and v_{drift} are inferred from $^{83\text{m}}\text{Kr}$ calibration data. A deviation of the observed width from Equation 5.9, called the *normalized width*, can be defined as [112]

$$r_{50}^{\text{norm}}(t_{\text{drift}}) = \frac{(r_{50})^2 - (r_{50}^{\text{SE}})^2}{r_{50}^{\text{mod}}(t_{\text{drift}})^2}, \quad (5.10)$$

where r_{50}^{SE} is a correction term introduced to match the width of very small S2 signals. In principle, Equation 5.10 captures the deviation of the measured S2 width from the S2 width modeled from the drift time. An r_{50}^{norm} that differs from one indicates that the drift time is wrong, and therefore the pairing of the S1 and S2 signal, meaning the event is an AC event.

The S2 width is different underneath the perpendicular wires, one of the reasons why the near wire selection described in subsection 3.6.2 is important. For the near wire events, there exists a different normalized S2 width model, that is by now also implemented in the *fuse* simulation framework. This, in principle, allows us to compute the S2 width cut acceptance on simulated data. The S2 width cut is defined in the normalized S2 width and S2 area parameter space, and derived by fitting the 1 % and 99 % percentiles of a simulated dataset. Figure 5.11 shows the distribution of normalized width and S2 area for ^{220}Rn calibration data (left), ^{214}Pb simulation data (center, see section 5.4 for details on the simulation), and the signal-like simulations (right) for the far-wire region, meaning the one unaffected by the perpendicular wires.

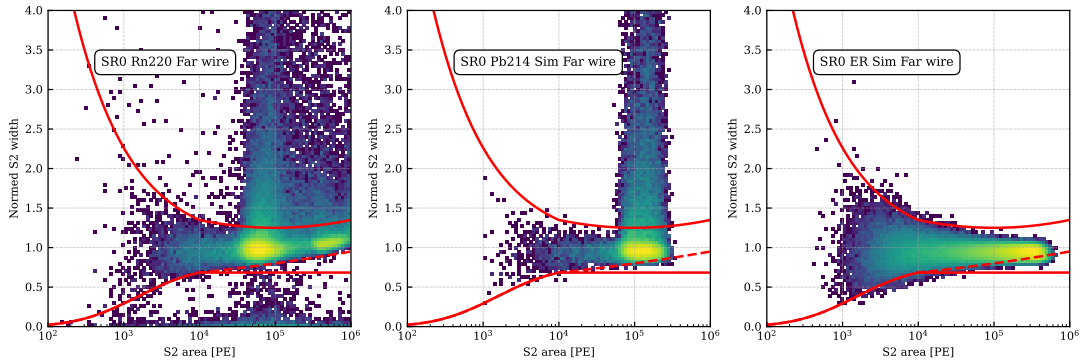


Figure 5.11: The distribution of the normalized S2 width against the S2 area for SR0. The shown datasets are ^{220}Rn calibration data (left), ^{214}Pb simulation data (center) and the signal-like simulations (right). In red are the cut boundaries: The red dashed is the cut definition used in [3] and the red solid is the new cut boundary, adjusted for this study. The same for SR1 can be seen in Figure C.9.

There are several issues visible: The calibration dataset and the simulated ^{214}Pb data show a large population of higher than expected width events for larger S2 areas. Those events are caused by the fact that ^{212}Pb and ^{214}Pb decays contain cascading branches, described in section 5.4. In such cascades, physical MS events can happen so close together in time and space as to be inseparable by the event reconstruction. However, the overlap of two independent energy deposits widens the S2 signals, leading to the observed increased normalized width. Therefore, using the calibration data for the cut acceptance would vastly underestimate the acceptance on signal-like events. However, the *fuse* simulation also slightly mismodels S2 signals: For the

largest energies, the simulated S2 width stays constant, while the main population in the calibration data increases. The red dashed line shows the cut boundary inherited from the analysis in [3], that mimics the increase of the S2 width at higher energies, is proper for data. But since this is not well modeled, the boundary would remove a large population of simulated events at high energies, leading to an underestimation of the cut acceptance. This is also visualized in Figure C.11, where the acceptance of the simulated data using the default cut boundaries is shown. As a solution, the cut boundary was adjusted for this study to the solid red line shown in Figure 5.11. Figure 5.12 shows the cut acceptance derived on simulated data using this new cut boundary. The cut acceptance is much more realistic than the one using the [3] definition, which in turn is shown in Figure C.11.

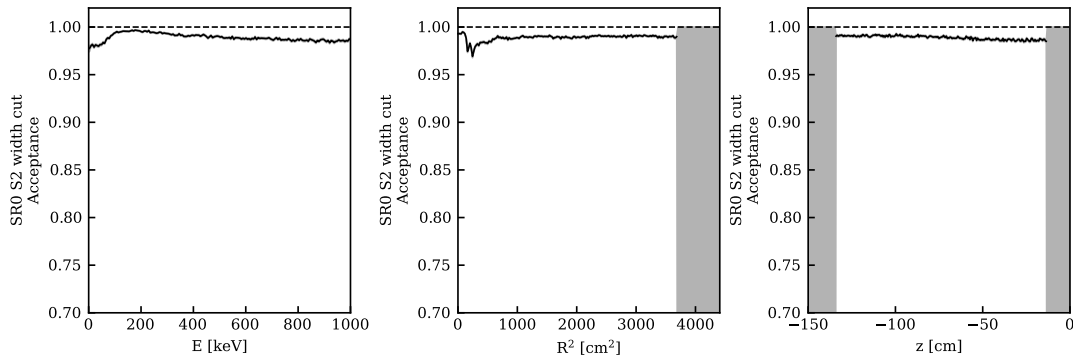


Figure 5.12: The S2 width cut acceptance for SR0 as a function of energy (left), squared radius (center) and vertical height (right), using the new boundary definition. The acceptance for SR1 is shown in Figure C.10. The acceptance using the cut boundary definition from [3] is shown in Figure C.11.

The uniformity of the normalized width in simulations, in combination with the observed increase in data, points to some additional unknown effect that inflates the S2 width at large energies. One possible explanation could be increased diffusion, driven by the Coulomb repulsion among the electrons inside the individual clouds, that is also discussed in subsection 3.6.5. The S2 width cut has an average acceptance of 98.9 % in both SR0 and SR1, and the acceptance as a function of energy, radius squared and vertical position can be seen in Figure 5.12 for SR0 and Figure C.10 for SR1.

5.2.10 Cumulative Cut Acceptance

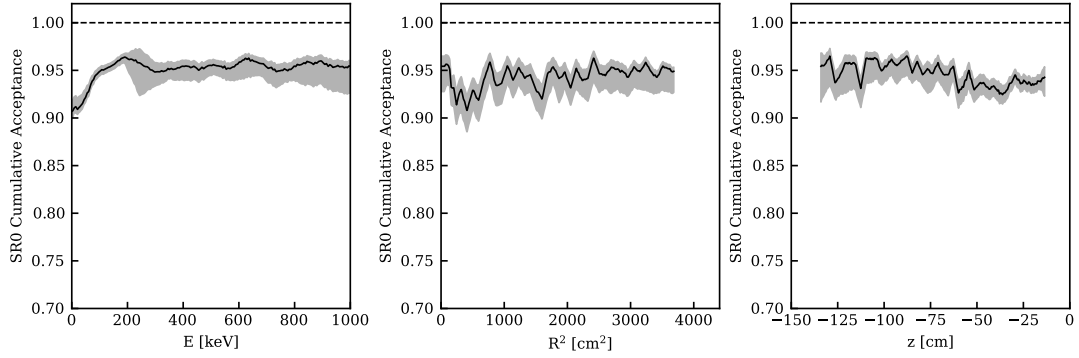
The final signal acceptance A_{tot} is the cumulative acceptance, the product of the individual cut acceptances $A_{\text{tot}} = \prod_i (A_i)$. The relative uncertainty of the cumulative acceptance is the root sum square of the individual relative uncertainties, making the total final acceptance uncertainty

$$\sigma_{\text{tot}} = \sqrt{\sum_i \left(\frac{\sigma_i}{A_i} \right)^2} \cdot A_{\text{tot}}, \quad (5.11)$$

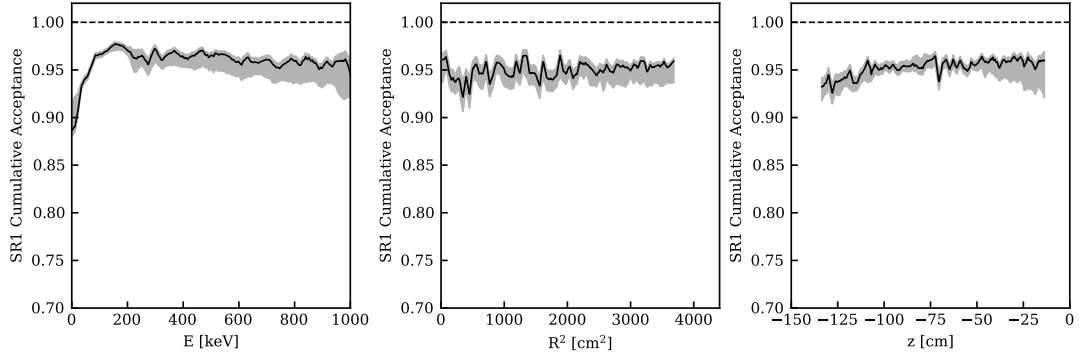
which is computed individually for the lower and upper uncertainties.

The resulting total cut acceptances for SR0 and SR1 are shown in Figure 5.13a and

Figure 5.13b. The acceptance is flat, bar statistical fluctuations, for all radii and heights.



(a) The cumulative cut acceptance for SR0 as a function of the energy (left), the squared radius (center) and the vertical height (right).



(b) The cumulative cut acceptance for SR1 as a function of the energy (left), the squared radius (center) and the vertical height (right).

Figure 5.13: The cumulative cut acceptances.

The final challenge is the distinction between real underlying effects varying the cut acceptance, and fluctuations from limited statistics. One approach is to fit an empirical function to the acceptance. However, there is no known function that catches all the actual physical effects, while disregarding the statistical noise, which makes the choice of function somewhat arbitrary in the end. In addition, an empirical function would have to be expressive enough to model the differences between the two science runs. For those reasons, the choice in this study fell to a smoothing using Gaussian kernel filters. A Gaussian filter is the convolution between a normal distribution and the original data, that is characterized by a standard deviation σ , also called kernel width. The obvious challenge in kernel smoothing is the choice of σ , a too small kernel will follow each small deviation in the acceptance and not smooth the curve at all, and a too large kernel will estimate the full acceptance as a constant and obscure actual physical changes in the acceptance.

To derive an optimal kernel width, a bootstrapping procedure was adopted. For 1000 trials, the energy dependent acceptance curve was modeled with a Gaussian filter of size σ , while 50 % of all energy acceptance value pairs, chosen at random, were hidden to the filter. The filter then predicts the acceptances $\hat{A}(E)$ at the hidden energies E , and the MSE is computed between those predicted values and the actual acceptances

$\langle (\tilde{A}(E) - A(E))^2 \rangle$. If this procedure is performed for many different values of σ , the distribution of MSE can assist with the choice of the kernel width for the smoothing of the final acceptance curve. It was observed that the smallest filters perform the best, as they capture all variation of the acceptance curve, even with as much as 50 % of the data hidden. However, it was assumed that variations in acceptance with a width in energy smaller than the energy resolution were caused by statistical fluctuations, and in general, a wider smoothing kernel should be preferred. The increase in MSE was only marginal until a kernel width of $\sigma = 20$ keV. For this reason a Gaussian smoothing with a standard deviation of $\sigma = 20$ keV was chosen to smooth the acceptance curves of SR0 and SR1. The result is shown in Figure 5.14.

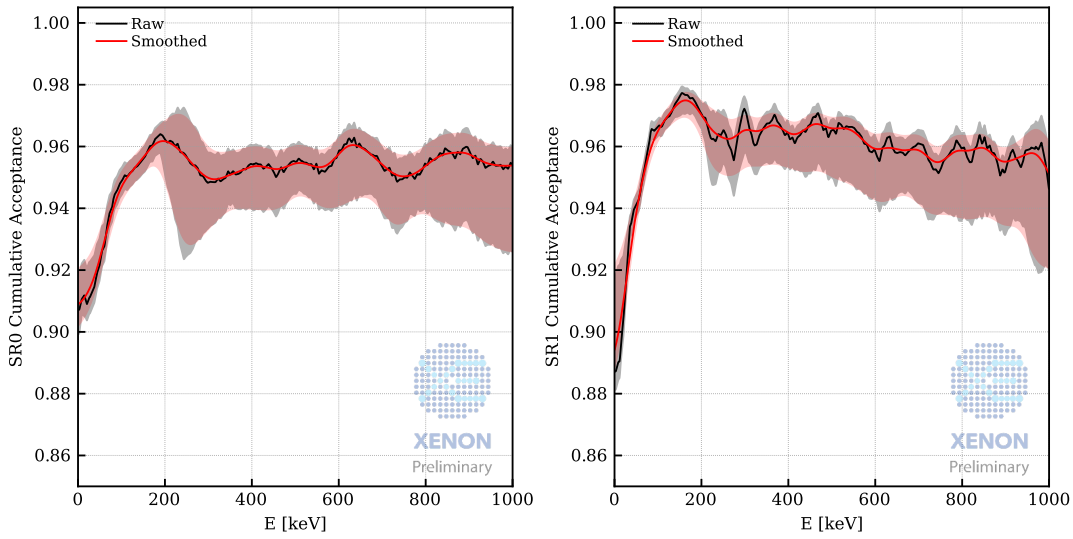


Figure 5.14: The cumulative (black) and smoothed (red) cut acceptances for SR0 (left) and SR1 (right) as function of energy. The smoothing was done with a Gaussian kernel filter of size $\sigma = 20$ keV.

5.3 Material Radioactivity and Fiducialization

5.3.1 Material Originating Background

Trace contamination of radioactive isotopes is present in every material that makes up the detector and amounts to a significant background, either via γ radiation and subsequent Compton scattering, and via α radiation and subsequent (α, n) processes. The particles emitted by the detector materials can reach different depths into the LXe, as outlined by Figure 2.9, and therefore, the extrinsic background caused by material radioactivity has a spatially dependent profile. All materials used in the construction of the XENONnT detector underwent excessive screening to determine the concentration of radioactive isotopes of all detector parts where pieces of the material were screened for their contamination with radioactive isotopes using gamma spectroscopy and mass spectroscopy [101]. Using the information from the material screening, large-scale simulations of the material originating backgrounds were performed, and the result of such a simulation is shown in Figure 5.15.

The most interesting feature of this background lies in its spatial distribution: The

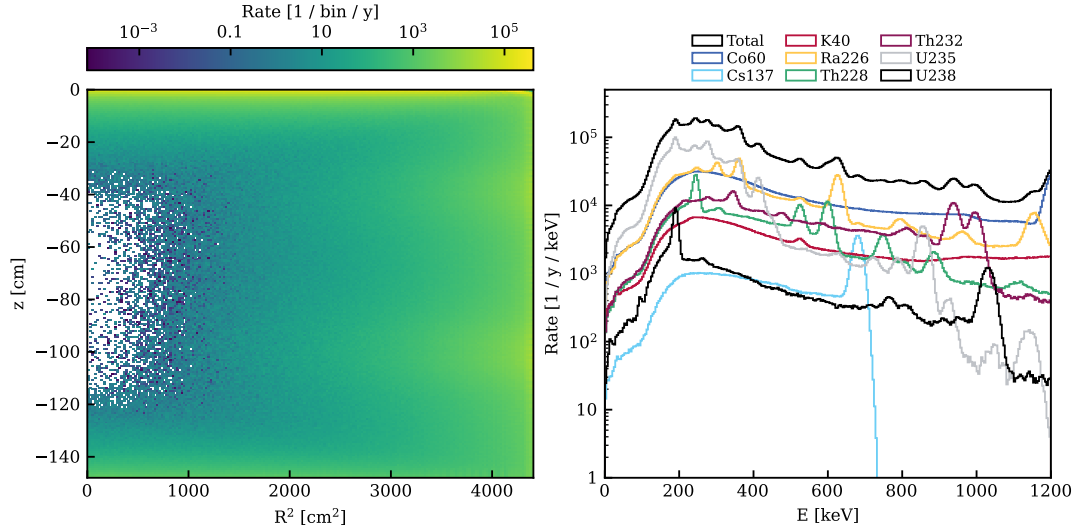


Figure 5.15: Left: The spatial distribution of background events originating from detector material radiation. Based on the LXe stopping power, the vast majority of these events are situated near the TPC walls. Right: The energy spectra, separated by isotope, of the material originating background. There are several peak-like features that can potentially obscure bosonic DM signals.

vast majority of material originating events are situated very close to the TPC wall, or the top and bottom PMT arrays. On a closer look, even the cryostat stiffening rings at $z \approx -40$ cm and $z \approx -110$ cm are visible in the event distribution. The energy spectrum is complex and features many peak-like structures, which have the potential to obscure a signal from bosonic DM, and in [72] for example, the material originating events pose the third greatest contribution to the total background. For this reason it is important to reduce this background, and know what remains as well as possible.

Because the material background has such a significant spatial dependence, the simulations used to study it are so-called *full-chain* simulations. Full-chain simulations include the propagation of initial particles, and the subsequent products from interactions in high detail, generated using the GEANT4 software [119]. After the derivation of the particle tracks via GEANT4, the detector physics and signal reconstruction were simulated using *fuse*. Since the material background simulations are necessary for a multitude of different analyses, and require a large amount of computing time, they were produced once and shared among the collaboration. However, in the course of this work, a minor source of mismodeling in the *fuse*-based simulation campaign was discovered, that affects events originating from outside of the TPC. For this reason, the material background simulation in this thesis makes use of a set of older simulations, where the detector response was simulated using the since deprecated *WFSim* software [211]. Given that, additional checks on the simulated material background were performed and are detailed in the following section.

5.3.2 Validation of the Material Background Simulation using Multi-Scatter Events

The SS selection, described in [subsection 5.2.2](#), was also applied to the simulated events, which permits the comparison of the simulation with data using only MS events. The first advantage of only using the MS events is that the SS events, that are to be used in the DM search, are not revealed in this preliminary study. In addition, the material originating rate of MS events is high enough that other backgrounds are negligible. The rate of MS events, split by originating isotope, in addition to the SR0 and SR1 MS data can be seen in [Figure 5.16](#).

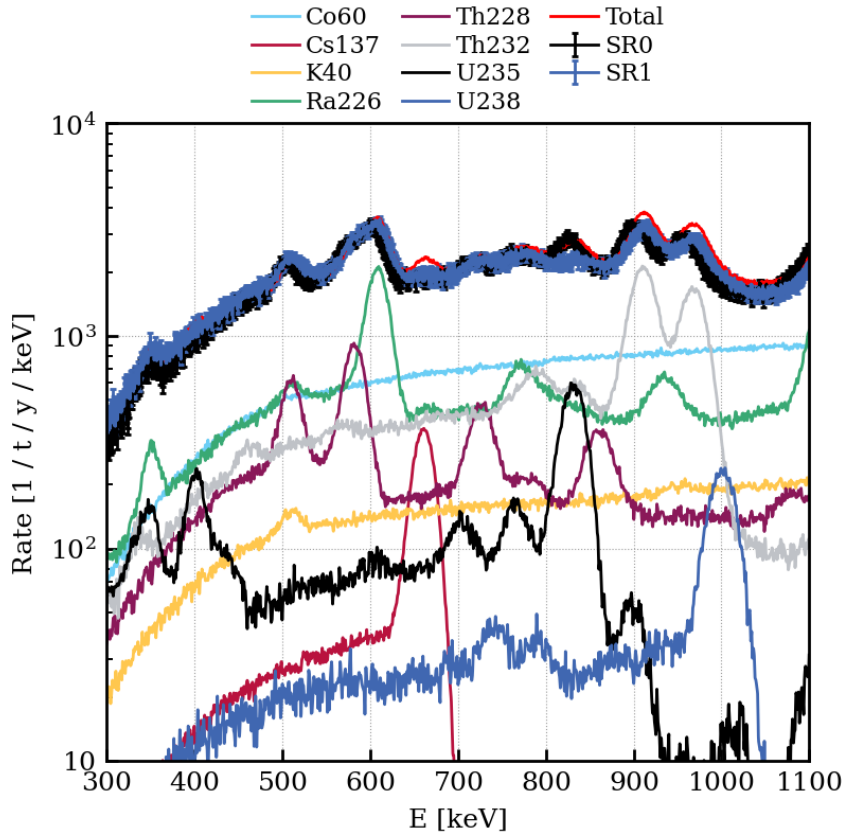


Figure 5.16: The MS event rates of SR0 and SR1, in addition to the MS part of the simulated material background, that is split by mother isotope.

There are three major discrepancies between the simulation and the data: At 662 keV, there is a prominent photo-peak from ^{137}Cs in the simulations, that can not be observed in either SR0 or SR1. In addition, at around 820 keV there is a peak originating in the decay chain of ^{235}U , which is visible in the simulated rates and in SR0, but not in SR1. The difference between the two science run can potentially arise from differences in the signal reconstruction and correction process. Finally, the rate of the two ^{232}Th peaks between 900 keV and 1000 keV is overestimated in the simulations, when compared to SR0 and SR1 data.

To make the material background model better resemble the data, the simulated rate is fitted to the data, by scaling the rates of the ^{137}Cs -, ^{235}U - and ^{232}Th -parts, while the remaining isotope rates are kept fixed to their nominal value. The resulting best-fit

scaling factors of these three isotopes are shown in Figure 5.17, where the dashed line at 1 signifies the nominal value. Scaling the contributions of ^{137}Cs , ^{235}U and ^{232}Th by their respective best fit factor results in the spectra shown in Figure 5.18. The scaling factor were propagated to the simulated SS event rate to derive the expected spectrum of material background events. In addition, half the difference between the nominal and the scaled rate was added as a systematic uncertainty of the rate.

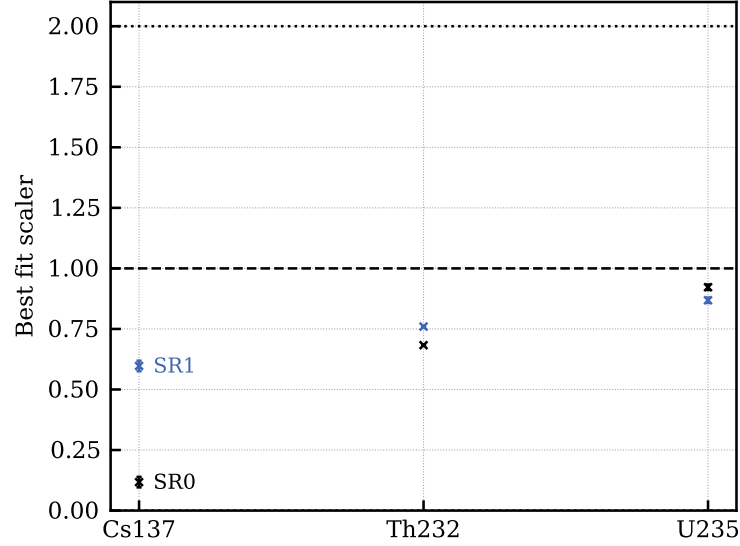


Figure 5.17: The scaling factors that result in the best fit between simulation and data. The black dashed line visualizes the nominal (non-scaled) value.

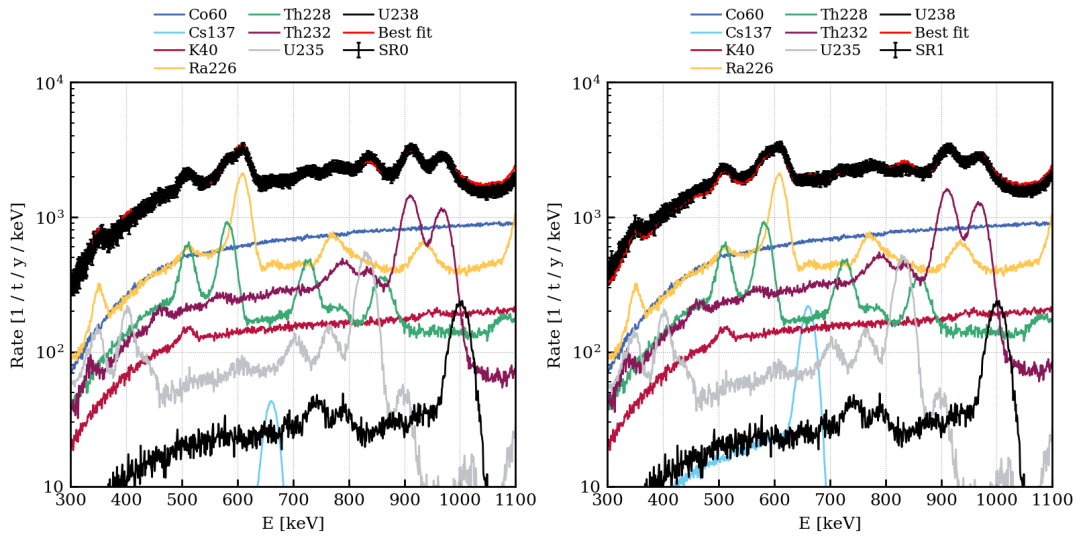


Figure 5.18: The MS event rates of SR0 (left) and SR1 (right), including the material background simulation, where the ^{137}Cs -, ^{235}U - and ^{232}Th -based rates were rescaled by the best-fit scaling factors.

5.3.3 Optimization of the Fiducial Volume

To counteract extrinsic backgrounds originating in the detector materials of the LXe skin, the self-shielding of LXe is utilized via fiducialization. Fiducialization is the

process of selecting a subvolume of the TPC, the Fiducial Volume (FV), as the sensitive detector and rejecting every event outside of this FV from the analysis. Doing so greatly reduces the material background, as is immediately visible from looking at its spatial distribution in [Figure 5.15](#).

Due to limited statistics, the spatial distribution of simulated events of the extrinsic backgrounds becomes more noisy for positions further inside the TPC. Since the event rates from the innermost part of the TPC are at least $\sim 10^6$ lower than those at the TPC edge, an excessive number of simulated primary events would be necessary to get a statistically significant rate in the TPC center. To avoid this, but still have a smooth distribution of the backgrounds, the spatial distribution is modeled via an AKDE [212]. AKDE is the approximation of a distribution with a product of multiple multivariate analytical kernels, such as normal distributions, which are used here. The *adaptive* part refers to the size of the kernels, meaning the width of the normal distributions, being variable across the fit region, as opposed to constant. The comparison of the spatial distribution of the background and the AKDE fit is shown in [Figure 5.19](#).

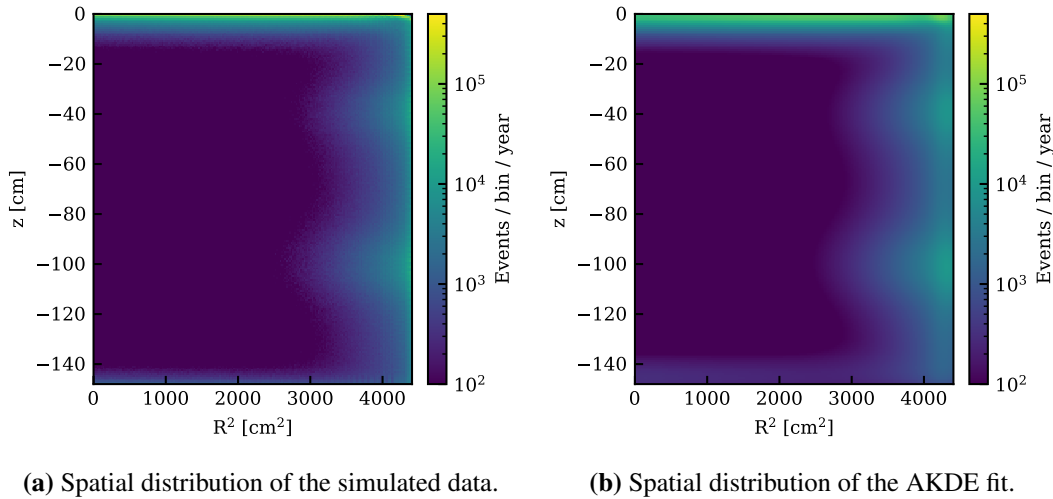


Figure 5.19: Left: An approximation of the total background, using the spatially dependent material and ^{214}Bi skin backgrounds, in addition to the two biggest intrinsic, and therefore spatially uniform, backgrounds from ^{136}Xe and ^{214}Pb . Right: An AKDE fit of the distribution of this background.

The AKDE fit was used to optimize the FV applied for the bosonic DM search. To make the AKDE model more realistic the two largest intrinsic, and spatially flat, backgrounds originating from $^{136}\text{Xe } 2\nu\beta\beta$ and ^{214}Pb were added to the model prior to the fit. Further, the ^{214}Bi LXe skin background, that is described in more detail later in [section 5.6](#), was also included when deriving the AKDE model. Finally, kernel convolutions are only defined if there is underlying data for the full width below the kernel. A kernel with a mean at $z = 0$ cm can not see data for the half of its width that is above $z > 0$ cm, which makes the behavior of the AKDE fit unreliable at the TPC edges. As a countermeasure, the values of the background were extrapolated as constants for 20 cm in each direction outside of the defined parameter range.

The reliability of the AKDE result was determined via bootstrapping. For 50 trials, a

sample of the same size as the background data was drawn at random from the distribution of the background in Figure 5.19a. In each trial an AKDE fit was performed on the sampled background distribution. The normalized mean AKDE fit value distribution is shown in Figure 5.20, as well as the 16 %, 50 % and 84 % percentiles of the mean, in black, and all the 50 bootstrap trials in red. The spread of the bootstrapped median and 84 % lines is minor, which is expected since there is a large extent of data available in those regions. Closer to the TPC center, where the 16 % percentiles lie, the spread is somewhat greater. However, the shape and positions of all the lines is still very much in agreement, showing that the mean AKDE fit can be relied on to model the background distribution well.

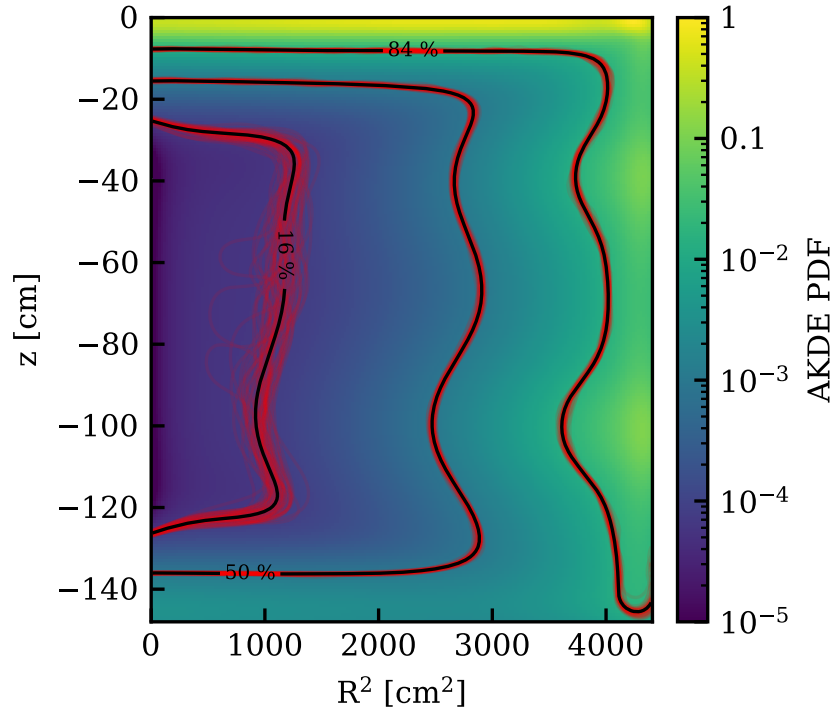


Figure 5.20: The distribution of the mean AKDE fit, and its 16 %, 50 % and 84 % lines in black. In red are the same percentiles of each of the 50 bootstrap trials.

To optimize the FV there is the need for a figure of merit that incorporates the decrease of background events, but also the decrease of usable detector mass that comes hand in hand with the shrinkage of the FV. Here, the median asymptotic significance from [122] was used

$$Z = \sqrt{2 \cdot \left((s + b) \ln \left(1 + \frac{s}{b} \right) - s \right)}, \quad (5.12)$$

where s are expected signal counts, and b are expected background counts. The optimum FV was determined using the following procedure:

1. The expected background distribution was predicted by the AKDE model for uniform binning in (R^2, z) .
2. The bins were sorted by decreasing value of s/b , where s was assigned to equal one.

3. The figure of merit Z was computed via [Equation 5.12](#), for all bins in a cumulative way, meaning that the value Z_i of bin i is computed using all previous bins that have by merit of the prior sorting a higher s/b than bin i .
4. The function of Z in dependence of the included bins i had a maximum $i_{\max}$, that signified all bins up to bin $i_{\max}$ should be included in the FV.

Using the above procedure, the resulting optimized FV is shown in [Figure 5.21](#).

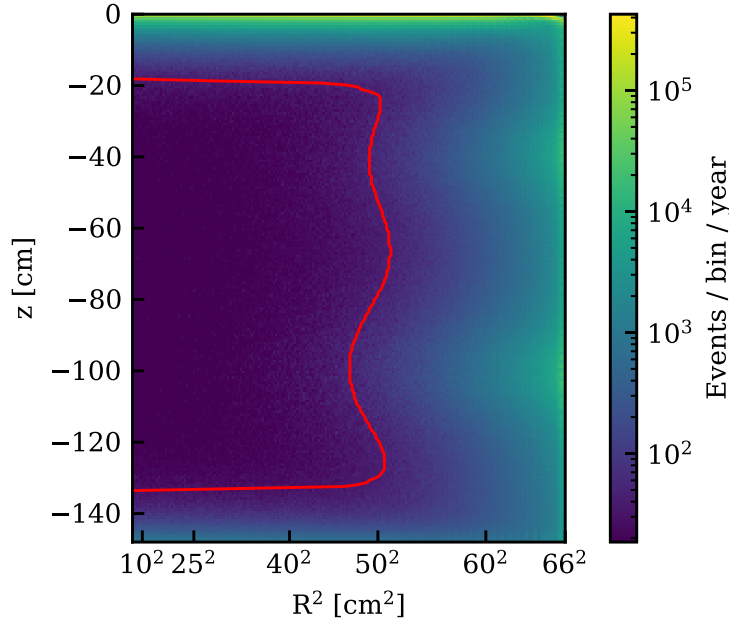


Figure 5.21: The resulting FV using the AKDE background model that optimizes [Equation 5.12](#). In addition, the background simulation used to fit the AKDE is shown.

5.3.4 Uncertainty of the Fiducial Volume

There are three sources of uncertainty in the derivation of the detector target mass of the FV. Firstly, the LXe density has a measurement uncertainty, second an uncertainty of the electrical field, and finally, there is a position correction uncertainty on the event positions. The LXe density was measured to be $\rho_{\text{LXe}} = (2861.5 \pm 3.1) \text{ kg} \cdot \text{m}^{-3}$ in SR0 and $\rho_{\text{LXe}} = (2859.5 \pm 3.3) \text{ kg} \cdot \text{m}^{-3}$ in SR1. From the modeling of the electrical field, there is a 3 % relative uncertainty on any FV, based on the differing size of CIV inside the detector, depending on potential charge-up effects of the PTFE walls [\[154\]](#).

The position reconstruction uncertainty arises due to the difference between the reconstructed S2 position at the liquid interface, and the actual position of the energy deposit inside the TPC. In the XENONnT analysis there exist two different sets of coordinates: the naive coordinates and the corrected coordinates. The naive coordinates are the output of the position reconstruction algorithm in (x, y) and the drift time multiplied with the electron drift velocity in the LXe. However, there are small inhomogeneities in the electric drift field, that bend the travel path of the electrons along the field lines. This leads to two effects: First, the electrons are not traveling in a straight line and therefore the z position is incorrect if modeled linearly. Second,

the horizontal position at the liquid interface, that is the output of the position reconstruction models described in [chapter 3](#), is different from the horizontal position of the actual energy deposition. This necessitates two corrections, one of the vertical and one of the horizontal position. The vertical (z) correction is based on simulations of the electrical field, where z is inferred from the drift time integrated along a curved field line as opposed to a straight line. In SR0, this z correction was not yet implemented. The horizontal correction is the so-called field distortion correction (FDC). The FDC is developed under the assumption that ^{83m}Kr calibration data is distributed uniform in the TPC. The naive positions of ^{83m}Kr events are stretched until their radial quantiles match a uniform distribution. This quantile matching is done in multiple bins of z and polar angle ϕ . Further, the electron survival probability, derived from electric field simulations, decreases when nearing the lower and outer edges of the TPC, creating a CIV at the far outside and bottom. This electron survival probability is convolved with the uniform target distribution for the development of the FDC in SR1. This whole procedure and its effect on the naive and corrected position is visualized in [Figure 5.22](#), with the use of simulated data.

To derive the uncertainty in the FV that originates from the position reconstruction and correction process, 1×10^6 events were simulated uniform in the TPC for each science run. The red and blue coloration is determined by events inside or outside the FV respectively, which is determined in corrected coordinates. The left hand column of [Figure 5.22](#) shows the naive positions of the simulated events, while the center column shows the corrected positions. It is well visible how there are no events at high R^2 and low z values in naive coordinates, since those events get drifted further to the inside of the TPC, which is corrected by the FDC, thus making those events appear correctly in the central column. In addition comparing the left and center columns between the top (SR0) and bottom (SR1) rows in [Figure 5.22](#) shows that the vertical z correction was not yet applied in SR0, as well as the CIV, which is clearly visible in the lower right corner of the corrected positions of SR1.

Finally, by nature of being simulated, the events also have a known true position that is shown in the right most column. In SR0 in particular, it is visible that the FV include events that have originated at much lower z values than intended from purely reconstructed positions. And since this same behavior is not observed in SR1, it can be assumed to be caused by the missing vertical correction. From the distribution of the true positions of those events included in the FV, the increase or decrease of the FV f_{FDC} when going from corrected to true positions is computed. The conversion values are $(94.40 \pm 0.22) \%$ for SR0 and $(99.50 \pm 0.23) \%$ for SR1. The fiducial volume is corrected for this factor and the uncertainty is propagated to the final fiducial mass.

The naive fiducial volume V_0 is determined numerically via MC, and the corrected mass is then computed via:

$$m_{\text{FV}} = V_0 \cdot f_{\text{FDC}} \cdot \rho_{\text{LXe}}. \quad (5.13)$$

The uncertainties are of each of these values, in addition to the 3 % uncertainty from the electrical field modeling, are summed in quadrature. The resulting fiducial masses are $m_{\text{FV}} = (2.35 \pm 0.07) \text{ t}$ in SR0 and $m_{\text{FV}} = (2.47 \pm 0.07) \text{ t}$ in SR1.

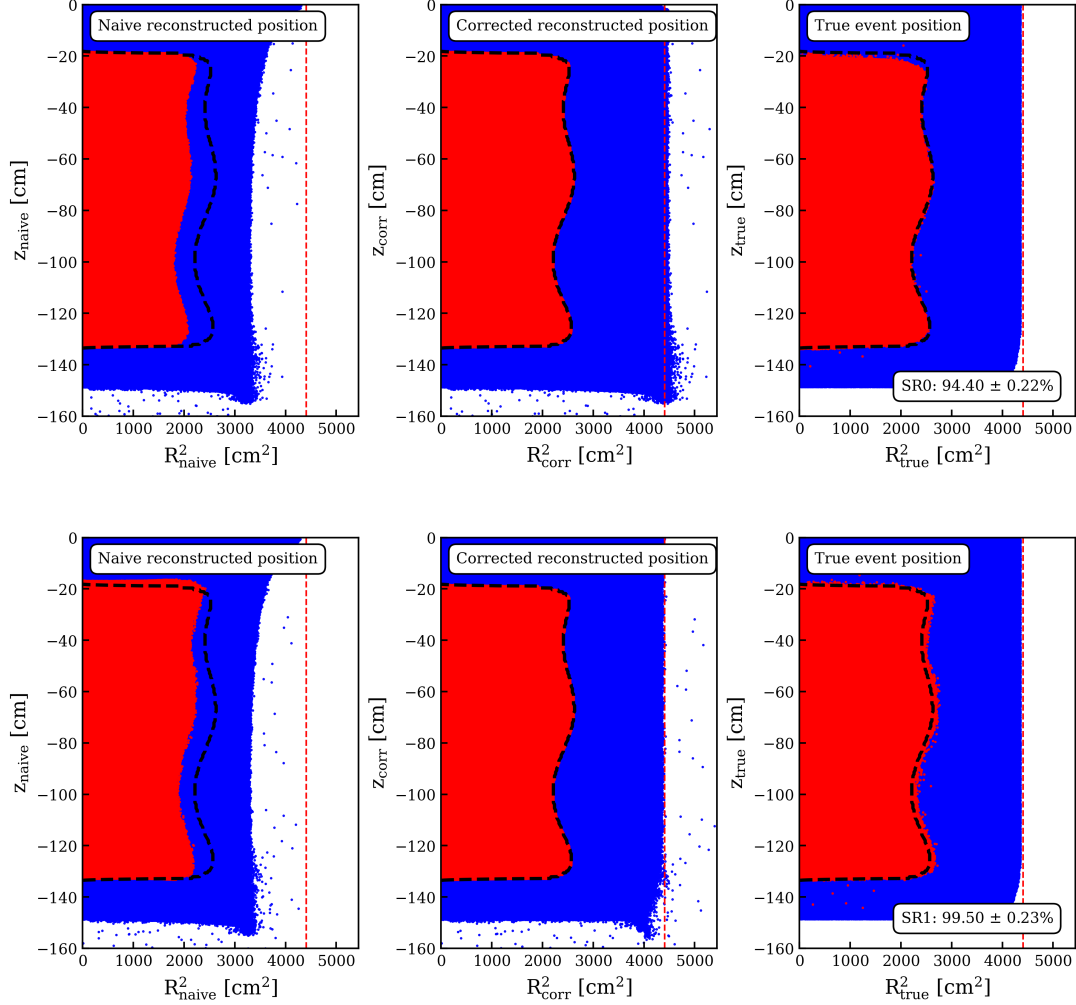


Figure 5.22: The size of the FV in naive reconstructed coordinates (left), corrected reconstructed coordinates (center) and true simulated coordinates (right). The volume in true simulated coordinates is $(94.40 \pm 0.22)\%$ as big as the one in corrected coordinates in SR0 (top) and $(99.50 \pm 0.23)\%$ in SR1 (bottom). The vertical z correction and the CIV correction were not yet implemented in SR0. Simulated events outside (inside) of the FV are drawn in blue (red). The outline of the FV, using the respective coordinates, is shown as a black dashed line. The red dashed line highlights the TPC wall. The FV cut is performed in corrected reconstructed coordinates. However, the distribution of the events inside the FV is slightly different when looked at in true simulated coordinates, which has to be accounted for when deriving the fiducial mass.

5.3.5 Remaining Material Background after Fiducialization

The spectral rate of the material originating SS events inside the optimized FV is depicted in [Figure 5.23](#), with the nominal rate from the simulations shown in black, and the scaled rates, using the best-fit factors for the ^{137}Cs -, ^{235}U - and ^{232}Th -based rates shown in dark blue (SR0) and light blue (SR1). The uncertainty on the rate is a combination of the systematic and statistical uncertainties, originating from the simulation, and half the difference between the nominal and scaled rates, all added in quadrature. The resulting event rates inside of the FV are $A_{\text{mat}} = (58\,133.24 \pm 8607.30) \text{ t}^{-1}\text{y}^{-1}$ in SR0 and $A_{\text{mat}} = (56\,312.44 \pm 7611.70) \text{ t}^{-1}\text{y}^{-1}$ in SR1, for the energy range between 1 keV and 1200 keV.

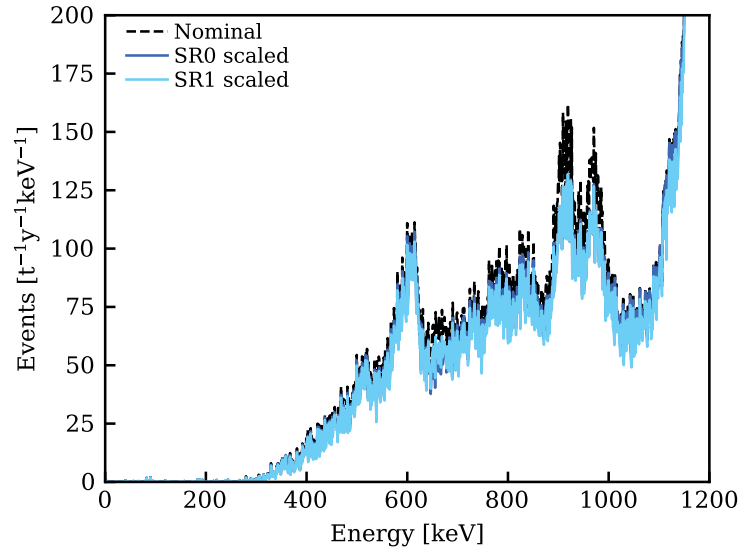


Figure 5.23: The spectra of the material originating background inside of the optimized FV. In black is the background rate, derived by using the nominal values of the simulation for each isotope. In blue are the rates that rescale the ^{137}Cs -, ^{235}U - and ^{232}Th -based rates by the best-fit factors, derived in [subsection 5.3.2](#).

5.4 The Decays of ^{214}Pb and ^{212}Pb

5.4.1 Lead Backgrounds: Origin and Mitigation

^{214}Pb and ^{212}Pb are parts of the decay chains of ^{238}U and ^{232}Th respectively, which are primordial nuclides and therefore are present in varying concentrations in all detector materials. Both decay chains involve Radon isotopes, ^{222}Rn and ^{220}Rn , which is a noble gas and can leave the detector materials and mix into the LXe itself, a process called outgassing. Because of that, radioactive isotopes below Radon in the decay chain, in particular ^{214}Pb and ^{212}Pb , can be found anywhere inside the detector, as opposed to the other material originating backgrounds (discussed in [section 5.3](#)). As described in [section 5.3](#), all detector materials underwent a radiopurity study, that determined their contamination with radioactive isotopes, and in addition, ^{222}Rn emanation measurements were performed to assess the outgassing of the gaseous radionuclides [[101](#), [213](#)].

In addition to material origin, ^{214}Pb and ^{212}Pb is also introduced into the detector by dedicated ^{222}Rn and ^{220}Rn calibration campaigns. The ^{222}Rn calibration in particular is made possible by the improved Radon removal system (RRS) of XENONnT [97], as without it, injecting ^{222}Rn into the TPC would pose too much of an additional background. The rate of ^{222}Rn induced events for XENONnT with and without the RRS, and for a selection of other LXe experiments, is shown in Figure 5.24. For example for the measurement of the solar neutrino rate in [3], it is crucial to minimize the activity of ^{222}Rn in the TPC.

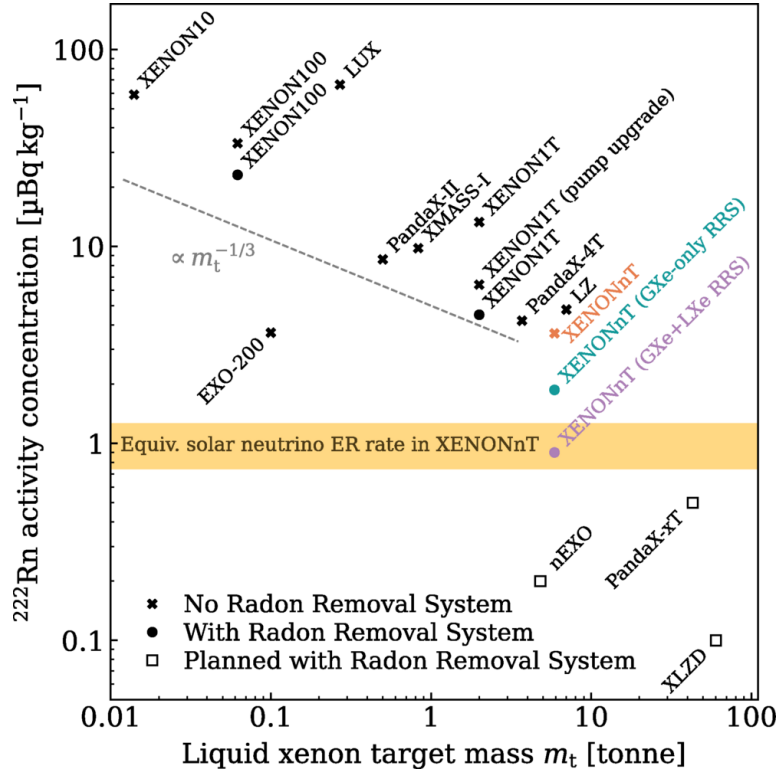


Figure 5.24: The ^{222}Rn activity in various LXe TPCs, from [97]. The ^{222}Rn activity in XENONnT using the RRS drops by a significant amount, compared to the activity in XENONnT without the RRS, and compared to other experiments of similar size.

5.4.2 Total ^{214}Pb and ^{212}Pb Activities

The concentration of ^{214}Pb inside of the TPC is constrained via the monoenergetic α events of ^{222}Rn and ^{214}Po . Since the α events are high energy NR interactions, they can be identified in an S1-only analysis, which allows for a pure selection of the individual decay events, and provides a good estimate on the ^{222}Rn event rate. The ratio of ^{222}Rn to ^{214}Pb is determined from ^{222}Rn calibration data, where the two rates are orders of magnitude higher than events from other sources.

For ^{212}Pb , this method is not applicable, as the involved α emitters have too short a lifetime to constrain the ^{212}Pb to ^{220}Rn ratio. Instead, the concentration of ^{212}Pb is constrained via the fit of a differential equation of the ^{212}Pb - ^{212}Bi - ^{212}Po decay chain, including a constant outgassing rate and effective removal rate.

The total activities of ^{214}Pb and ^{212}Pb are determined to be $A_{\text{Pb}214} = (43738.8 \pm$

$2145.9) \text{ t}^{-1} \text{ y}^{-1}$ and $A_{\text{Pb}212} = (1199.2 \pm 220.9) \text{ t}^{-1} \text{ y}^{-1}$ in SR0 and $A_{\text{Pb}214} = (23\,983.8 \pm 1073.0) \text{ t}^{-1} \text{ y}^{-1}$ and $A_{\text{Pb}212} = (1073.0 \pm 135.7) \text{ t}^{-1} \text{ y}^{-1}$ in SR1. However, these rates are not immediately applicable to the study presented here, because of the β - γ event topology of the decays, as explained in the following section.

5.4.3 The Challenge in Constraining Pb β - γ Events

The ^{214}Pb (^{212}Pb) β -decay can result in an excited state of ^{214}Bi (^{212}Bi), resulting in a closely following γ emission. Schematics of the decays, including energies and branching ratios, are shown in Figure 5.25, taken from [214].

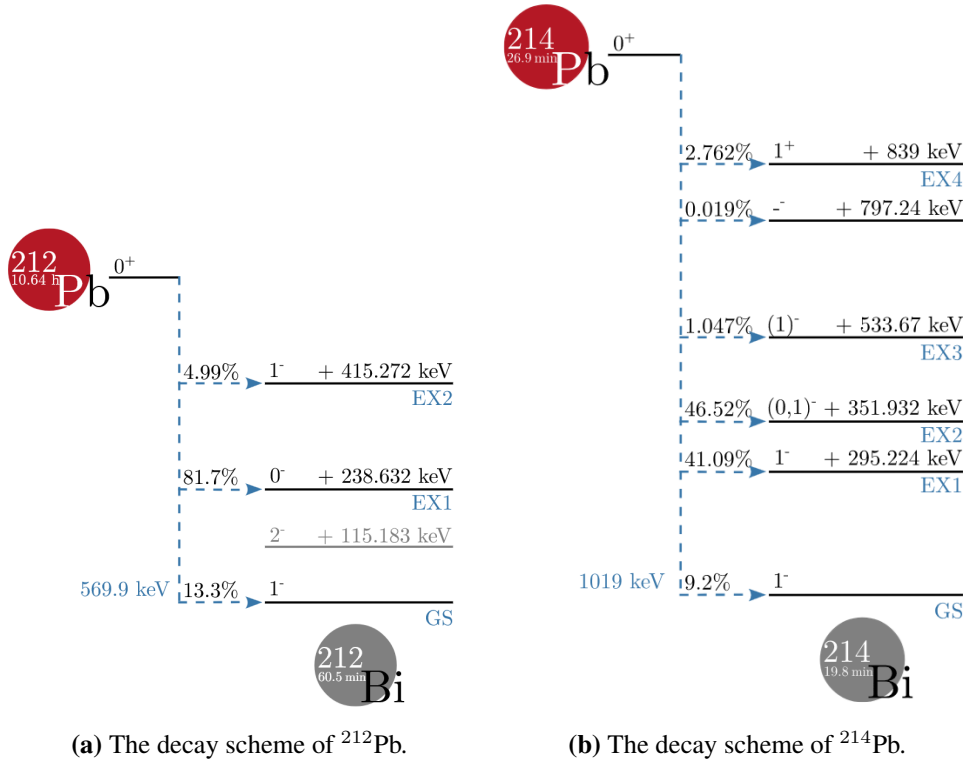


Figure 5.25: The decay schemes of the two lead isotopes, taken from [214], with the ground state labeled as GS and the excited states labeled as EX.

For low energy analyses such as the one described in [3], only the ground state branches are of concern. If the β and γ decays can be resolved as separate, the event will be removed via the SS selection, and if they are merged into a single event, the energy will reside outside of the ROI. For the search for bosonic DM presented here, the major goal was the extension of the energy ROI up to 1 MeV, where all the branches of both decay chains become relevant. The β - γ cascade events can be seen as true MS events, however, since they are happening so close together they have a high chance to be reconstructed as a SS event. This particular event topology makes the ^{214}Pb and ^{212}Pb backgrounds special, as the relevant background is not the total rate of events, but rather the fraction of events that are *reconstructed* as SS. For this reason a so-called *full chain simulation* was performed for the ^{214}Pb and ^{212}Pb backgrounds, that included realistic β and γ propagation in the TPC simulated with GEANT4, followed by signal generation and reconstruction simulated with *fuse*. Provided that the simulation models reality well, it can be used to infer the spectral shape and the number of

Pb events reconstructed as SS.

5.4.4 Comparison of ^{212}Pb Full Chain Simulation and ^{220}Rn Data

The simulation of ^{212}Pb , and the corresponding ^{220}Rn calibration data, is used to verify the full chain simulation, since for ^{214}Pb simulations a ^{222}Rn calibration is required, which only exists for SR1, not for SR0.

To get a pure selection of ^{212}Pb events in calibration data, and not be affected by material background events, a cylindrical FV cut of $R < 30$ cm and -130 cm $< z \leq -20$ cm was applied. In addition, the ER band selection was applied to both datasets, while the data quality cuts described in [section 5.2](#) were applied to the calibration data. The SS selection was explicitly not applied to any data at first. For the full chain simulations, 10^8 primary ^{212}Pb decays were simulated, using the branching ratios derived in [\[214\]](#). After reconstruction and selection, there were 4.5×10^7 reconstructed simulation events in SR0 and 4.8×10^7 reconstructed simulation events in SR1 left, that were compared to 1.5×10^5 and 3.6×10^5 calibration events respectively.

[Figure 5.26](#) shows the energy spectra of the selected events, for simulation (blue) and calibration data (black) for SR0 ([Figure 5.26a](#)) and SR1 ([Figure 5.26b](#)), as well as the relative deviations between simulation and calibration. The SR0 calibration data were taken shortly after an AmBe calibration, leading to a contamination of neutron activated ^{131m}Xe and ^{129m}Xe isotopes, that emit monoenergetic gamma-rays at 163.9 keV and 236.2 keV. At the high energy tails, there is an excess in calibration data, which is expected since the ^{212}Pb spectrum ends at 569.9 keV and the calibration data is not a 100 % pure selection of ^{212}Pb events. Given these constraints, the simulated spectrum matches the calibration data well.

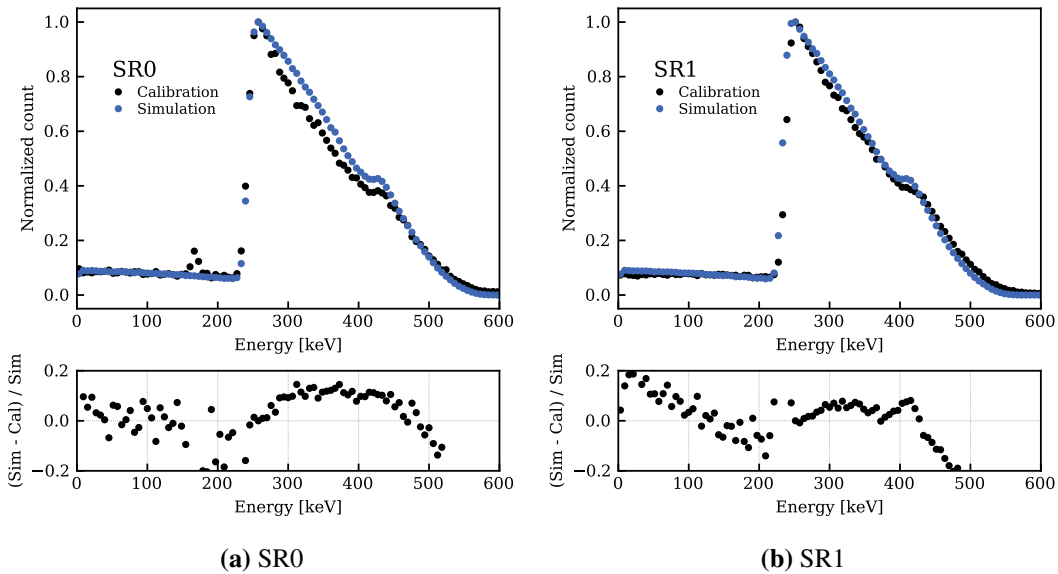
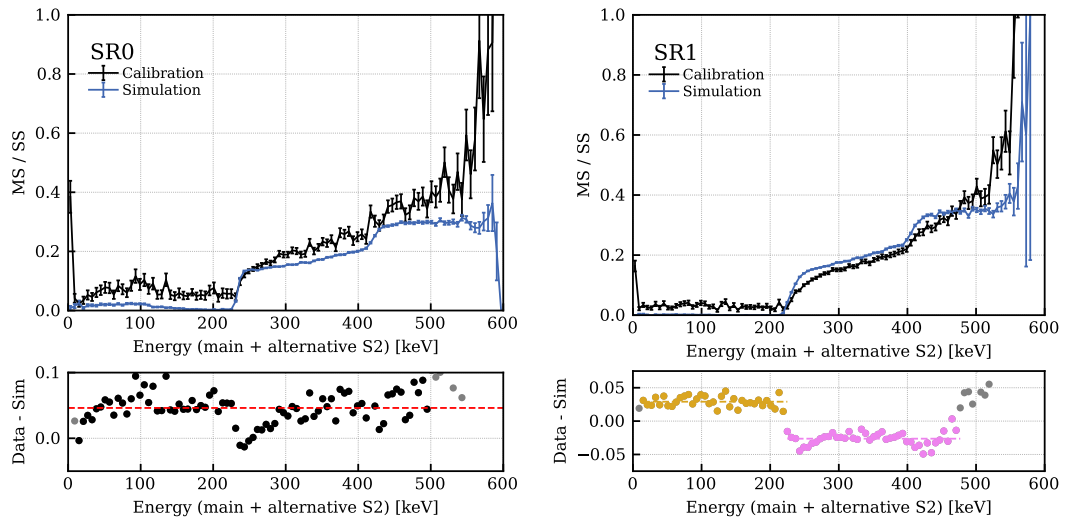


Figure 5.26: The normalized spectra of ^{212}Pb events from simulated (black) decays and from calibration data (black). The bottom panels show the relative deviations between simulation and calibration. In SR0, the ^{220}Rn calibration happened close to an AmBe calibration, which is why there is a contamination of ^{131m}Xe and ^{129m}Xe monoenergetic gamma events.

To derive the lead backgrounds from simulations, it is necessary to verify that the SS selection works the same on simulated data and calibration data. The ratio of the number of events reconstructed as MS or SS is shown in Figure 5.27. The energy of the individual events was not reconstructed with Equation 2.22, but rather with an adjusted version that takes into account the area of the alternative S2 signal

$$E = W \cdot \left(\frac{cS1}{g_1} + \frac{(cS2_{\text{main}} + cS2_{\text{alt}})}{g_2} \right). \quad (5.14)$$

The reasoning behind this is that the number of interest is the MS probability for an event of a given *total* energy from the β and γ decays. Using only the main S2 area would discard the information about the energy of the secondary energy deposit in the case that it is reconstructed as alternative S2.



(a) SR0: There is a bias in the MS rate of the calibration data of 4.6 % on average (red).

(b) SR1: At low energies, the MS ratio is higher in the calibration data by approximately 2.9 % (gold). At higher energies, the ratio is greater in simulated data by 2.6 % (violet).

Figure 5.27: The MS to SS ratio of ^{212}Pb events, compared for calibration and simulation data. The bottom panel shows the deviation between the MS to SS ratios between data and simulation. In SR0 there is a constant bias across all energies. In SR1 there seem to be two distinct energy regions where the simulation is underestimating, and overestimating the MS rate.

From Figure 5.27 it can be seen that the distinction between SS and MS events is similar between simulated and measured data, and the deviation between both is in general below 10 %. In SR0, the MS fraction in the calibration data is higher than that in simulations by an approximately constant rate of 4.6 %. The constant bias is shown as a red line in Figure 5.27a, and gray points were excluded when deriving it. This can be explained by data contamination, meaning that the selection of calibration events was not a pure ^{212}Pb selection, but contained additional background components, including MS backgrounds. For SR1 however, this behavior is only seen for energies below ~ 225 keV. Above that energy, the MS rate in the simulated data is actually greater than that in the calibration data. This is visualized by the golden and violet

Table 5.4: The simulated SS fractions of ^{214}Pb and ^{212}Pb .

	$f_{\text{Pb}214} [\%]$	$f_{\text{Pb}212} [\%]$
SR0	65.5 ± 5.7	85.2 ± 22.6
SR1	62.9 ± 3.1	84.0 ± 12.8

Table 5.5: The reconstructed SS activities of ^{214}Pb and ^{212}Pb .

	Expected ^{214}Pb activity [$\text{t}^{-1}\text{y}^{-1}$]	Expected ^{212}Pb activity [$\text{t}^{-1}\text{y}^{-1}$]
SR0	28648.9 ± 2862.0	1021.7 ± 330.0
SR1	15085.8 ± 1004.1	901.3 ± 178.5

data points and dashed lines in [Figure 5.27b](#), below 225 keV there is a bias of 2.9 %, and above of -2.6 %. These deviations, in the case of SR1 the greater of the two, will be propagated as uncertainties to the calculation of the background rate. The fraction of SS events $f = N_{\text{SS}}/(N_{\text{SS}} + N_{\text{MS}})$ in the simulated spectra for ^{214}Pb and ^{212}Pb is listed in [Table 5.4](#).

Aside from the spectrum and activity used in the background model, the full chain simulation and the corresponding calibration data can be used to validate which parameter spaces are well modeled in simulations, and which are not. This is of interest for example in the cut acceptance study in [section 5.2](#), to determine whether or not a cut can be developed and the acceptance derived on simulated data. The comparison of parameter spaces can be found in [Appendix D](#).

5.4.5 Resulting ^{212}Pb and ^{214}Pb Spectra

The activity of lead background events reconstructed as SS can be computed from the total activity of the lead decays multiplied with the fraction of SS events recovered from the simulation, listed in [Table 5.4](#). The resulting SS event rate of lead background events is listed in [Table 5.5](#).

The values for ^{214}Pb were also derived via full chain simulation, however, as an uncertainty on the SS fraction f , the value derived from the comparison between simulated ^{212}Pb values and the calibration data were used, based on a lack of calibration data for ^{214}Pb . [Figure 5.28](#) shows the resulting energy spectra of the lead decays after applying the FV, ER band and SS selections.

These spectra will be used for the lead backgrounds in the fit for bosonic DM interaction search. However, since they were derived via full chain simulations, the energy bias and smearing are already applied, which needs to be kept in mind for the inference in [section 5.7](#).

5.5 Neutron Activated Isotopes

Some xenon isotopes can interact with neutrons via neutron capture or inelastic scattering, forming metastable or unstable short lived isotopes, that make up an additional background for the DM search:

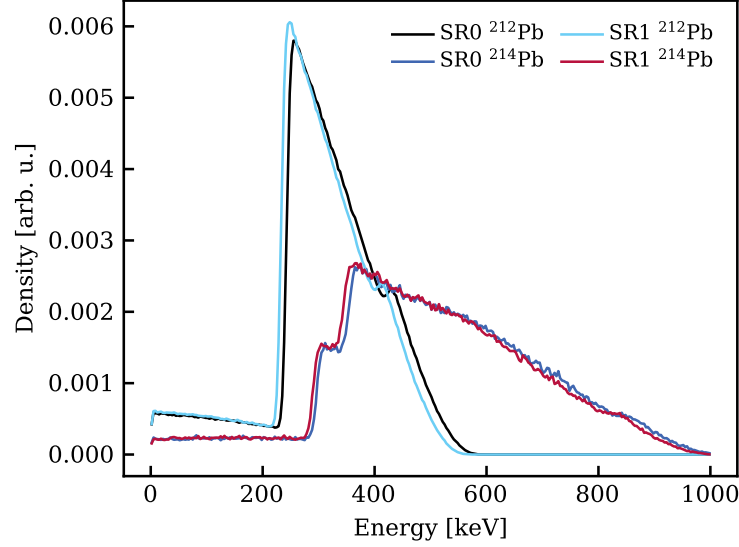


Figure 5.28: The simulated SS spectra of the ^{214}Pb and ^{212}Pb decays for SR0 and SR1.

- ^{133}Xe [215]: ^{133}Xe decays via β -decay, with a half-life of $\tau_{1/2} = 5.3$ d and an endpoint energy of 346.4 keV to an excited state of ^{133}Cs , which in turn emits a γ of 81.0 keV. The γ emission follows the β decay practically immediately, causing the two energy deposits to be reconstructed as a single signal, resulting in an effective spectrum of ^{133}Xe that starts at 81.0 keV and ends at 427.4 keV.
- ^{129m}Xe and ^{131m}Xe [216, 217]: The metastable state of ^{129m}Xe (^{131m}Xe) decays after a half-life of $\tau_{1/2} = 8.9$ d ($\tau_{1/2} = 11.8$ d) by emitting a monoenergetic γ with an energy of 236.2 keV (163.9 keV).
- ^{127}Xe [218]: ^{127}Xe decays, with a half-life of $\tau_{1/2} = 36.4$ d, into one of two possible excited states of ^{127}I , releasing an energy of 375.0 keV with a branching ratio of 47 % or 202.9 keV with a branching ratio of 53 %. The excited states of ^{127}I can also decay via two branches, releasing photons with energy 33.2 keV with 85 % probability or 5.2 keV with 12 % probability. Since the two events of the decay chain happen so close in time, they are also expected to be merged in all cases, which leads to effectively four different observable monoenergetic signals at 208.1 keV (6.4 %), 236.1 keV (45.1 %), 380.2 keV (5.6 %) and 408.2 keV (39.5 %).
- ^{125}Xe [219]: With a half-life of $\tau_{1/2} = 59.4$ d, ^{125}Xe decays to an excited state of ^{125}I , releasing an energy of 188.4 keV, 243.4 keV or 453.9 keV. Subsequently, the excited ^{125}I emits a photon of 33.2 keV, the signal of which is again merged with that of the initial ^{125}Xe decay, resulting in three monoenergetic signals of energies 221.6 keV (25.4 %), 276.6 keV (66.6 %) and 487.1 keV (4.7 %).
- ^{125}I [219]: ^{125}I is produced by the decay of ^{125}Xe , and can itself decay again via electron capture (EC) with a half life of $\tau_{1/2} = 59.4$ d. The produced excited states of ^{125}Te will decay, leading to signals of 36.5 keV (3.5 %), 40.4 keV (15.6 %) or 67.3 keV (80.1 %). However, ^{125}I is a special background for two

Table 5.6: The decay energies and half lives of neutron-activated isotopes.

Isotope	$\tau_{1/2}$ [d]	E [keV]	Branching ratio [%]
^{125}Xe [219]	0.7	[221.6, 276.6, 487.1]	[25.4, 66.6, 4.7]
^{127}Xe [218]	36.4	[208.1, 236.1, 380.2, 408.2]	[6.4, 45.1, 5.6, 39.9]
^{129m}Xe [216]	8.9	236.2	-
^{131m}Xe [217]	11.8	163.9	-
^{133}Xe [215]	5.3	81.0 – 427.4	-
^{125}I [219]	59.4	[36.5, 40.4, 67.3]	[3.5, 15.6, 80.1]

additional reasons. First, since it is not an isotope of xenon, it will be removed by the purification system, which leads to an effective observable half life that is lower than the literature value. Furthermore, since it is a daughter of the ^{125}Xe decay, it is continuously produced as long as ^{125}Xe is present in the detector.

The literature values for the half-lives, energies and where applicable the branching ratios (BRs) of these so called *neutron activated* isotopes are listed in Table 5.6.

In XENONnT, these isotopes are produced during AmBe calibrations, of which there was one during SR0 and two during SR1, and their concentration in the detector can be constrained via the time dependence of the observed event rate in days following such a calibration. This time dependence of the events is visualized in Figure 5.29, where the rate of events in energy ranges corresponding to specific isotopes is shown as a function of time for data taken during 40 days following an AmBe calibration. One challenge in the derivation of the rate is that individual runs can have no counts at all in the specified energy range. For this reason, multiple runs were grouped together into time windows, that cover a cumulative lifetime of 12 h and 24 h for SR0 and SR1 respectively. The cumulative lifetime refers to the time in which data was recorded, as opposed to physical time passed. For example, if a time interval of 24 h contains only 5 runs of 3 h each, the cumulative lifetime would be 15 h, even though a full day passed between the start of the first run and the end of the last run. This procedure ensured that each datum contains the same exposure of recorded events, but leads to unequal distances in time. In SR0, there was less data available, due to the AmBe calibration being followed immediately by a ^{83m}Kr calibration, limiting the number of runs suitable for the analysis of the neutron activated isotopes. This resulted in the time wise spread of the data being particularly visible in SR0. To take the non uniform time distribution into account, each window was assigned an uncertainty in time, that is computed from the spread of the time of the runs included in the window, weighted by the counts of each run.

Below, the strategy to constrain the rate of events originating from neutron activated isotopes is described, first for ^{131m}Xe in SR1, which serves as an example for all of the Xe isotopes, and then separately for ^{125}I , since ^{125}I requires slightly different modeling. The corresponding figures of the remaining isotopes, in particular those for the SR0 analysis, can be found in Appendix E.

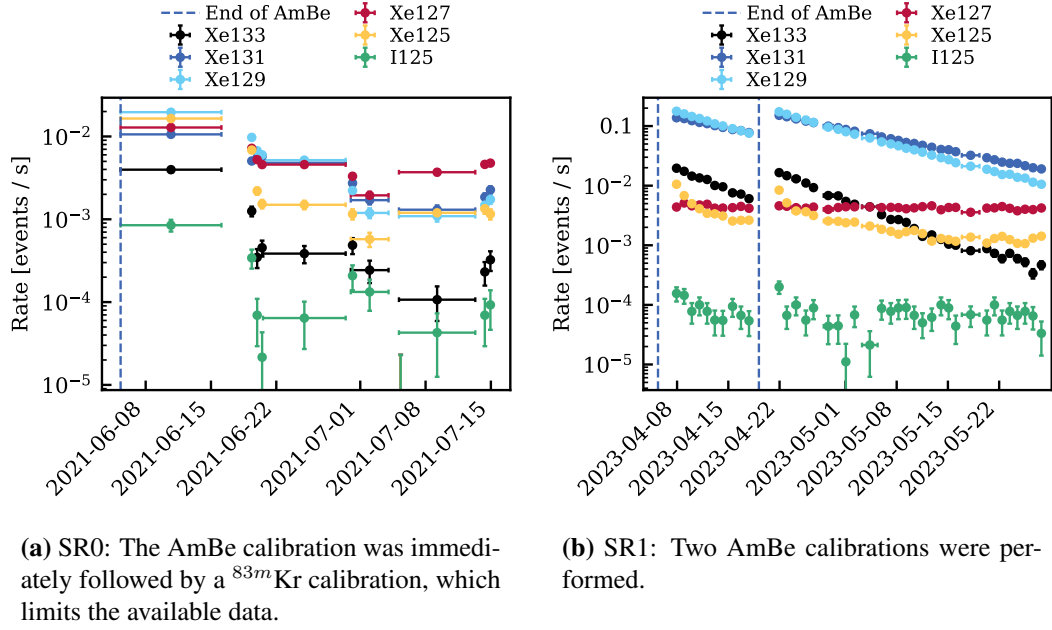


Figure 5.29: The rate of events in 40 days following the AmBe calibrations (blue dashed) of the various neutron activated isotopes. Individual runs have been grouped into time windows of cumulative lifetime of 12 h and 24 h for SR0 and SR1 respectively, and the error bars on the time axis indicate the spread of the runs per time window, weighted by the count of each. The uncertainty of the rate is the Poissonian uncertainty of the event count of each isotope.

5.5.1 The Rate Constraint of Neutron-activated Xenon Isotopes on the Example of ^{131m}Xe

The first step to derive the event rate originating from the ^{131m}Xe decay is to count the events that are within 2σ of the literature value of the decay energy, where σ is defined via the energy resolution of the experiment, described in subsection 5.1.1². The event selection is visualized in Figure 5.30, where the rate of events in the vicinity of the ^{131m}Xe decay energy (red solid) is shown. The red dashed lines signify the 2σ range to each side of the peak, and the rate of ^{131m}Xe events is computed by dividing the count of events within the two red dashed lines by the sum of the lifetime of all runs included in the count.

The rate is computed for the 24 h (12 h in SR0) time windows described above, and fitted with the decay function

$$R_{\text{Xe133}}(t) = \sum_i [\Theta(t - t_i) \cdot N_i \cdot \exp(-\lambda_{\text{Xe133}} t)] + c, \quad (5.15)$$

where the index i indicates the different AmBe campaigns, t_i is the end time of the respective campaign, Θ is the Heavyside θ function, c is a constant background rate, N_i is the number of activated atoms of ^{131m}Xe and $\lambda_{\text{Xe133}} = \ln 2 / \tau_{1/2}$ is the decay constant. The rate as a function of time, including the best fit, is shown in Figure 5.31.

The free parameters in the fit are the initial activities after the calibration, N_i , the

²For the energy ranges that are used in the case of multiple possible peaks, or the continuous β spectrum of ^{133}Xe , refer to Appendix E.

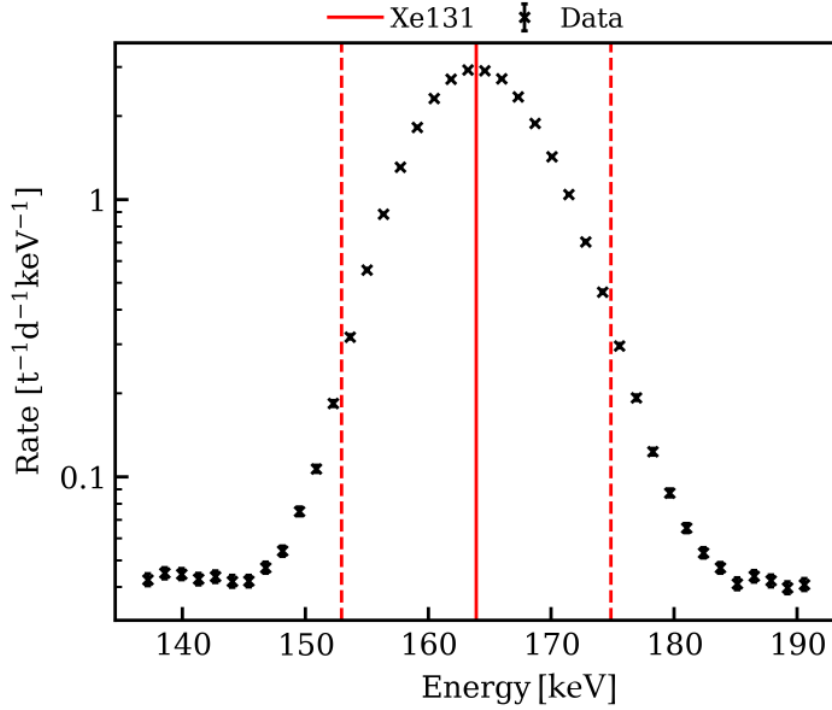


Figure 5.30: The energy spectrum of the SR1 ^{131m}Xe data. The red solid line indicates the literature value of the decay energy, and the red dashed lines show the 2σ interval, where σ is defined via the experiment's energy resolution.

decay constant, λ , and the constant background rate c . The fit is performed using the *scipy* [151] package, which provides best fit parameters as well as a covariance matrix. To properly take into account the correlations among the fit parameters, the uncertainties on the best fit are computed via bootstrapping. The bootstrapping works by sampling parameters from a multivariate normal distribution, whose mean is comprised of the best fit values and whose covariance is the covariance matrix calculated in the fit procedure. The best fit values and their uncertainties are detailed in [Table 5.7](#). The best fit half-lives do not match the literature values within their respective uncertainties, with only the values of SR1 ^{129m}Xe , ^{131m}Xe and ^{133}Xe coming close. One reason for the discrepancy is that the energy ranges used to compute the rate overlap with other time dependent backgrounds, in particular ^{133}Xe spans the energy regions of many other isotopes. Another reason for the mismatch is the low amount of available data, which is exemplified by the SR0 results being further apart from the literature than the SR1 values. This is why the expected number of events for each isotope has to be determined carefully and with conservative uncertainties, which are explained below, following the description of the constraints on ^{125}I .

5.5.2 The Rate Constraint of ^{125}I Decays

The case of ^{125}I is different from the xenon isotopes, since it is produced even after the end of the AmBe calibrations via the decay of ^{125}Xe , and because it is removed from the detector continuously via purification. The second point results in the rate of ^{125}I events decreasing via an effective decay constant $\lambda_{\text{eff}} = \lambda_{\text{I125}} + \lambda_{\text{pur}}$, with the decay constant λ_{I125} and the purification rate λ_{pur} . In addition, because ^{125}I continues to be

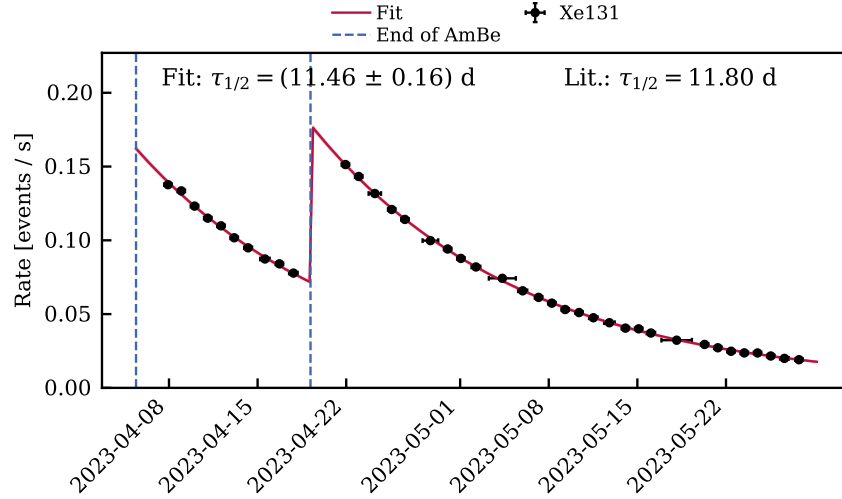


Figure 5.31: A fit of the ^{131m}Xe decay. The blue dashed lines show the end points of AmBe calibration campaigns. Right after the AmBe calibration, there is an increased concentration of the n-activated isotope ^{131m}Xe , which decays over the course of the following weeks. The uncertainty in time is the weighted standard deviation of the run times included per data point. The uncertainty in the rate is propagated from the Poissonian uncertainty on the count.

Table 5.7: The best fit activities and half-lives of neutron-activated Xe isotopes. There are two initial activities N_i in SR1, corresponding to the concentration after the first and second AmBe calibration.

	^{125}Xe	^{127}Xe	^{129m}Xe	^{131m}Xe	^{133}Xe
$\tau_{1/2,\text{lit}}$ [d]	0.7	36.4	8.9	11.8	5.3
$\tau_{1/2,\text{SR0}}$ [d]	3.4 ± 0.1	4.0 ± 0.1	3.8 ± 0.1	4.3 ± 0.2	3.2 ± 0.2
$\tau_{1/2,\text{SR1}}$ [d]	3.1 ± 0.2	8.3 ± 6.9	8.2 ± 0.1	11.5 ± 0.2	5.7 ± 0.1
N_{SR0} [mBq]	49.4 ± 2.8	38.4 ± 2.2	58.9 ± 2.8	31.8 ± 1.9	11.9 ± 1.5
$N_{0,\text{SR1}}$ [mBq]	12.4 ± 0.6	0.7 ± 0.2	210.3 ± 0.7	160.3 ± 0.6	26.4 ± 0.4
$N_{1,\text{SR1}}$ [mBq]	9.9 ± 0.6	0.3 ± 0.2	147.9 ± 0.7	106.8 ± 0.6	18.1 ± 0.4

produced during the period used to constrain the rate, the decay has to be modeled using a Bateman equation [220]

$$N_{1125}(t) = N_{\text{Xe}125} \cdot \frac{\lambda_{\text{eff}}}{\lambda_{\text{eff}} - \lambda_{\text{Xe}125}} \cdot [\exp(-\lambda_{\text{Xe}125}t) - \exp(-\lambda_{\text{eff}}t)], \quad (5.16)$$

where $N_{\text{Xe}125}$, λ_{eff} , $\lambda_{\text{Xe}125}$ and c were kept as free parameters during the fit.

The remainder of the study follows that described in the previous section. The rate as a function of time was fitted, using Equation 5.16, and the fit parameters were bootstrapped to derive the uncertainty. The results are showcased in Figure 5.32. The derived half-lives have large uncertainties, as do the activities $N_{\text{SR0}} = (2.5 \pm 19.7)$ mBq, $N_{0,\text{SR1}} = (0.4 \pm 1.2)$ mBq and $N_{1,\text{SR1}} = (0.3 \pm 1.2)$ mBq, leading to the large uncertainty bands. In addition, assuming that the ^{125}I decay has a half life equal to the literature value of $\tau_{1/2} = 59.4$ d, it is possible to compute the purification lifetime of $\tau_{\text{pur}} = (2.86 \pm 3.68)$ d in SR0 $\tau_{\text{pur}} = (1.71 \pm 6.44)$ d in SR1³.

³The purification lifetime regarding oxygen contamination was estimated to be approximately

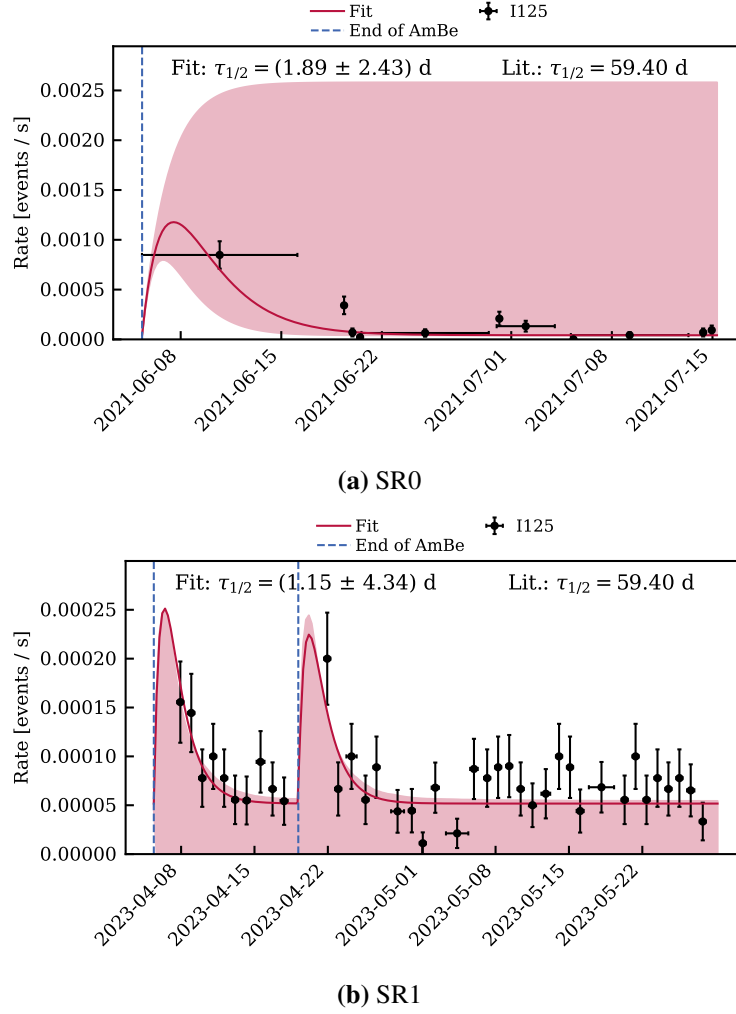


Figure 5.32: A fit of the ^{125}I decay. The blue dashed lines indicate the end of AmBe calibration campaigns. The red solid line shows the best fit values using the Bateman equation, and the red shaded region shows one standard deviation around the best fit. The uncertainty on the fit was derived via bootstrapping.

5.5.3 The Resulting Background Rates from Neutron Activation

The above studies result in functions of the expected number of events depending on time after the AmBe calibration for each isotope, which in turn can be integrated over the lifetime of each run used in the search for bosonic DM. But to do this, the rates derived from the fit have to be corrected for the coverage of the energy range and the LXe mass used to derive the counts. The energy selection in Figure 5.30 for example covers a 2σ interval, corresponding to about 95.5 % of the total counts, for which the fit derived number has to be corrected. For decays with multiple branches, the BR has to be taken into account in addition, and finally for the continuous ^{133}Xe spectrum, the rate has to be converted according to the fraction of the β spectrum used for the initial count. To compute the upper limit on the number of expected events, the sum of best fit and uncertainty of the fit parameters is used. In case the derived upper limit of the half life was found to be lower than the literature value, the literature value was used

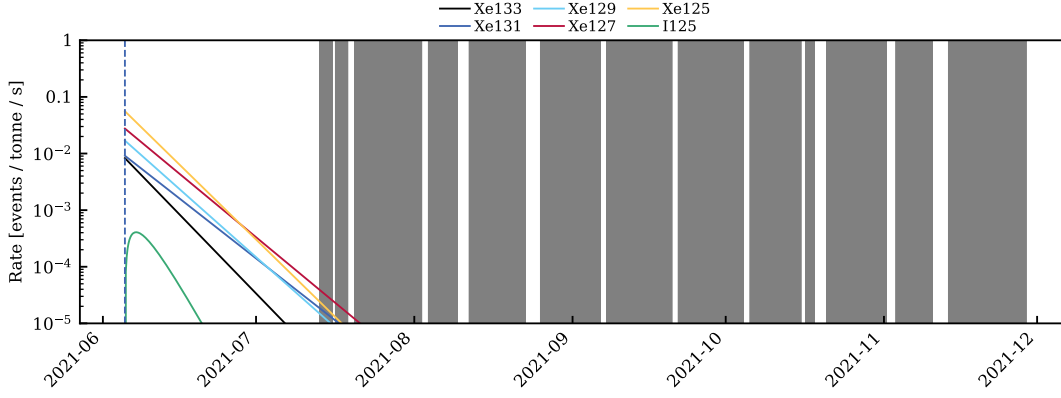
$\tau_{\text{pur}} \approx 1.8 \text{ d}$ [221].

Table 5.8: The expected rates of background events from neutron-activated isotopes, listed for the best-fit (BF) and the upper limit (UL) in units of $\text{t}^{-1}\text{y}^{-1}$. The number of events is computed for the full spectrum of the respective decay. Therefore, for the energy region split in SR1, the actually measured numbers will be reduced by the fraction of the spectrum that is included in the analysis.

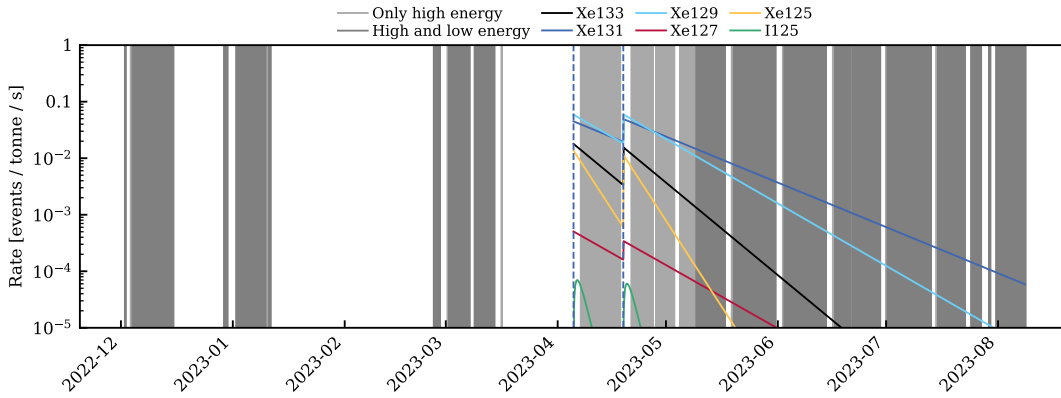
	^{125}I	^{125}Xe	^{127}Xe	$^{129\text{m}}\text{Xe}$	$^{131\text{m}}\text{Xe}$	^{133}Xe
SR0 BF	0.0	26.2	50.2	17.8	25.3	2.4
SR0 UL	0.0	30.4	136465.5	2352.0	3437.0	84.2
SR1 E_{low} BF	0.0	139.4	195.4	32300.8	60064.1	2846.5
SR1 E_{low} UL	0.0	175.5	2772.6	41145.0	64011.1	3011.2
SR1 E_{high} BF	-	14629.9	1406.8	-	-	39633.0
SR1 E_{high} UL	-	14849.2	3936.8	-	-	39825.2

instead.

The energy coverage corrected best-fit rates, as functions over time, can be seen in [Figure 5.33](#), where the ends of the AmBe calibration campaigns are shown as blue dashed lines, and shaded in gray are the runs that are used in the search for bosonic DM. In SR0, due to the successive $^{83\text{m}}\text{Kr}$ calibration campaign, there is a gap between the AmBe calibration and the science data, that is wide enough for most of the neutron activated isotopes to have decayed already. In SR1, there exists a significant amount of data before the first AmBe calibration, which is assumed to be free of any neutron activated backgrounds, however, the data immediately after the AmBe calibration contains a large amount of neutron activated decays. Because of this substantial background contribution, the analysis of the SR1 data was split into two: A low energy region which contains all events with $E \leq 300$ keV, and a high energy region with the events with a greater energy up to $E \leq 1000$ keV. Since the vast majority of the neutron activated background is in the low energy region, this split enables the removal of the most affected runs from the analysis, while still maintaining a maximum amount of data in the unaffected energy regions. To achieve a good overall sensitivity, the low energy region analysis did not include runs that were taken within 20 days of an AmBe calibration, while the high energy analysis included those runs. In addition, the run list for the high energy region does also contain runs after $^{83\text{m}}\text{Kr}$ calibrations, whereas all runs within 10 h after $^{83\text{m}}\text{Kr}$ were removed from the low energy region. This split is visualized by the two shades of gray in [Figure 5.33b](#). The expected numbers of events, normalized to the respective lifetimes of $T_{\text{SR0}} = 102.7$ d, $T_{\text{SR1,low}} = 101.4$ d and $T_{\text{SR1,high}} = 129.5$ d, using the best-fit and the upper limit are listed in [Table 5.8](#). The difference between the best-fit and the upper limit for the SR0 analyses is far larger than that for the SR1 case. Therefore, in the inference described in [section 5.7](#) the upper limit parameters from the fits were used to compute the nominal expectation of the rates of $^{129\text{m}}\text{Xe}$, $^{131\text{m}}\text{Xe}$ and ^{133}Xe in SR0, to be conservative due to the unsatisfying fit of those three isotopes.



(a) SR0: There is a gap between the AmBe calibration and the science data, that is wide enough that most of the neutron activated backgrounds have decayed away already.



(b) SR1: There exists a significant amount of data before the first AmBe calibration, which is assumed to be free of any neutron activated backgrounds. However, immediately after the AmBe calibration the rates of the neutron activated backgrounds are so large that the science data for SR1 is split into two: a low energy and a high energy region, and the runs in a time window of 20 days after AmBe are not included in the low energy analysis.

Figure 5.33: The best fit event rates from neutron activated isotopes as a function of time, for SR0 (Figure 5.33a) and SR1 (Figure 5.33b). The blue dashed lines indicate the end of AmBe calibration campaigns. The background data runs are shaded in gray.

5.6 Other Backgrounds

5.6.1 ^{85}Kr β Decay

^{85}Kr is an anthropogenic isotope of Kr that originates from nuclear activities and is found in the atmosphere, and as a contaminant in the LXe of XENONnT. ^{85}Kr decays via β decay with an endpoint energy of 687 keV and a half life of $\tau_{\text{Kr}85} = 10.8$ y with a BR of 99.563 %. The other possible branch is a β -decay of 173 keV followed quickly by a γ of 514 keV, with an intermediate lifetime of 1.015 μs . The concentration of natural Kr and ^{85}Kr inside the TPC is determined via rare gas mass spectrometer (RGMS) and validated using two data-driven methods, that both take advantage of the rare β - γ branch via delayed coincidence search.

The RGMS measurement indicated a Kr concentration in LXe of $c_{\text{Kr}} = (82 \pm 23)$ ppq for SR0 and $c_{\text{Kr}} = (165 \pm 66)$ ppq for SR1 and a ^{85}Kr per Kr abundance of $c_{\text{Kr}85} =$

$(2.12 \pm 0.13) \times 10^{-11}$ mol/mol for SR0 and $c_{\text{Kr}85} = (2.31 \pm 0.41) \times 10^{-11}$ mol/mol for SR1. The latter is close to the literature result for natural abundance of ^{85}Kr in natural Kr $c_{\text{Kr}85} = 2.33 \times 10^{-11}$ mol/mol [222].

The ^{85}Kr event rate⁴ can be computed via

$$A_{\text{Kr}85} = \frac{\ln 2}{\tau_{\text{Kr}85}} \cdot N_{\text{Xe}} \cdot c_{\text{Kr}} \cdot c_{\text{Kr}85} \cdot \text{BR}. \quad (5.17)$$

The final rate of expected background events from ^{85}Kr are $A_{\text{Kr}85} = (512.4 \pm 146.8) \text{ t}^{-1} \text{ y}^{-1}$ for SR0 and $A_{\text{Kr}85} = (1123.4 \pm 492.8) \text{ t}^{-1} \text{ y}^{-1}$ for SR1.

5.6.2 ^{124}Xe Double Electron Capture

^{124}Xe can undergo double electron capture (DEC), resulting in undetected neutrinos and a cascade of X-rays and Auger electrons caused by the filling of vacant electron shells. The relevant shells are the KK-shell leading to an energy of 64.3 keV with a BR of 76.5 %, the KL-shell with an energy of 37.4 keV and a BR of 22.0 % and finally the LL-shell with an energy of 10.0 keV and a BR of 1.6 %. The activity can be restrained from the isotopic abundance $c_{\text{Xe}124} = (10.13 \pm 0.56) \times 10^{-4}$ mol/mol and the half life of the DEC $\tau_{\text{Xe}124} = (1.1 \pm 0.3) \times 10^{22} \text{ y}$ [86]. The activity per branch is

$$A_{\text{Xe}124} = \frac{\ln 2}{\tau_{\text{Xe}124}} \cdot N_{\text{Xe}} \cdot c_{\text{Xe}124} \cdot \text{BR}, \quad (5.18)$$

leading to an expected activity (all branches combined) of $A_{\text{Xe}124} = (292.8 \pm 81.4) \text{ t}^{-1} \text{ y}^{-1}$.

5.6.3 ^{136}Xe double β Decay

In the rare event regime of the bosonic DM search one of the main background contributions comes from the two neutrino double beta decay ($2\nu\beta\beta$) of ^{136}Xe . The ^{136}Xe activity can be constrained just like the ^{124}Xe DEC from the natural abundance and the lifetime. The abundance $c_{\text{Xe}136}$ was measured to be (0.0897 ± 0.0016) mol/mol, and the lifetime is taken from the EXO200 experiment, $\tau_{\text{Xe}136} = (2.165 \pm 0.075) \times 10^{21} \text{ y}$ [223]. However, since the ^{136}Xe $2\nu\beta\beta$ endpoint energy of 2.46 MeV is beyond the energy ROI considered in this study, it is necessary to take into account the relative fraction of the spectrum that is considered. Therefore, the activity is

$$A_{\text{Xe}136} = \frac{\ln 2}{\tau_{\text{Xe}136}} \cdot N_{\text{Xe}} \cdot c_{\text{Xe}136} \int_{E_{\min}}^{E_{\max}} \frac{d\tilde{A}_{\text{Xe}136}}{dE} dE. \quad (5.19)$$

where $\tilde{A}_{\text{Xe}136}$ is the normalized spectrum of the $2\nu\beta\beta$. This results in a total expected activity of $A_{\text{Xe}136} = (81\,403.2 \pm 3171.8) \text{ t}^{-1} \text{ y}^{-1}$ below 1 MeV.

5.6.4 ^{83m}Kr γ Emission

^{83m}Kr is a double γ emitter that is used as a calibration source in XENONnT. The decay happens in two steps, releasing γ rays of 32.1 keV and 9.4 keV, with an intermediate lifetime of 154 ns in between [224]. Because of this short intermediate

⁴Of the β decay branch only, the β - γ branch is assumed to be removed by SS selections.

lifetime, the vast majority of ^{83m}Kr cascades are reconstructed as SS events of energy 41.5 keV.

The ^{83m}Kr is dissolved in the LXe and distributes uniformly in the TPC, which makes it an excellent calibration source to study spatially dependent effects. However, this also makes ^{83m}Kr an intrinsic background present everywhere in the detector. Because the half life of ^{83m}Kr is only $\tau_{1/2} = 1.83$ h, this background source dissipates quickly on its own. For that reason, 10 h after each ^{83m}Kr calibration session were removed from the data analyzed in this study, which leaves only 2.3 % of the initial ^{83m}Kr atoms. The final list of runs included in the analysis was discussed in [section 5.5](#). The background contribution of remaining ^{83m}Kr is assumed to be zero.

5.6.5 ^3H β Decay

In 2020, the XENON1T experiment reported an excess of low energy ER events, that was compatible with different BSM models, but was suspected to be originating from ^3H dissolved in the LXe [225]. In 2022, the low energy ER analysis of SR0 data, presented in [54], concluded that the XENON1T excess was not present in XENONnT data, and therefore does not stem from new physics, but rather from contamination with tritium atoms. In 2024, XENONnT published the measurement of coherent scattering of solar neutrinos, using SR0 and SR1 data, [73], where it was discovered that there is again a low energy excess visible in SR1 data, that matches a background contribution from ^3H β decay. It is suspected that the tritium was introduced into the TPC during an operation between SR0 and SR1.

The concentration of tritium in the LXe is taken from a study performed in the course of [3], that investigated data in a 1 cm region outside of the FV⁵. Using this sideband offers the advantage that it consisted of data not used in the main analysis, and can therefore be fitted without unblinding the study. Doing so provides a best fit result of $A_T = 152.61^{+78.72}_{-70.37} \text{ t}^{-1}\text{y}^{-1}$, which is consistent with the estimation used in [65].

5.6.6 Accidental Coincidences

So-called accidental coincidences (ACs) are incorrect pairings of S1 and S2 signals that are not from the same physical interaction. The AC events pose a dominant background for low energy analyses, but for energies above 20 keV they are practically nonexistent.

The number of AC events was determined by studying isolated S1 and S2 peaks, meaning peaks where the corresponding partner got lost and not recorded. Since the AC events consist mainly of such isolated peaks getting paired, the overall rate can be constrained by determining the rate of the isolated S1s and S2s. What remains is the efficiency of our cuts, mainly the ER band selection and the S2 width cut, at removing ACs, which is determined via so-called *salting*. In this process, isolated peaks are extracted from the data, paired with a randomly chosen drift time and planted back into the data stream at random times, as artificial ACs and finally propagated through

⁵That is the FV from [3], not the one derived for this study.

the standard analysis pipeline. Via the ratio of salted to surviving ACs it is possible to determine the rate of remaining AC events after cuts.

For the study in [3], the rate of AC events was computed to be $A_{AC} = (5.22 \pm 1.74) \text{ t}^{-1}\text{y}^{-1}$ for SR0 and $A_{AC} = (3.74 \pm 0.53) \text{ t}^{-1}\text{y}^{-1}$ for SR1. However, since [3] uses additional AC removal cuts that are discarded in this study, the rate of AC events will be left unconstrained in the final fit.

5.6.7 Solar Neutrinos

The recoil energy spectrum of solar neutrino induced events in the detector is

$$\frac{dR_i}{dE_R} = \sum_{j \in \{e, \mu, \tau\}} \int P_{j \rightarrow e} \frac{d\Phi_i}{dE_\nu} \frac{d\sigma_j}{dE_R} dE_\nu, \quad (5.20)$$

where i is the label of the process producing the neutrinos, such as proton-proton fusion (pp) or ^8B decay, j denotes the neutrino flavor, $P_{j \rightarrow e}$ is the oscillation probability of oscillating into electron flavor, or in the case of emitted electron neutrinos the survival probability, $d\Phi_i/dE_\nu$ is the differential neutrino flux from source process i and finally $d\sigma_j/dE_R$ is the interaction cross section.

The search for bosonic DM uses the free electron assumption (FEA) including the so-called stepping approximation to compute the interaction cross section. At low energies, there exist more accurate approximations such as the Relativistic Random Phase Approximation (RRPA) [226], however the difference between the two approximations is not relevant for the work presented here. The FEA assumes that all 54 electrons of the xenon atoms are free particles, while the stepping approximation models the atomic structure as steps in the energy range where more electrons become available to participate in the scattering.

The solar neutrino rate is dominated by the pp fusion process, with the electron capture of ^7Be as the second largest contributor being one order of magnitude smaller. The next greatest solar neutrino production process in terms of rate is nearly two more orders of magnitude smaller, nearly 3 orders of magnitude smaller than the pp induced neutrino rate. For this reason, the solar neutrino induced background is dominated by the pp neutrino based background for energies up to $\sim 300 \text{ keV}$, after which the pp neutrino rate drops to zero and the remaining neutrinos originate from ^7Be .

The solar neutrino flux and survival probabilities were taken from the measurements of the Borexino experiment [227, 228]. Following the procedure of [49], a 3 % uncertainty is assigned to the rate of solar neutrino events and the final rate was determined to be $A_\nu = (410.7 \pm 12.3) \text{ t}^{-1}\text{y}^{-1}$.

5.6.8 ^{214}Bi Skin Background

The so-called ^{214}Bi skin background is a spatially dependent background, similar to the background originating from detector material radiation. ^{214}Bi is a daughter of the ^{222}Rn decay chain, and as such is distributed everywhere inside the LXe TPC,

like the lead backgrounds described in [section 5.4](#). The decay of ^{214}Bi is a β -decay to ^{214}Po , which has a half life of $\tau_{1/2} = 164\mu\text{s}$, before it decays to ^{210}Pb . Most β - α decay cascades are either identified as a MS event and subsequently removed from the analysis, or if they are merged into one signal are far outside the energy ROI. ^{214}Bi also has the chance to decay to an excited state of ^{214}Po , releasing a practically immediate γ . For this γ , the same basic considerations are valid, it is either identified as an MS or outside of the ROI. However, there exist uninstrumented LXe areas, called the *skin*, in between the PTFE and the cryostat, and below the cathode. A ^{214}Bi decay in the skin can happen, where the β and α particles are absorbed by the detector materials, while the γ travels into the TPC, depositing its energy there and leading to a non-filtered background event.

The ^{214}Bi skin background can be constrained via full-chain simulations of ^{214}Bi decays in the uninstrumented LXe volume, the spatial extent of which is illustrated in [Figure 5.35](#), where the initial positions of 1×10^5 simulated ^{214}Bi skin events are shown. The black lines indicate the gate and cathode electrodes at the top and bottom respectively, and the PTFE wall on the right, which form the sensitive volume of the detector.

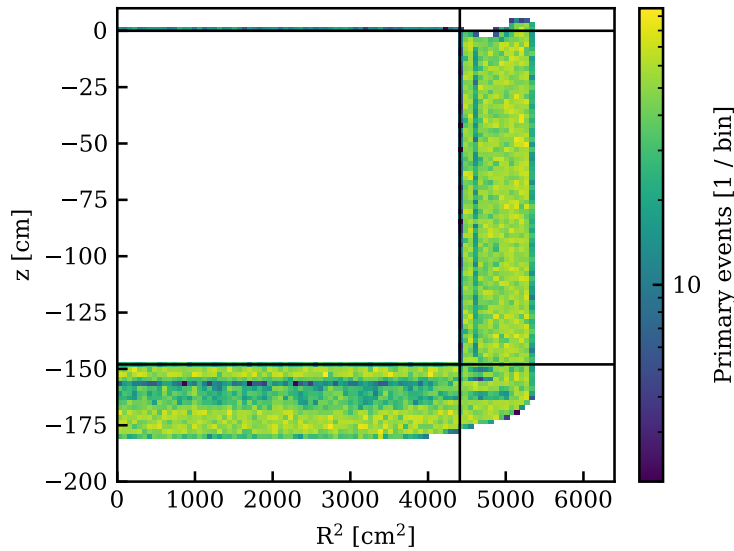


Figure 5.34: The spatial distribution of the primary positions of 1×10^5 ^{214}Bi skin decay sites. The black lines indicate the gate at the top, the cathode at the bottom and the TPC wall at the outer edge.

In total, 1×10^7 decays were simulated and propagated through the full processing chain. Due to the way that the simulation software is set up, storing the primary positions is resource intensive. For this reason, the initial positions used for visualization in [Figure 5.34](#) were only stored for a subset of the simulations. Of the 1×10^7 total simulated initial decays, 460156 in SR0 and 519705 in SR1 resulted in an SS event inside of the TPC. That means the chance for a ^{214}Bi skin decay to produce a measurable, meaning not absorbed by any detector part, SS event is $f_{\text{skin}} = (4.60 \pm 0.01) \%$ in SR0 and $f_{\text{skin}} = (5.20 \pm 0.01) \%$ in SR1. Assuming that the activity of ^{214}Bi is the same as that of ^{214}Pb , the rate of ^{214}Bi skin events can be computed via

$$A_{\text{Bi}214} = A_{\text{Pb}214} \cdot m_{\text{skin}} \cdot f_{\text{skin}}, \quad (5.21)$$

where $A_{\text{Pb}214}$ is the ^{214}Pb activity in units of event per tonne per year, derived in [section 5.4](#) and m_{skin} is the mass of the LXe skin volume, which is $m_{\text{skin}} = (2.220 \pm 0.006) \text{ t}$ in SR0 and $m_{\text{skin}} = (2.219 \pm 0.006) \text{ t}$ in SR1. The differences between SR0 and SR1 stem from differences in the LXe density and a different liquid level above the gate, which was lowered in SR1. The mass uncertainty of the skin volume is taking the uncertainty of the LXe density and the statistical uncertainty on the volume into account, not the position correction or electric field uncertainties that were discussed in [section 5.3](#). Evaluating [Equation 5.21](#) results in a rate, $A_{\text{Bi}214}$, in units of events per *detector* per year, i.e. a rate that is not yet scaled to the mass of LXe used in the simulation.

The rate of expected SS events originating from the ^{214}Bi skin background for SR0 is shown in [Figure 5.35](#) distributed in space (left) and as a function of energy (right). The majority of the ^{214}Bi skin background originates from the LXe volume in the lower dome below the cathode. The rates of events inside the FV are $A_{\text{Bi}214} = (13.3 \pm 1.4) \text{ t}^{-1}\text{y}^{-1}$ in SR0 and $A_{\text{Bi}214} = (7.5 \pm 0.6) \text{ t}^{-1}\text{y}^{-1}$ in SR1.

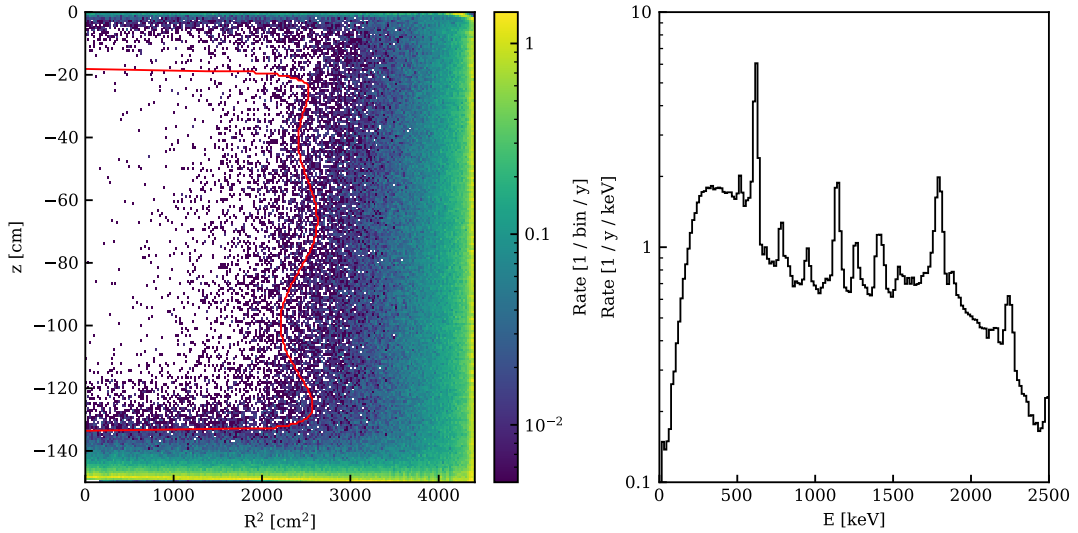


Figure 5.35: Left: The spatial distribution of background events originating from ^{214}Bi decays in the LXe skin in SR0. The optimized FV is shown in red. Right: The energy spectrum of the ^{214}Bi skin background.

5.7 The Sensitivity Projection for Bosonic Dark Matter in XENONnT

The previous sections describe the derivation of the signal and background models, that were used to compute the 90 % CL exclusion sensitivity for bosonic DM. The number of expected events was based on SR0 and SR1 data of XENONnT, using a FV of mass $m_{\text{FV}} = (2.35 \pm 0.07) \text{ t}$ in SR0 and $m_{\text{FV}} = (2.47 \pm 0.07) \text{ t}$ in SR1, and lifetimes of $T_{\text{SR0}} = 102.7 \text{ d}$, $T_{\text{SR1,low}} = 101.4 \text{ d}$ and $T_{\text{SR1,high}} = 129.5 \text{ d}$. The inference procedure, using the *alea* software, and the projected sensitivity are detailed in this section, starting with a summary of the expected signal and background models and their implementation into the statistical framework.

Table 5.9: The expected activities, A_c , of the background components, c , in units of $[\text{t}^{-1}\text{y}^{-1}]$ in the respective energy ROI.

Source	SR0	SR1 $E \leq 300 \text{ keV}$	SR1 $E > 300 \text{ keV}$
^{85}Kr	512.4 ± 146.8	701.1 ± 307.6	422.1 ± 185.1
^{124}Xe	292.8 ± 81.4	292.8 ± 81.4	0
^{136}Xe	50391.1 ± 1963.4	5080.4 ± 198.0	45275.0 ± 1764.1
^{83m}Kr	0	0	0
^3H	0	$152.6^{+78.7}_{-70.4}$	0
AC	5.2 ± 1.7	3.7 ± 0.5	0
Solar ν	410.7 ± 12.3	410.7 ± 12.3	0
^{214}Pb	28648.9 ± 2862.0	1478.6 ± 98.4	13392.8 ± 891.4
^{212}Pb	1021.7 ± 330.0	414.0 ± 82.0	444.1 ± 88.0
^{133}Xe	84.2 ± 84.2	2626.0 ± 151.9	3004.1 ± 14.6
^{131m}Xe	3437.0 ± 3411.7	60064.1 ± 3947.0	0
^{129m}Xe	2352.0 ± 2334.2	32300.8 ± 8844.3	0
^{127}Xe	50.2 ± 50.2	104.2 ± 104.2	656.8 ± 656.8
^{125}Xe	26.2 ± 4.2	92.3 ± 23.8	4946.8 ± 74.2
^{125}I	0	0	0
Materials	39809.9 ± 5894.3	35.3 ± 4.8	38526.5 ± 5207.6
^{214}Bi	2.5 ± 0.3	0	1.4 ± 0.1

5.7.1 Summary and Formulation of the Signal and Background Models

The background model, μ_{bgd} , consists of the sum of all the background components, c , that were listed in the previous sections

$$\mu_{\text{bgd}}(E) = \sum_c A_c \cdot f_c(E), \quad (5.22)$$

where A_c denotes the expected activity per component and $f_c(E)$ the respective spectrum. The expected activities of all components, including their uncertainties, are listed in Table 5.9. The listed values may differ from the ones described in the above sections, because the table here presents the expected rate in the energy ROI of the individual partitions, SR0, SR1 low energy and SR1 high energy. For example, section 5.6 lists the activity expected from the $2\nu\beta\beta$ of ^{136}Xe as $A_{\text{Xe}136} = (81\,403.2 \pm 3171.8) \text{ t}^{-1}\text{y}^{-1}$. However, of the integrated spectrum of the $2\nu\beta\beta$, only $\sim 62\%$ lie within the ROI of SR0, leading to $(50\,391.1 \pm 1963.4) \text{ t}^{-1}\text{y}^{-1}$ expected ^{136}Xe events in SR0.

The background models for SR0 and SR1 are shown in Figure 5.36 and Figure 5.37 respectively, in dependence of the reconstructed energy, meaning the spectra are shifted by the energy bias given in subsection 5.1.1. The SR1 background model shows the expected rates of the low and high energy regions combined, which leads to the gap at 300 keV. At energies below 500 keV, the background spectra are complex, with features originating from a multitude of different sources, while at higher energies the only remaining relevant background sources are ^{136}Xe , ^{214}Pb and the detector materials.

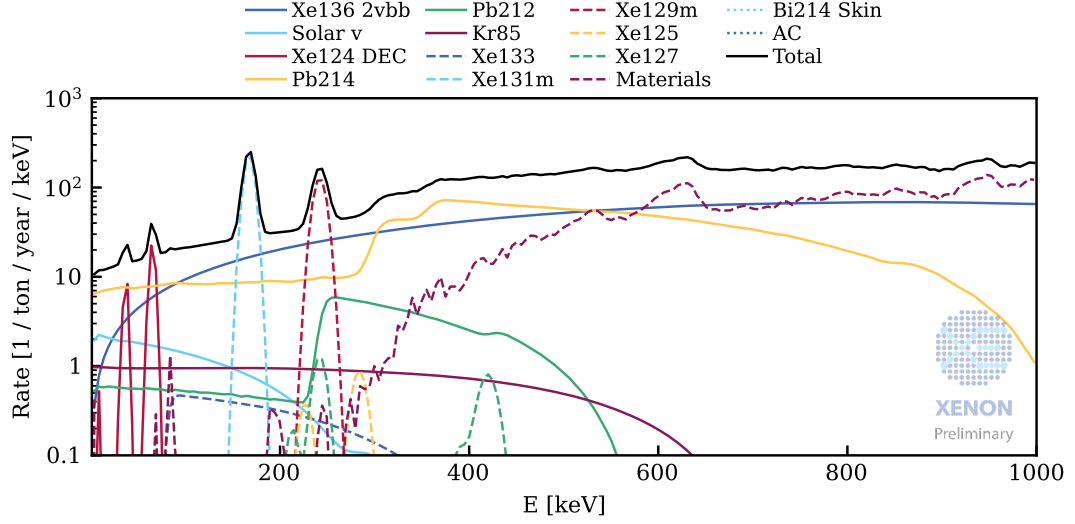


Figure 5.36: The background model of SR0.

The hypothetical signal to be searched for is the absorption of a bosonic DM particle, for instance an ALP or DP. The signature of such an absorption is a monoenergetic energy deposit, the spectrum of which is a Skew Gaussian of the form

$$\mu_{\text{sig}}(E, N) = N \cdot \mathcal{L}_{\text{SG}}(\mu = E, \sigma = \rho(E), \alpha = \alpha(E)), \quad (5.23)$$

where $\mathcal{L}_{\text{SG}}$ is the PDF of a Skew Gaussian distribution and $\rho(E)$ is the energy resolution. The energy resolution and the skewness parameter α are detailed in [subsection 5.1.1](#). Furthermore, E is the mass of the particle and N is the number of expected events, that can be converted to the coupling strength. Both, potential signal and background model, have to be multiplied with the signal efficiency. The signal efficiency is the product of the thresholds described in [subsection 5.1.2](#), the EBE that is detailed in [subsection 5.1.3](#) and the cut acceptance explained in [section 5.2](#), with each having an upper and lower uncertainty, that is added in quadrature to reach a combined uncertainty. The signal efficiency and its uncertainty are implemented in the inference framework via a *shaper* variable, f_{eff} , that is a free parameter in the fit. [Figure 5.38](#) visualizes the working principle of the shaper. The figure shows the SR0 signal efficiency as a black line, and the combined uncertainty shaded in gray. The resulting shaped efficiency is computed via

$$\epsilon_{\text{sig}}(E) = \hat{\epsilon}(E) + f_{\text{eff}} \cdot \sigma_{\text{sig}}, \quad (5.24)$$

where $\hat{\epsilon}$ is the mean signal efficiency and $\sigma_{\text{sig}}(E)$ the uncertainty.

5.7.2 Statistical Framework

A binned likelihood maximization was used in this analysis, implemented in the *alea* software, using the likelihood test formulae described in [\[122\]](#) and [\[113\]](#). The parameter of interest, N , is defined as the number of expected events from absorption of bosonic DM. The base hypothesis H_0 , called the *background-only* hypothesis, is the statistical model devoid of any DM signal, meaning $N = 0$.

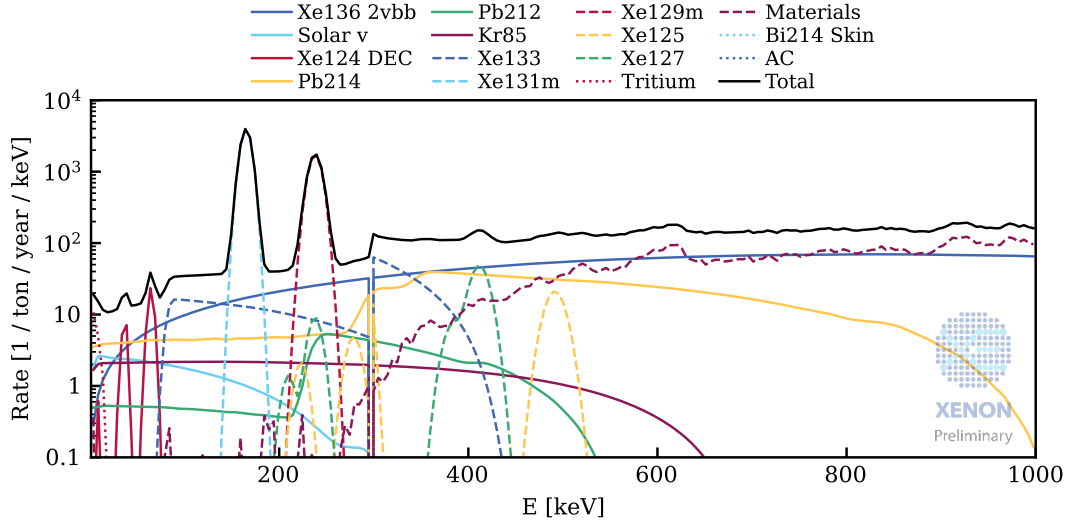


Figure 5.37: The background model of SR1, with the low and high energy regions combined. The split in low and high energy is the cause of the gap at 300 keV.

Given the signal- and background model explained in [subsection 5.7.1](#), the number of expected events, μ , is

$$\mu(E|\theta, N) = \epsilon_{\text{sig}}(E, N) \cdot [\mu_{\text{sig}}(E) + \mu_{\text{bgd}}(E)], \quad (5.25)$$

where θ is the set of all nuisance parameters, namely the efficiency shaper f_{eff} and the activities of each background component A_c . The likelihood of observing data, x , given the expectation value, μ , can be expressed as the product of Poissonian probabilities per bin

$$\mathcal{L}(x|\theta, N) = \mathcal{L}(x|\mu(\theta = \theta, N = N)) = \prod_i \frac{\mu_i^{x_i} \cdot \exp(-\mu_i)}{x_i!}. \quad (5.26)$$

The values μ_i and x_i are the expected and measured number of events in bin i respectively. The negative logarithm of [Equation 5.26](#) is minimized to find the best-fit parameters, $\hat{\theta}$, and $\hat{N}$, which are those that maximize the likelihood $\mathcal{L}(x|\theta, N)$. The test statistic $q_\mu(N)$ is the PLR between the likelihood assuming a fixed value for N , and the best-fit likelihood

$$q(N) = -2 \ln \left(\frac{\mathcal{L}(x|\theta_{\text{max}}, N = N)}{\mathcal{L}(x|\hat{\theta}, \hat{N})} \right), \quad (5.27)$$

where θ_{max} are those nuisance parameters that maximize [Equation 5.26](#) for the given value of N .

Under the asymptotic assumption of Wilk's theorem, [\[124\]](#), $q(N)$ follows a χ^2 distribution with one degree of freedom. However, especially in bins with a low total event count, this asymptotic assumption does not always hold. For the derivation of the sensitivity, the asymptotic assumption is still assumed, meaning that the 90 % CL threshold to exclude a signal strength is computed from the 90 % percentile of the χ^2 -CDF, which is at $q(N) = 2.71$. In [Appendix F](#) it is shown that doing so slightly overestimates the upper limit. The sensitivities derived using the asymptotic assumption

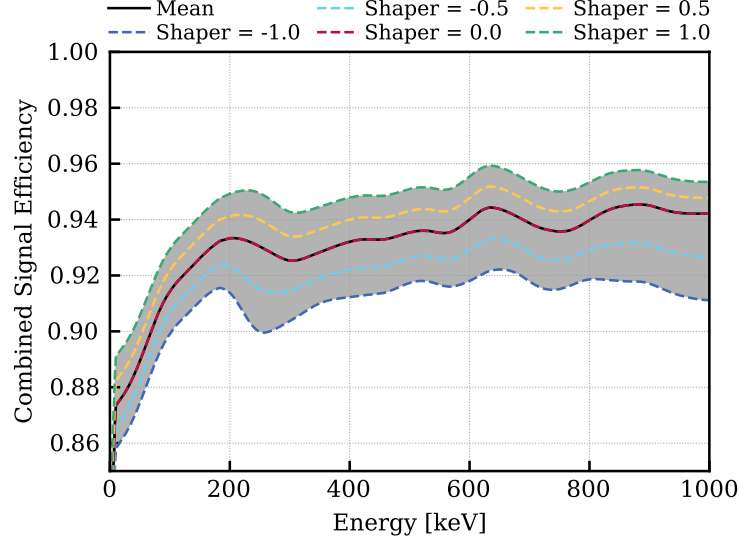


Figure 5.38: The signal efficiency of SR0 and a visualization of the shaper variable. The mean efficiency is shown in black, with the shaded region indicating the uncertainty. The colored dashed lines show the resulting efficiency when using different values for the shaper.

can therefore be seen as conservative limits. The confidence interval is determined by sweeping across the parameter N and computing at which value of N the test statistic is $q(N) = 2.71$. Such a parameter sweep is visualized in Figure 5.39 on the example of two toy models. The first example, which resulted in the black curve, shows the values of $q(N)$ for a model with a best-fit of $\hat{N} = 12$. The 90 % confidence interval is determined by those values of N , where $q(N) = 2.71$, visualized by the red line. The interception between red and black lines is at $N < 0$, however a negative number of signal events is unphysical. Therefore the lower limit of the 90 % confidence interval is set at $N = 0$, and the upper limit at the right-hand interception between black and red lines, indicated by the black dashed line. For the second example with $\hat{N} = 20$, shown in blue, both lower and upper limit can be determined by the interception between the $q(N)$ curve and the threshold.

5.7.3 Sensitivity Projection

The projected sensitivity was computed via toy MC simulations. The *alea* software allows to sample events according to the background model, μ_{bgd} , providing realistic data resembling the background-only hypothesis, H_0 . For each mass of potential bosonic DM particle, 1000 toy MC datasets were generated, and the parameter sweep to compute the 90 % confidence interval was performed, as described in subsection 5.7.2. Doing so resulted in 1000 upper limits for the number of interactions, N , per mass, that can be excluded with a 90 % CL, provided the data follows H_0 . Given N , the upper limit can be converted to the axioelectric coupling and the kinetic mixing by rearranging Equation 2.16 and Equation 2.17 respectively, resulting in

$$g_{\text{ae}} = \sqrt{\frac{N}{5.5 \times 10^{24} \text{ t}^{-1} \text{ y}^{-1}} \cdot A \cdot \left(\frac{\text{keV}/c^2}{m_{\text{a}}} \right) \frac{\text{b}}{\sigma_{\text{pe}}}} \quad (5.28)$$

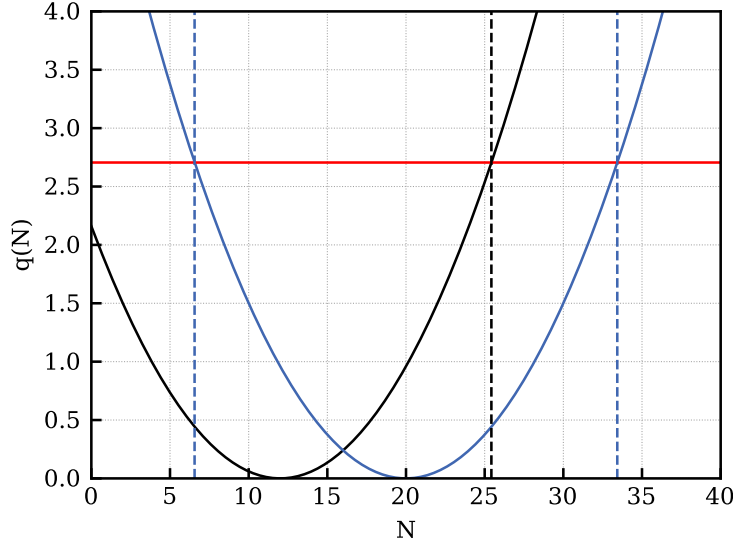


Figure 5.39: The determination of the 90 % confidence interval, based on the test statistic $q(N)$. The red line depicts the threshold below which 90 % of all points of a Poissonian PDF with one degree of freedom lie. Two example distributions of $q(N)$ are shown: The black curve shows a model with a best-fit of $\hat{N} = 12$. Since only values of $N \geq 0$ are allowed, and the lower limit of the 90 % confidence interval is at $N = 0$. The upper limit is at that N at which the curve crosses the red threshold, indicated by the black dashed line. The blue curve shows an example with $\hat{N} = 20$, for which both the lower and upper limit of the confidence interval are indicated by the N values at which the blue curve intercepts the red line.

and

$$\kappa = \sqrt{\frac{N}{1.7 \times 10^{29} \text{ t}^{-1} \text{ y}^{-1}} \cdot A \cdot \left(\frac{m_{A'}}{\text{keV}/c^2} \right) \frac{\text{b}}{\sigma_{\text{pe}}}}, \quad (5.29)$$

where A is the number of nucleons of xenon and σ_{pe} is the photoelectric cross section from [89].

The obtained 90 % CL exclusion sensitivity is visualized in Figure 5.40 for the axioelectric effect and in Figure 5.41 for the kinetic mixing. The median 90 % CL exclusion limit is depicted as a black line, and the 1σ and 2σ bands are shaded in green and yellow respectively. The 1σ (2σ) band is determined by the 16 % and 84 % (2.5 % and 97.5 %) percentiles of the upper limits. The current best limits are from the XENONnT SR0 analysis, [54], drawn in blue and from the PandaX analysis, [72], in purple. Below 140 keV for g_{ae} and 200 keV for κ , the best limits are those from XENONnT SR0, and a comparison to this study is best performed by examining the left-hand figures. In general, the comparison yields that including SR1 data in the analysis does not improve the sensitivity in the low energy region, and the best limit of the SR0 study lies even below the 2σ band of this study for the lowest potential masses. One reason for the worse sensitivity of this study when compared to the SR0 result at the low energy edge is the non-negligible contamination of tritium present in SR1. Another reason is the conservative approximation of the 90 % threshold that is used in this work and detailed in Appendix F. However, a significant improvement of this study was the inclusion of the energy region around the $^{83\text{m}}\text{Kr}$ peak at 41.5 keV. By removing data recorded immediately after the $^{83\text{m}}\text{Kr}$ calibrations from the analy-

sis, the sensitivity in the energy vicinity of the ^{83m}Kr peak was enhanced, resulting in a potential new world leading limit.

At energies above 140 keV for g_{ae} and 200 keV for κ , the best published limits come from the PandaX experiment, detailed in [72]. The 90 % CL upper limit presented in this thesis is consistently lower than that value. There are two sharp edges observable in the upper limits determined in this study, as well as in the PandaX results, that correspond to the energies of the ^{131m}Xe and ^{129m}Xe neutron activated decays at 163.9 keV and 236.2 keV respectively. It can be assumed that removing more data following the AmBe calibrations would improve the sensitivity in this energy region even further. Using the same logic, further constraining the rate of ^{127}Xe (^{125}Xe) can be expected to improve the sensitivity around the 408 keV (487 keV) peak. The remaining increments in the sensitivity originate from the material background, in particular from ^{214}Bi in the ^{226}Ra decay chain around 609 keV and ^{228}Ac in the ^{232}Th decay chain at 911 keV and 969 keV. Further lowering the material background requires shrinking the FV, which lowers the total exposure of the study.

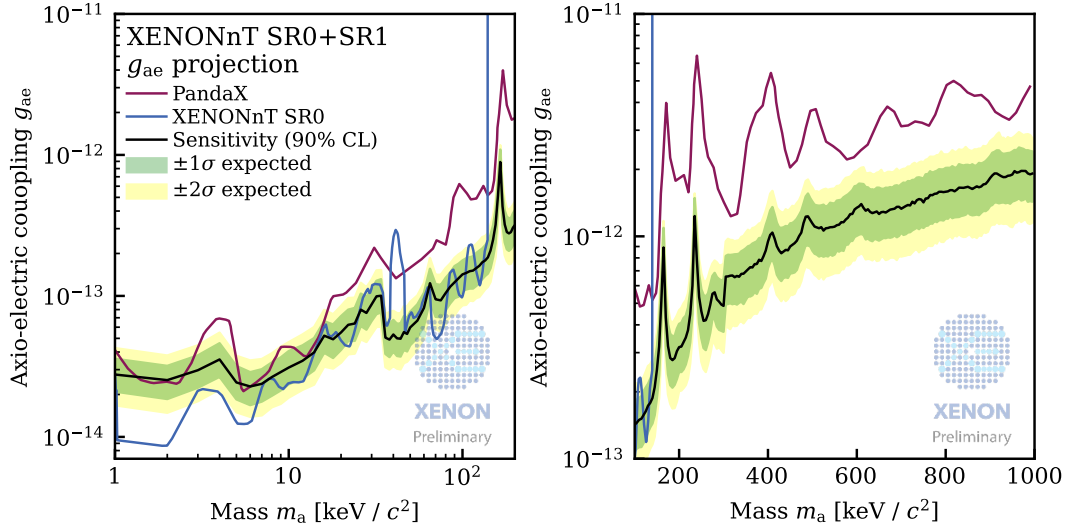


Figure 5.40: The projected sensitivity of XENONnT for the axioelectric coupling constant g_{ae} , using data from SR0 and SR1. The median 90 % CL exclusion limit is shown as a solid black line, with the 1σ and 2σ bands in green and yellow respectively. The purple line indicates the PandaX upper limit from [72], and the blue line that of XENONnT SR0 from [54]. The left panel shows the energy region between 1 keV and 200 keV in a logarithmic scale, while the right panel presents the energy region between 100 keV and 1000 keV in a linear scale.

5.8 Conclusion and Outlook

This chapter describes all the components that are necessary to conduct a search for bosonic DM signals in an energy range from 1 keV to 1000 keV, using the SR0 and SR1 data of XENONnT. The interaction is assumed to be analogous to the photo-effect, leading to a monoenergetic DM signal, whose shape is determined by the energy reconstruction process described in subsection 5.1.1. The utilized signal model is affected by two further factors, the EBE and the cut acceptance. The EBE was estimated by using the peak shadow approximation, a procedure developed in particular

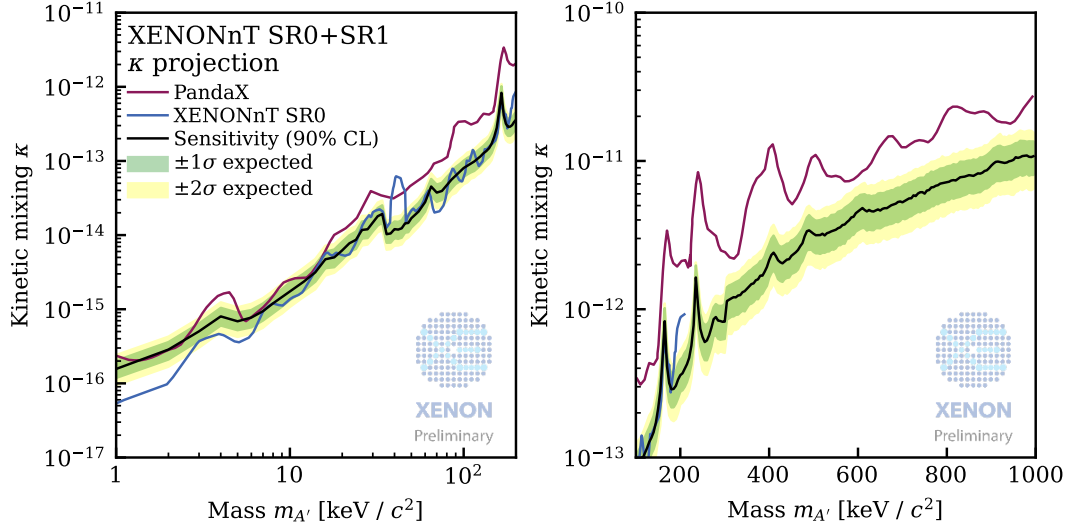


Figure 5.41: The projected sensitivity of XENONnT for the kinetic mixing between dark photons and SM photons, κ , using data from SR0 and SR1. The median 90 % CL exclusion limit is shown as a solid black line, with the 1σ and 2σ bands in green and yellow respectively. The purple line indicates the PandaX upper limit from [72], and the blue line that of XENONnT SR0 from [54]. The left panel shows the energy region between 1 keV and 200 keV in a logarithmic scale, while the right panel presents the energy region between 100 keV and 1000 keV in a linear scale.

for this study, that provides a quickly computable estimate for the EBE for high energy regions. The cut acceptance was evaluated on a combination of ^{220}Rn and ^{232}Th calibration data, and a full-chain simulation of signal-like events. The final cumulative acceptance was smoothed using a Gaussian kernel.

For the background estimation, a large number of components could be adapted from the analysis of the solar neutrinos from proton-proton fusion detailed in [3]. Moreover, dedicated studies were performed for the material originating background, the decays of ^{212}Pb and ^{214}Pb , short-lived isotopes from neutron activation and the ^{214}Bi skin background. The estimation of the material background was based on simulation campaigns and the radioactivity screening of the detector components. The simulated backgrounds were validated on the MS spectrum, and slight discrepancies between data and simulation were counteracted by scaling the non-matching rates belonging to the isotopes ^{137}Cs , ^{235}U and ^{232}Th by least-squares minimization. This simulated material background was also utilized to develop a fiducial volume that optimizes the potential significance of a DM signal. Next, the decays by the lead isotopes ^{212}Pb and ^{214}Pb were analyzed. These lead isotopes decay via a β - γ cascade, which can get reconstructed as two separate events, or as a single merged event. To determine the fraction and energy spectrum of events that get reconstructed as SS, and therefore pose a background to the bosonic DM search, full chain simulations of the decays were performed and evaluated. The same procedure was performed for the ^{214}Bi LXe skin background, events originating by a ^{214}Bi decay in an uninstrumented xenon volume of the detector, that can lose part of their signal to absorption by detector materials. Finally, the decays of neutron activated isotopes were analyzed and their rates constrained by fitting the decrease in event count over time with decay functions.

The signal and background models were implemented in the *alea* software, which allows for the sampling of toy MC data according to the model's PDF. A computation of the 90 % confidence interval was performed for the event rate from interactions induced by bosonic DM in a range of DM masses from 1 keV to 1000 keV. The resulting projected median 90 % CL exclusion limits for the axioelectric coupling g_{ae} and the DP kinetic mixing parameter κ are shown in [Figure 5.40](#) and [Figure 5.41](#) respectively. The analysis presented in this thesis demonstrates, that by extending the energy region and including data from SR1, XENONnT is sensitive to an axioelectric coupling constant and a dark photon kinetic mixing parameter lower than any published results in the energy region from 140 keV to 1000 keV. Lastly, the full analysis framework described in this chapter and required to perform the final fit of model to data was provided to the XENON collaboration.

Summary

Across all observable scales, there exist measurements of astronomical phenomena that suggest that the universe includes much more mass than the visible baryonic matter can account for. The nature of this additional *Dark Matter* (DM) is one of the major puzzles of modern physics. Theories that explain DM as beyond the standard model (BSM) particles include weakly interacting massive particles (WIMPs), or bosonic DM such as axion-like particles (ALPs) or dark photons (DPs). One of the world-leading experiments for the direct detection of particle DM is the XENONnT experiment, located at the LNGS underground laboratory in Italy.

The XENONnT detector is a liquid xenon (LXe) dual-phase time projection chamber (TPC). The TPC consists of a large cylindrical volume of LXe that serves as the main target, and a smaller volume of gaseous xenon (GXe) on top. A particle can interact and deposit energy on the xenon atoms, causing them to scintillate and ionize. The scintillation photons are recorded by photomultiplier tubes (PMTs) at the top and bottom of the TPC, and the resulting signal is called the S1. The ionization electrons are drifted along the electric drift field to the top of the TPC, where they are extracted by a second electric field into the GXe, eventually leading to electroluminescence. This secondary light pulse is also recorded by the PMTs, and called the S2 signal. The energy of the initial energy deposit can be determined from the signal strength, called *area*, of the S1 and S2, while the ratio of the two signals provides information about whether the interaction was an electronic recoil (ER) or a nuclear recoil (NR). The vertical depth of the event is inferred from the time between S1 and S2, and the horizontal position is determined from the pattern recorded by the top PMTs caused by the S2, called the *S2 hitpattern*.

6.1 Probabilistic Position Reconstruction using Conditional Normalizing Flows

The position reconstruction, which is the derivation of the horizontal position from the S2 hitpattern, was the topic of [chapter 3](#). The goal was the development and

implementation of a probabilistic machine learning (ML) model, that provides position probability information across the full horizontal plane, as opposed to the point-estimate that was implemented previously. In particular, the purpose of the enhancement of the position reconstruction model was to gain additional information beyond the position. Improvements in the accuracy of the model were only a secondary aim.

At first, I implemented two methods that can extend any arbitrary feed-forward neural network to make it predict a probabilistic output and tested them on an example multilayer perceptron (MLP). The One-Hot (OH) method splits the output space into distinct bins, and treats the position reconstruction as a classification task, with each bin representing a separate class. The per-class pseudo-probabilities can then be interpreted as a probability distribution of the event position. The second method was replacing the point-like output of the network with the parameters of a predefined analytical probability density function (PDF), and using (the negative log of) that PDF as a loss function. The advantage of the OH model is that it is completely agnostic towards the probability distribution, while the parameterization approach requires a fixed analytical PDF. On the other hand, the parameterized model requires only a minor increase in parameters, compared to a base model, as opposed to the OH model, where the number of parameters scales with the inverse square of the bin width. A very fine binned OH model significantly slows down the training and inference process, as well as drastically increases the chance of overtraining.

To take advantage of both approaches, a fine-binned OH model was trained, and its predictions of example hitpatterns were fitted with a selection of pre-defined analytical PDFs. The best-matching function was a Skew Gaussian (SG) distribution, which was then taken as the loss function of a parameterized model. This resulting SG model was slightly more accurate in predicting the position. In addition, it provided a PDF that was shown to have the correct uncertainty coverage within 5 %. The ellipses predicted by the SG model were circular in the TPC bulk, and elongated in radial direction when close to the TPC walls.

Another approach, the one that was ultimately implemented in SR2 of XENONnT, is to model the position reconstruction with a conditional normalizing flow (CNF), a type of ML model that is different from the classical neural networks. A CNF works by mapping a known and simple base distribution to an unknown and complex target distribution via the use of bivariate transformations. The resulting CNF position reconstruction model is both, small in terms of required parameters and thereby fast, and able to predict an arbitrary probability distribution. The CNF is the most accurate position reconstruction model so far, in particular near the TPC edges. While accessible to analysts, the full predicted position PDF was not stored by default in SR2, but rather, a polygon that corresponds to the equiprobability contour at one standard deviation. It was shown that a polygon with $N = 16$ corners achieves extremely well coverage, with a maximum overcoverage of only 2 %.

So far, the position reconstruction models were trained and validated on simulations only, and for this reason two data-driven approaches were used to validate the resolution. S2 signals originating from below the perpendicular wires, or from the so-called hotspots in the TPC were assumed to be caused by point-like sources (or line-like in

the case of the perpendicular wire events). By fitting the reconstructed positions of those events with Gaussian distributions, the spatial spread of the predicted positions was determined. The data-driven and simulations driven resolution, as well as the uncertainty predicted by the CNF itself, matched within 0.05 cm for S2s with an area greater than 100 PE. A secondary result of this study was that the hotspot S2s do indeed originate from a single position, while the perpendicular wire S2s seem to have some additional effect widening their spatial distribution.

The additional position uncertainty information provided by the CNF model has already been adapted by other XENONnT scientists into their analyses. The iterative Bayesian S2 merging and the fuzzy spatial selection are two examples of that, summarized in this work. In addition, proof of principle studies were presented in this thesis of potential future use-cases of the CNF model, including using the position uncertainty as a data quality measure, using it to identify potential multi-scatter (MS) events, and attempting to derive the horizontal diffusion constant of electrons in LXe.

6.2 Quality Assurance of the XENONnT Upgrade Cathode

In summer 2025, the XENONnT experiment underwent an upgrade campaign, which included the replacement of its electrodes. The new cathode is a hexagonal etched mesh, that received in-depth quality assurance prior to installation, which was presented in [chapter 4](#). My contribution to the quality assurance was the development of computer vision (CV) and ML methods for the optical identification of anomalous regions, *defects*, on the cathode mesh.

The first step towards this was the construction of a pixel segmentation algorithm, capable of distinguishing between mesh pixels and background pixels in images of the cathode. Two ML models, a random forest classifier (RFC) and an MLP, were trained on intensity, gradient and texture features, with the RFC performing better at the segmentation task than the neural network. Using the segmentation RFC, image sections of the individual sides of the hexagonal cells that make up the mesh, the *legs*, were cut from bigger, zoomed-out images.

The main challenge to overcome in the automatic defect detection is the fact that defects are extremely rare and apriori unknown, which makes it impossible to train a detection model in a supervised fashion. Rather, a variational autoencoder (VAE) model was trained in an unsupervised fashion, with the goal of learning the pixel distribution of a normal, i.e. non-anomalous, leg. A VAE consists of two models working in accord, an encoder model and a decoder model. The former encodes the input into a *latent space*, with a greatly reduced dimensionality, and therefore expressiveness. The decoder receives this latent encoding and attempts to decipher it, inflating it back into the original size, trying to mimic the input. The idea behind the VAE based anomaly detection is that defects are occurring too rarely for the model to learn how to encode and decode them to and from the latent space. Therefore, a non-anomalous leg will be reconstructed nearly perfectly, while a defect will appear as a clear difference between input and output.

Different VAE architectures were trained, incorporating different CV preprocessing pipelines, different network layers such as skip-connections, or adaption of the loss function, for example via annealing. The best-working mode, based on the area under the curve (AUC) of the receiver operator characteristic (ROC), was the baseline-model, including only a standard-scaling preprocessing and convolutional and pooling layers in alternation. A final number of 18 anomalous legs were identified in the investigated mesh and repaired where deemed necessary. The repair was conducted either via grinding and polishing away excess materials, or surgically cutting off the defective area and laser welding in healthy sections from a spare mesh. The CV and ML methods implemented for the defect detection will be used in future projects towards the development of large-scale electrodes, such as for the future XLZD experiment.

The optical inspection can identify potential defects, but does not provide insight into whether or not those defects are actually detrimental to the performance of the electrode. To test this, a high voltage (HV) test stand was set up at KIT, to which I contributed image analysis software. In the final stage of the HV test, there were no sparks or electrical breakdowns observed on the actual cathode mesh.

6.3 The Search for Bosonic Dark Matter with XENONnT

The final project conducted in the course of this thesis was the projection of the sensitivity of XENONnT in regard to absorption signals induced by bosonic DM, for example ALPs or DPs, using SR0 and SR1 data, that is described in [chapter 5](#). The main tasks towards this goal were the development and validation of detector response, signal acceptance and background models in an extended energy region of interest (ROI) from 1 keV up to 1000 keV. Many individual pieces of the analysis, such as the energy resolution and bias, were adapted from previous XENONnT studies, and are not further described in this summary.

The event building efficiency (EBE) quantizes the probability to correctly pair recorded S1 and S2 signals originating from the same physical energy deposit into events. For low energy studies, the EBE is typically derived via large-scale salting studies, where S1-S2 pairs are injected into the regular data stream, and it is analyzed how many of the injected pairs get reconstructed accurately. However, for high energy studies, and studies spanning a wide energy range, a salting study is time consuming and not always feasible. For this reason, I developed the *peak shadow* approximation, that can estimate the EBE in a quick manner for large energy ROIs. The peak shadow approximation works by counting S1 or S2 peaks with a larger area than some target value, and deriving the probability of such a larger signal happening within the event building time window. Within the energy range where there exists a salting EBE study, the peak shadow approximation was validated, and it agreed well within uncertainty with the salting method.

All cuts and selections were either adapted as they existed from other studies or received minor adjustments to make them applicable in the ROI of this work. The ER band selection used in this thesis was a combination of the one made for the solar neutrino search, and the one used in high energy analyses. Slight adjustments to the

were made, to ensure that the ER band boundaries are smooth. The S2 single-scatter (SS) selection was taken from the high energy ER study. The boundary of the S2 width cut was slightly loosened, to enable the validation on simulated data at high energies. The remaining data quality cuts were taken from the solar neutrino analysis. For the validation of the cuts and the computation of the acceptance, a combination of ^{220}Rn and ^{232}Th data was used, in addition to signal-like full-chain simulations. The cut acceptance followed the same computational framework as that used in the solar neutrino search, and the cumulative acceptance was smoothed with a Gaussian kernel filter.

The background originating from radiation in the detector materials has a complex spatial dependence, due to the propagation distances in the TPC of different particles of various energies. For this reason, large-scale material background simulations were conducted, based on activities derived during a radioactivity screening. The MS spectrum of the simulations was compared to the MS spectrum seen in data, and minor inconsistencies were corrected by scaling the contributions of the responsible isotopes in the simulation. The fiducial volume (FV) was optimized on this material background simulation. To ensure a spatially smooth spectrum, even in the center of the TPC where only few events were available, the material background was fitted with an adaptive kernel density estimation (AKDE). Finally, for the uncertainty of the FV, the event position reconstruction, the LXe density and the uncertainty of the electric field were taken into account.

The decays of ^{212}Pb and ^{214}Pb pose an additional challenge, as they have the chance to decay into short-lived excited states, causing a β - γ cascade. The two events can either be reconstructed together as a SS event, or be rejected by the SS selection as a MS event. However, because of this event topology, the exact fraction surviving and the spectrum of Pb background events reconstructed as SS had to be determined via full-chain simulations.

Another background that received more attention in this work were the neutron-activated isotopes, that can be generated for example during AmBe calibrations. Those isotopes form short-lived states with half lives on the order of days. Their concentration inside the detector was derived from the rate development over time after the AmBe calibration runs. However, since many of these neutron-activated background events are rare events themselves, the constraint on their rate is accompanied by large uncertainties. To reduce their overall impact, SR1 was split into two different sections below and above 300 keV. The high energy section, where there are only minor contributions from neutron activated backgrounds, used all available science data, while for the low energy section, the data within 20 days after an AmBe calibration was not used in the analysis.

The decay of ^{214}Bi happens in a three-step cascade. In general, these decays are either removed by the SS selection, or if they are accidentally merged into a single event, they are outside the energy ROI of this study. However, there is a volume of uninstrumented LXe in the detector, beneath the cathode and in-between the PTFE paneling and the field cage. A ^{214}Bi decay in this *skin volume* can lose part of its signal, leading to the remaining event being energetically inside the ROI or being missed by

the MS removal. To determine the influence of the ^{214}Bi -skin background, another set of full-chain simulations was conducted, with the result that this background is largely negligible. All remaining background components were taken from other XENONnT analyses.

The above described detector response, signal and background models were implemented into a statistical model, built using the *alea* software, with the aim of performing a profile likelihood ratio (PLR) test. The sensitivity was computed from Monte-Carlo (MC) simulations, drawn from the signal and background models, under the asymptotic assumption of Wilk's theorem, which presented a conservative upper limit in this case. The derived projections predict world leading sensitivity at 90 % CL for the axio-electric coupling g_{ae} and the dark photon kinetic mixing κ in the energy range between 140 keV and 1000 keV.

The results in this thesis demonstrate how ML enhances signal reconstruction and background suppression, and how ML can be used to efficiently identify anomalies in large-scale electrodes. Furthermore, this work also underlines the world-class sensitivity of XENONnT towards physics BSM by potentially detecting new DM candidates such as axion-like particles or dark photons.

S2 Hitpattern Simulation using Fuse

This section details the *fuse* simulation procedure, that was used to simulate the hitpatterns for the training and validation of the position reconstruction models in [chapter 3](#). The data required for the ML based position reconstruction is a set of S2 top hitpatterns $\{\vec{h}\}$ with corresponding known positions $\{(x, y)\}$. For XENONnT, this dataset was simulated using the *fuse* software, which is capable of simulating every step of the signal generation and reconstruction of XENONnT. Since for the position reconstruction only S2 signals are required, the energy deposition and quanta generation steps were skipped in the simulation. Instead, a given number of electrons were started at a given position in the TPC, which were then propagated to the liquid interface, where they produced the S2 photons. Those photons were distributed to the individual PMTs according to a so-called *optical map*. Afterwards, the XENONnT signal reconstruction pipeline was applied to the simulated S2s, to attain realistic hitpatterns.

The first point to note is that the initial position of the electrons is not necessarily the position at which the S2 photons are emitted. The electrons simulated in *fuse* were drifted from the initial position, (x_0, y_0, z_0) , to the liquid interface according to the electric drift field, resulting in a position $(x_1, y_1, z_{\text{interface}})$. To ensure a uniform distribution of training data in the horizontal plane, which in practice means $x_0 = x_1$ and $y_0 = y_1$, the horizontal drift, caused by electric field, had to be avoided. This was achieved by setting the starting depth of the electrons as $z_0 = -0.1$ cm, where the non-uniformity of the electric field is no longer implemented in *fuse*.

Furthermore, the position reconstruction models are required to be accurate across a broad range of S2 top area values. To enable this, the training data was generated to be uniform in $A_{10} = \log_{10}(A_{\text{top}})$, the logarithm of the S2 top area, A_{top} . The mapping from A_{10} to the number of initial electrons was derived empirically, by simulating signals with different numbers of initial electrons and fitting the resulting A_{top} values. [Figure A.1](#) shows this empirical relation, and the resulting fit function. Using the inverse transformation technique, this mapping was used to generate simulated hitpatterns that are uniform in A_{10} . The distribution of the area of the S2 top hitpatterns that were used for the training of the position reconstruction models is shown in [Fig-](#)

ure A.2. At approximately 10 PE there is a minor peak, which is the approximate S2 top area of a single electron S2.

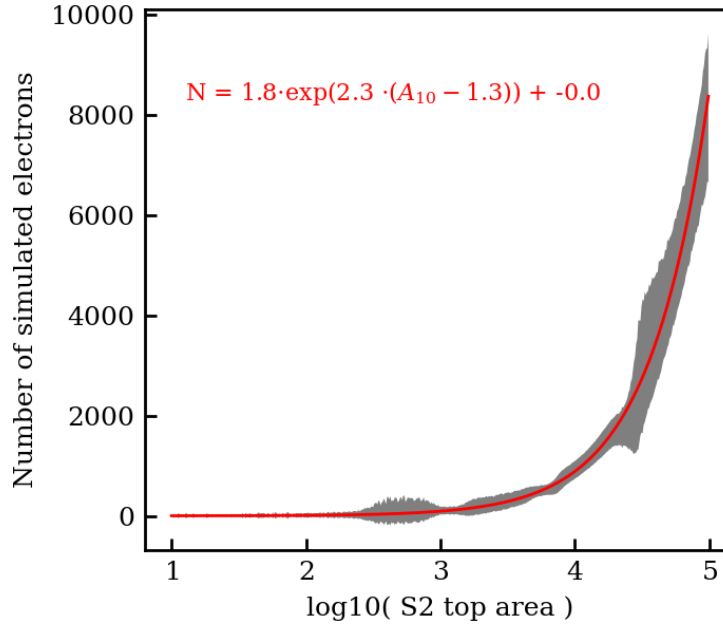


Figure A.1: The empirical mapping between the number of simulated electrons and the logarithm of the S2 top area.

Finally, the optical map plays an important role in the simulation of the hitpatterns, and therefore also in the training of the position reconstruction models. In parallel to the development of the CNF model, detailed in this thesis, there was XENONnT internal work on the improvement of the optical map [229]. The optical map employed during SR0 and SR1 did not properly take into account the generation of the S2 photons in dependence of the electric field generated by the anode wires. This led to a pattern coined the *anode shadow*, which was visible in data that should be, in theory, spatially uniform, such as ^{83m}Kr calibration data. The position distribution of ^{83m}Kr calibration data is shown in Figure A.3, derived by two different CNF models. The old model on the left was trained on S2 hitpatterns that were simulated using the SR0 and SR1 optical map. The anode shadow is the stripe-like pattern of higher event density parallel to the anode wires. The distribution on the right is provided by a model trained with hitpatterns that were simulated using the updated SR2 optical map. While there is still some pattern visible, it is far less pronounced than before.

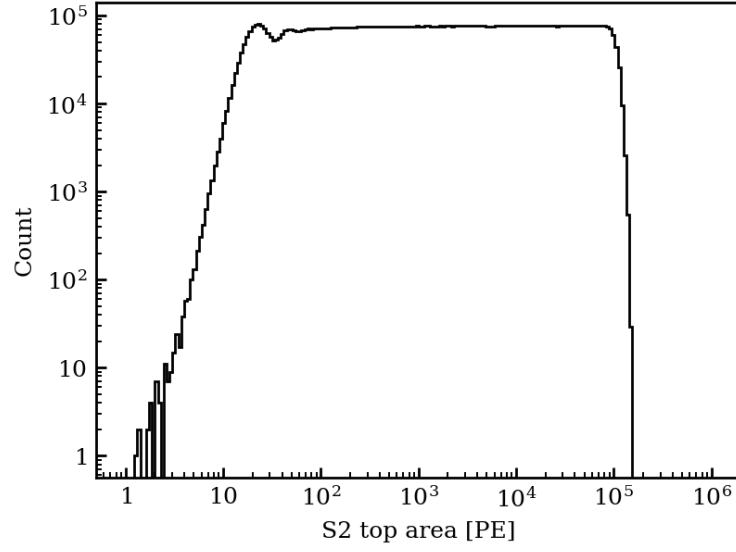


Figure A.2: The S2 top area distribution of the simulated hitpatterns used to train the position reconstruction models. The peak at approximately 10 PE is the average S2 top area of a single electron signal.

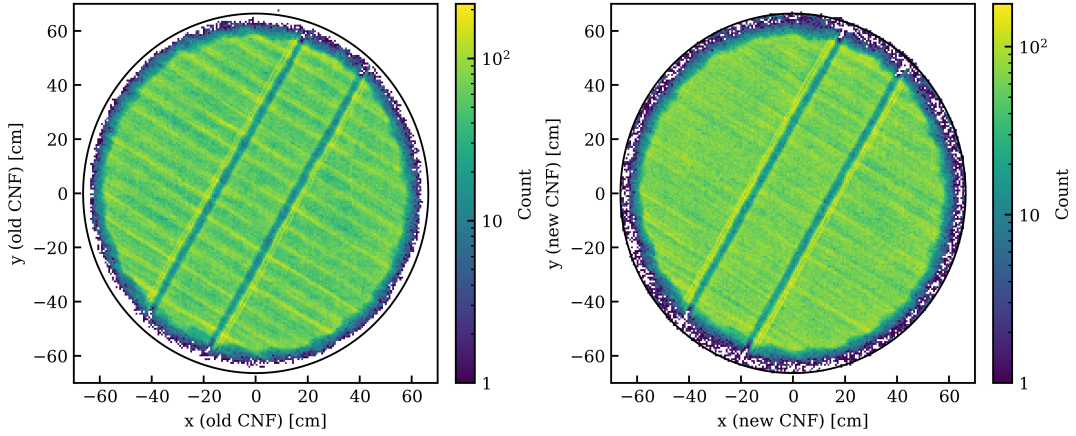


Figure A.3: The predicted position distributions of S2 signal from ^{83m}Kr calibration runs, using a position reconstruction CNF model that was trained using either the old optical map (left) or the new optical map (right).

Regular Hexagonal Mesh Algorithm

A regular hexagon can be defined by either one of the small diagonal d which is the distance between two opposite flat edges, the large diagonal D which is the distance between two opposite corners, or the length of one of the sides a . These measurements are related via $a = D/2 = d/\sqrt{3}$. The following constructs a hexagonal mesh, where D is vertical, that fills the positive quadrant of the horizontal plane. This is done by defining *anchor points* $\{x_i, y_i\}$, the points where three hexagonal edges of neighboring cells meet. At a given y , there is an anchor point every d distance, meaning $x_j = j \cdot \sqrt{3}a$ for j is either $[0, 1, \dots]$ or $[\frac{1}{2}, \frac{3}{2}, \dots]$ depending on y . For a given x , the distances between the y_j coordinates are a and D alternately.

For a physical mesh with a given a pitch p_{hex} and wire width w_{hex} , the small diagonal is $d = p_{\text{hex}} + w_{\text{hex}}$. Given the coordinates of the anchor points in units of pixels, the hexagonal pattern is produced by drawing lines between each anchor point and its 3 nearest neighbors, and the line width is reproduced by using a square dilution kernel of size w_{hex} .

In addition, the orientation of a mesh on a picture can be defined by offsets dx, dy in horizontal and vertical directions respectively, and by a rotation angle α from the horizontal line. Using these five parameters, the image of any hexagonal mesh can be reproduced as a binary mask. The full algorithm is described in [Appendix B](#).

The pixel classification RFC provides a binary mask of wire pixels for each base image. The parameters $p_{\text{hex}}, w_{\text{hex}}, dx, dy$ and α are optimized for each base image to achieve an optimal overlay of the algorithm created hexagonal mesh and the binary mask of wire pixels.

```

1
2 def get_mesh(pitch, wire_thickness, dx, dy, alpha,
3             shape, color):
4     '''
5     Generates anchors (x-y positons of the nodes of the mesh)

```

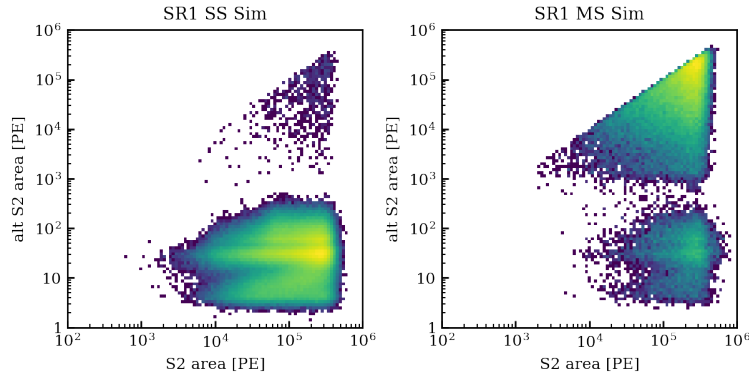
```
57                                     (cc<shape[1]))
58         valid_rows = rr[valid_ids]
59         valid_cols = cc[valid_ids]
60         img[valid_rows,valid_cols] = color
61
62         drawn_lines.append(pair)
63
64     dilution_kernel = cv2.getStructuringElement(cv2.MORPH_RECT,
65         (wire_thickness, wire_thickness)
66     )
67     img = cv2.dilate(img, dilution_kernel)
68     return img, anchors_in
69
```

Additional Figures on the Data Selection Cuts

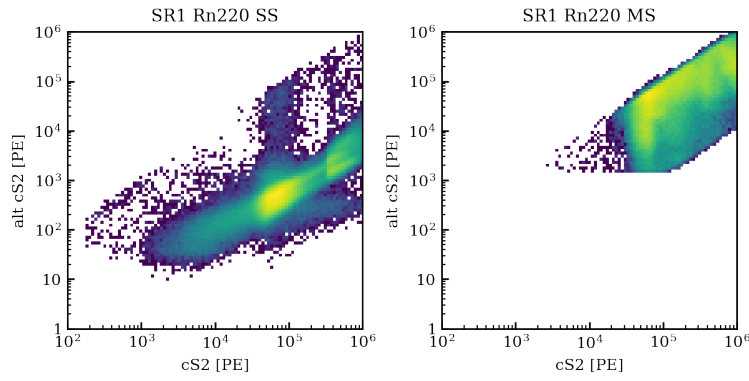
This chapter lists additional figures used in the derivation of the cut acceptance in [section 5.2](#).

C.1 S2 Single Scatter

The S2 SS selection is defined in the parameter space of alternative and main S2 areas. [Figure C.1b](#) shows this parameter space for simulated data (top) and calibration data (bottom). The top panel shows the distributions of signals that were simulated as either SS or MS, without applying the SS selection. This was done to highlight the actual different distributions in the parameter space. The bottom panel shows the alternative and main S2 areas for ^{220}Rn calibration data that were *reconstructed* as either SS or MS by the S2 SS cut. Furthermore, [Figure C.2](#) shows the acceptance of the SR1 S2 SS cut, which is for practical purposes equal to one for all energies, radii and depths.



(a) The main and alternative S2 strength distribution for SR1 simulations. The left panel shows signal-like SS simulations. The right panel shows events that were explicitly simulated as MS.



(b) The main and alternative S2 strength distribution for SR1 ^{220}Rn calibration data, reconstructed as SS (left) and MS (right).

Figure C.1: The main and alternative S2 strength distributions of different datasets: ^{220}Rn calibration data *reconstructed* as SS or MS (top), and simulation data *simulated* as SS or MS (bottom)

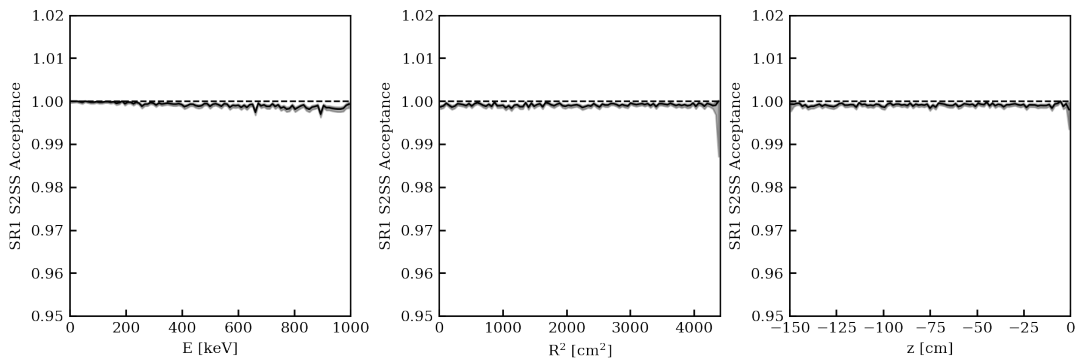


Figure C.2: The acceptance of the SS selection in SR1, determined on signal-like simulations as a function of energy (left), radius squared (center) and vertical position (right).

C.2 S1 Width

Figure C.3 shows the parameter space and cut boundaries of the S1 width cut for SR1, just as Figure 5.10 does for SR0. The cut acceptance of the S1 width cut is displayed in Figure C.4. Since the cut is only defined for S1 areas below 2000 PE, the cut acceptance becomes equal one at increasing energy. For low energies, there is a decrease in acceptance, in particular in SR0.

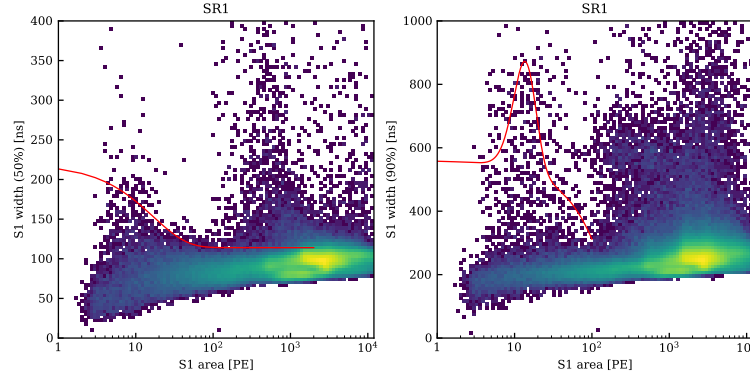


Figure C.3: The boundaries of the S1 width cut, shown over SR1 ^{220}Rn calibration data.

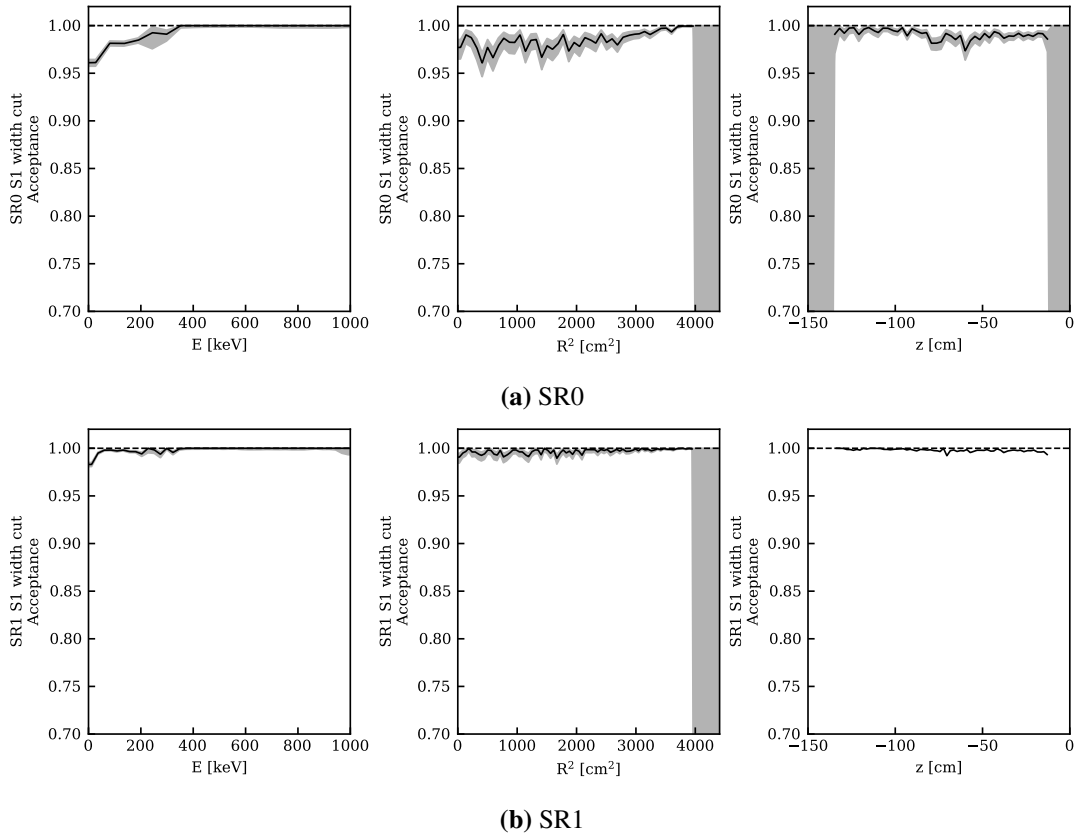


Figure C.4: The S1 width cut acceptance

C.3 S1 Area Fraction Top

The acceptance of the S1 AFT cut is shown in [Figure C.5](#). In SR0, the cut acceptance is flat in all dimensions. However, in SR1, there is a slight decrease in acceptance for increasing energy, and for lower values of z . One potential reason for this is the elevated PI rate of SR1, which affects the AFT by absorbing a fraction of the S1 photons during their propagation through the LXe. A PI-based effect would be especially pronounced for the lowest z values, since those events have the furthest travel distance for the S1 photons from initial event site to PMT array. Nevertheless, since the effect is on the same order as the statistical fluctuations of the acceptance, it does not necessitate additional in-depth evaluations of the SR1 cut.

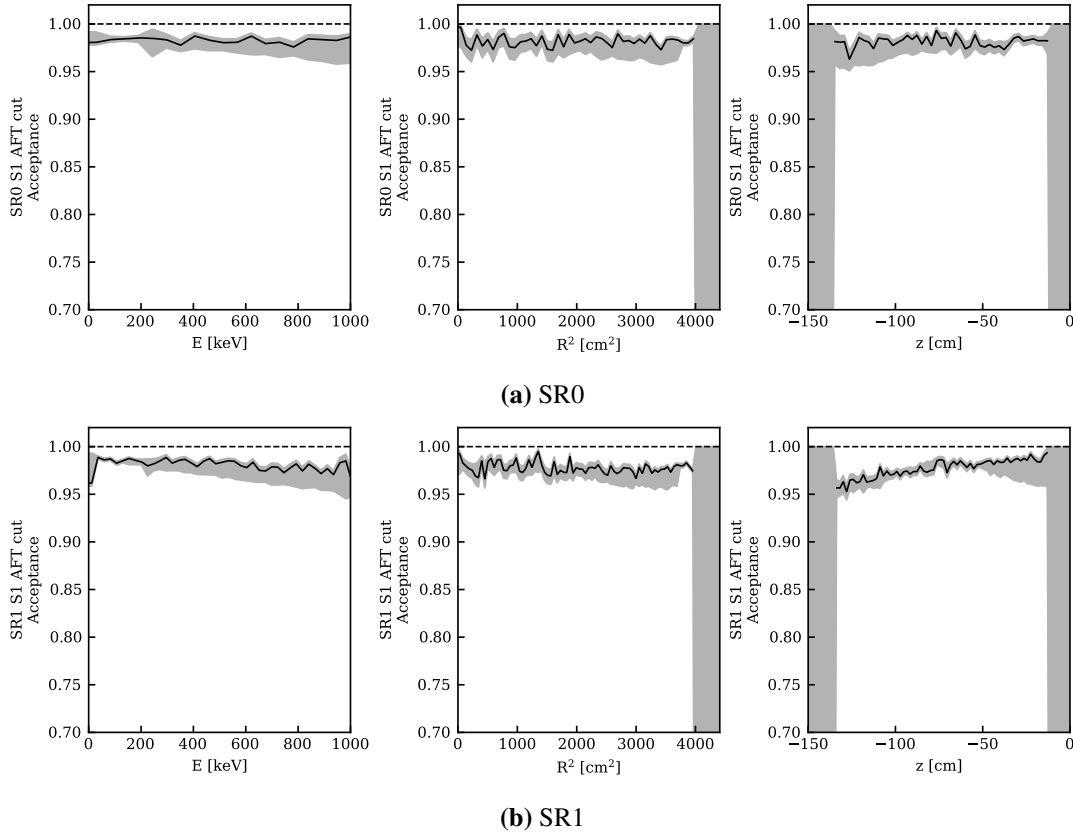


Figure C.5: The S1 AFT cut acceptance

C.4 S1 Max PMT

The S1 max PMT cut is meant to remove a very particular event population, with only a minimal amount of total events. For this reason, the cut acceptance, shown in [Figure C.6](#), is very nearly constant at 1 for all dimensions.

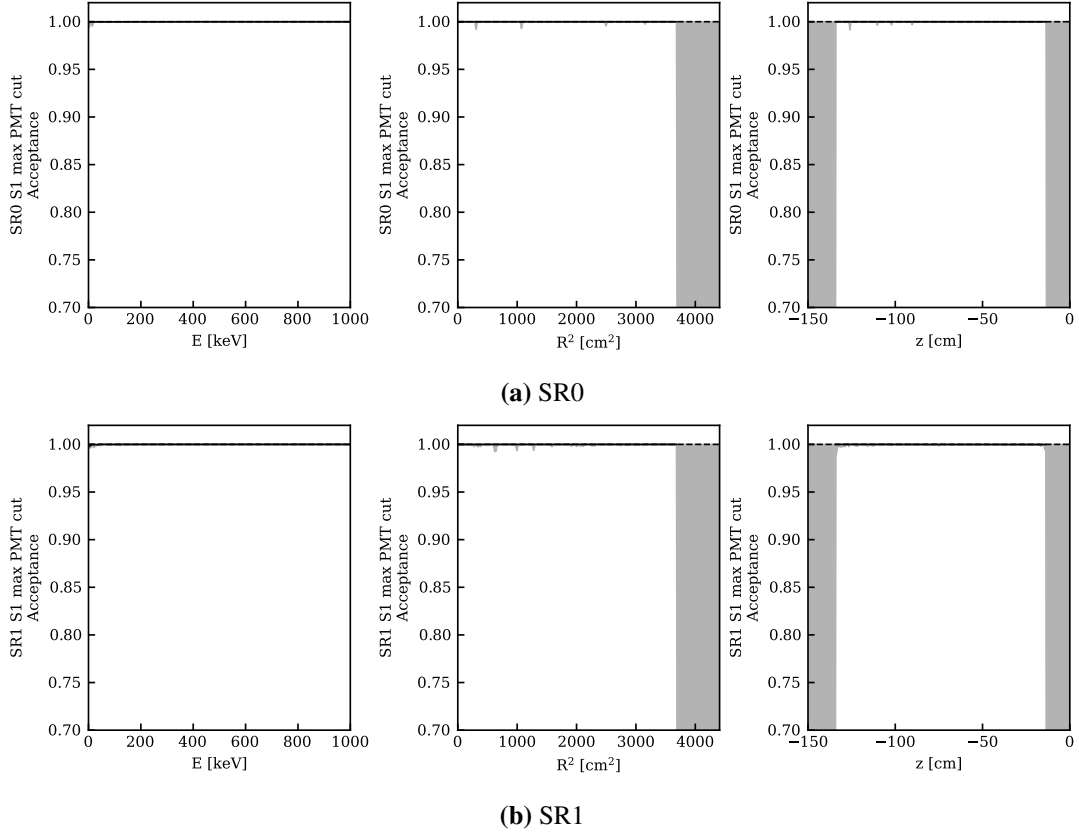


Figure C.6: The S1 max PMT cut acceptance

C.5 S2 corrected Area Fraction Top

The acceptance of the S2 AFT cut is shown in [Figure C.7](#). In both SR0 and SR1, there is a small decrease in cut acceptance visible at approximately 200 keV, and at the very largest radii. Both decreases are more pronounced in SR0 than in SR1. However, given the minimal extent of these decreases, it was decided that there is no requirement for an additional in-depth analysis of this cut.

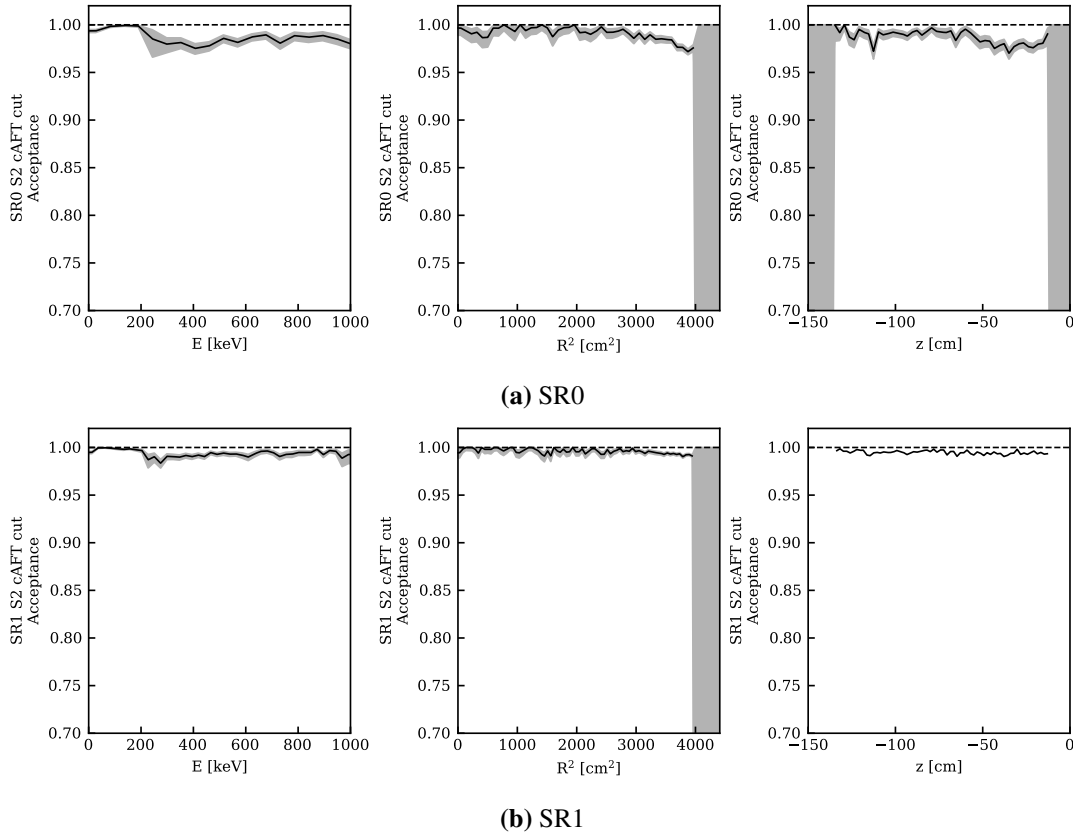


Figure C.7: The S2 AFT cut acceptance

C.6 S2 Pattern

The acceptance of the S2 pattern cut is depicted in [Figure C.8](#). There is an increase in acceptance for low energies leading up to approximately 200 keV, which is expected to some extent. The reason for this is that the S2 hitpattern is more significantly influenced by noise or merging with small uncorrelated signals for lower S2 areas, meaning lower energies.

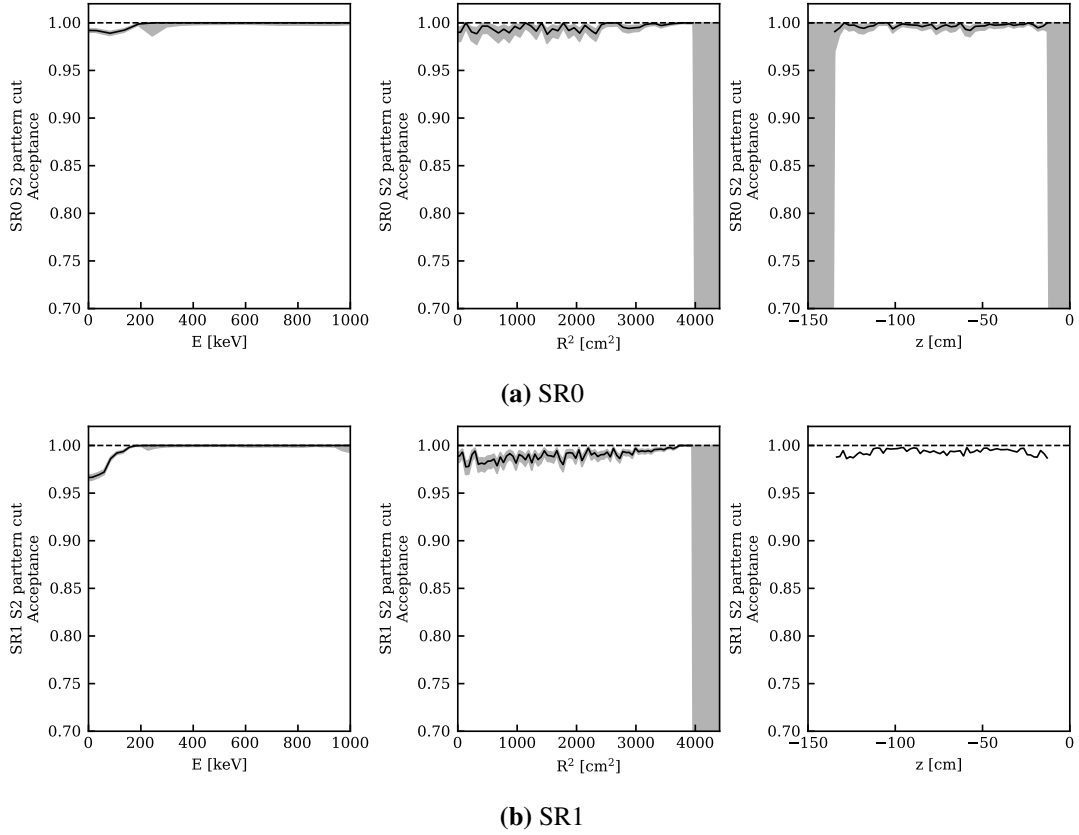


Figure C.8: The S2 pattern cut acceptance

C.7 S2 Width

The S2 width cut boundaries were adjusted for the bosonic DM search, as described in subsection 5.2.9. Figure C.9 shows the adjusted boundaries for the SR1 parameter space, and Figure C.10 shows the SR1 cut acceptance. The cut acceptance of the previous implementation of the S2 width cut, which is the definition used in [3], is shown in Figure C.11. This is portrayed here to highlight the necessity of adjusting the S2 width cut for energies beyond 200 keV.

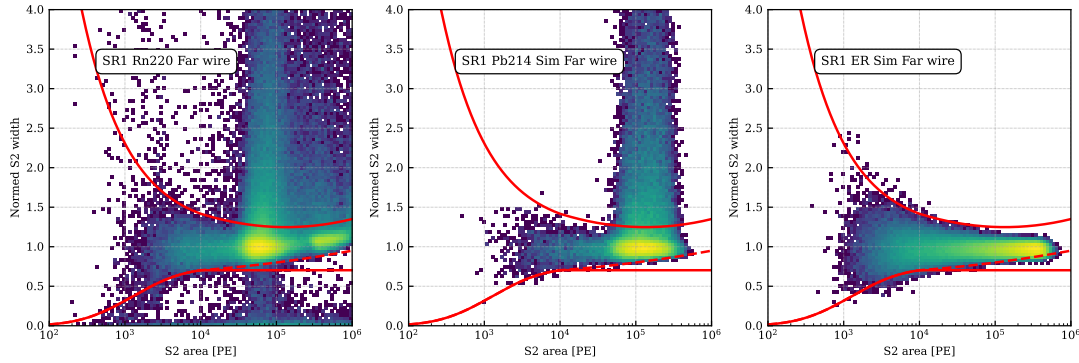


Figure C.9: The distribution of the normalized S2 width against the S2 area for SR1. The shown datasets are ^{220}Rn calibration data (left), ^{214}Pb simulation data (center) and the signal-like simulations (right). In red are the cut boundaries: The red dashed is the cut definition used in [3] and the red solid is the new cut boundary, adjusted for this study.

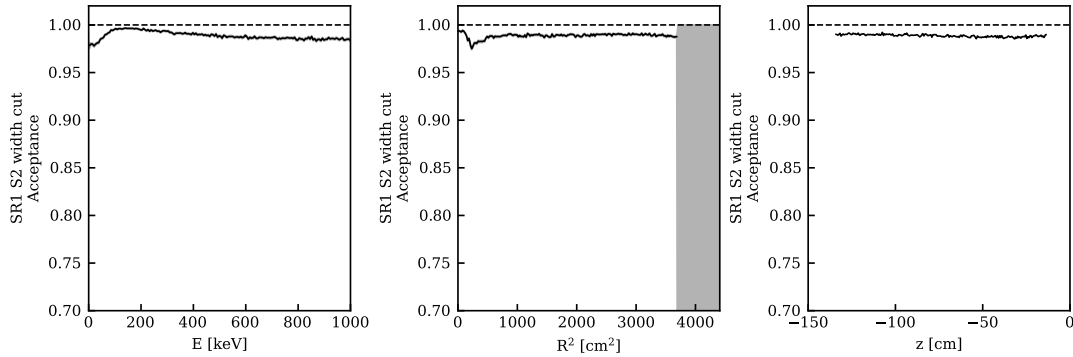


Figure C.10: The S2 width cut acceptance for SR1.

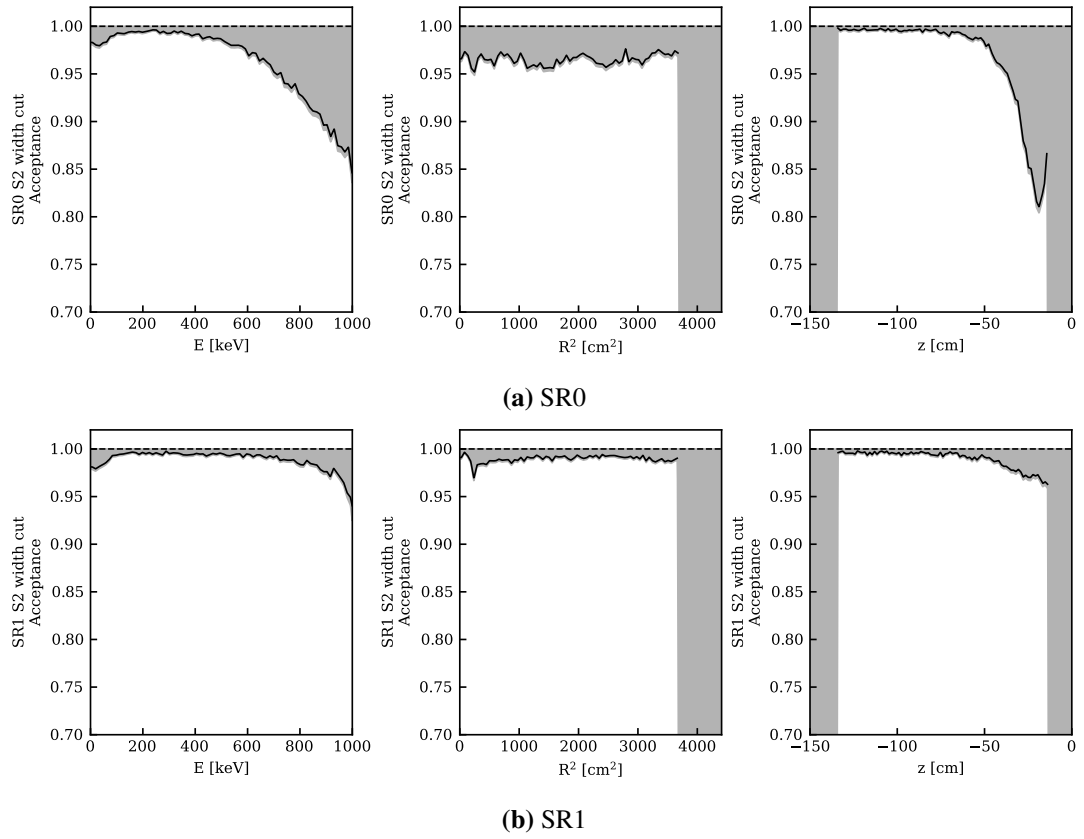


Figure C.11: The S2 width cut acceptance, using the cut definition from [3].

Parameter Comparison between ^{212}Pb Full Chain Simulation and ^{220}Rn Calibration Data

This appendix details the comparison between various parameter spaces between full-chain simulated ^{212}Pb events and ^{220}Rn calibration data, which consists majorly of ^{212}Pb decay events. The parameter comparisons relevant to the main work of this thesis are explained in [section 5.4](#). The remaining parameter spaces are compared here solely for completeness.

D.1 S1 and S2 areas

A comparison of the S1 and S2 areas is shown in [Figure D.1](#). The S1 area of SR0 matches well between simulated events and calibration data. For SR1, there is a minor shift in the simulated events towards lower S1 areas. For S2 areas, the simulations display a wider tail towards higher areas than is observed in calibration data. In addition, in SR1 there is again a visible shift, this time towards higher areas. The opposite shifts in S1 and S2 areas in SR1 can be explained by slightly mismodeled LY and CY in the simulation framework.

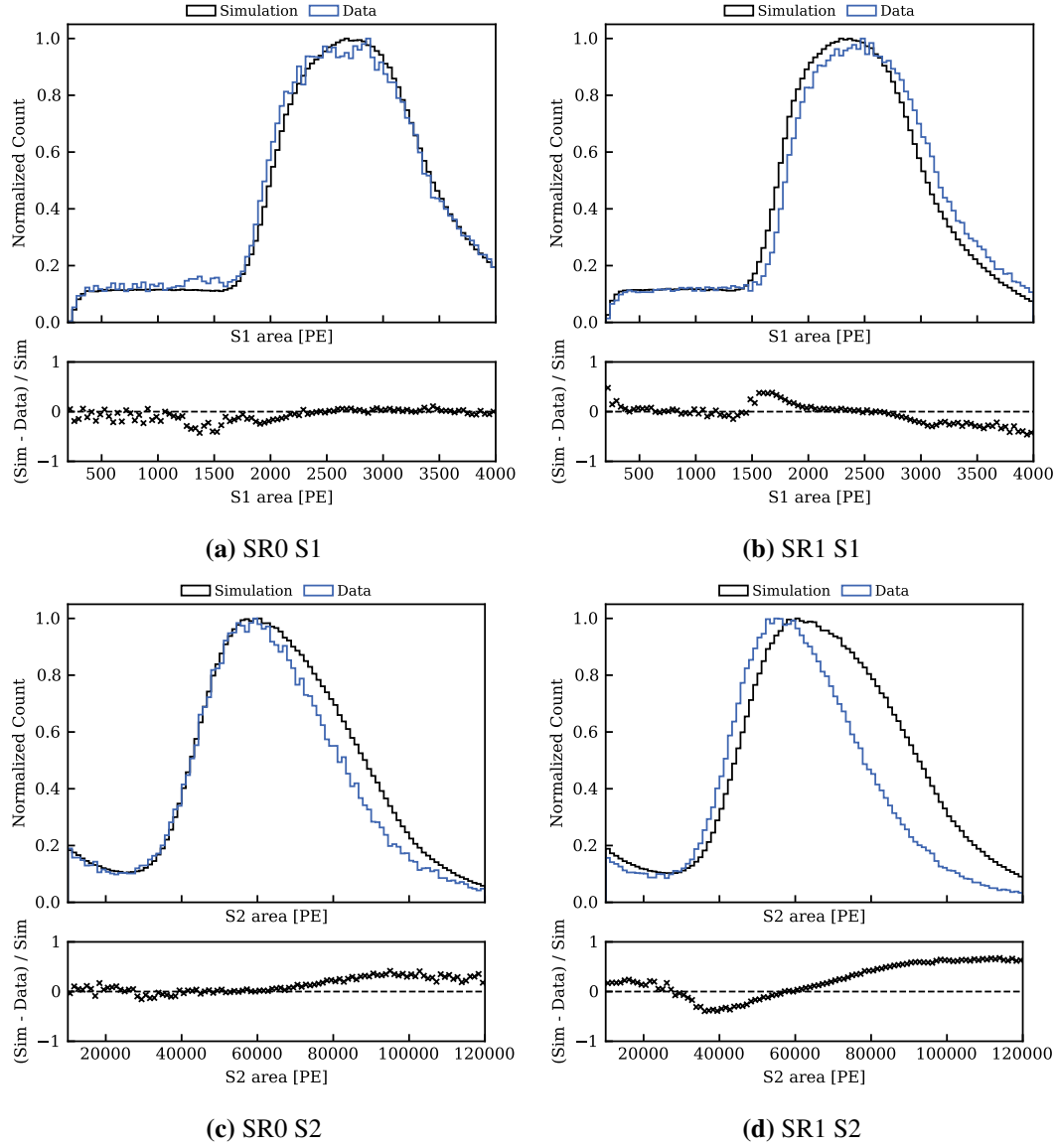


Figure D.1: A comparison of the S1 and S2 areas of the ^{212}Pb full chain simulation and ^{220}Rn calibration data.

D.2 S1 Parameters

The S1 AFT is shown in Figure D.2b, and matches well between data and simulation for both SR0 and SR1. The S1 width however, does not match, which is highlighted in Figure D.3. This points to significant mismodeling of the time distribution of S1 photons in the simulated events. For this reason, the S1 width cut can not be evaluated on simulated data.

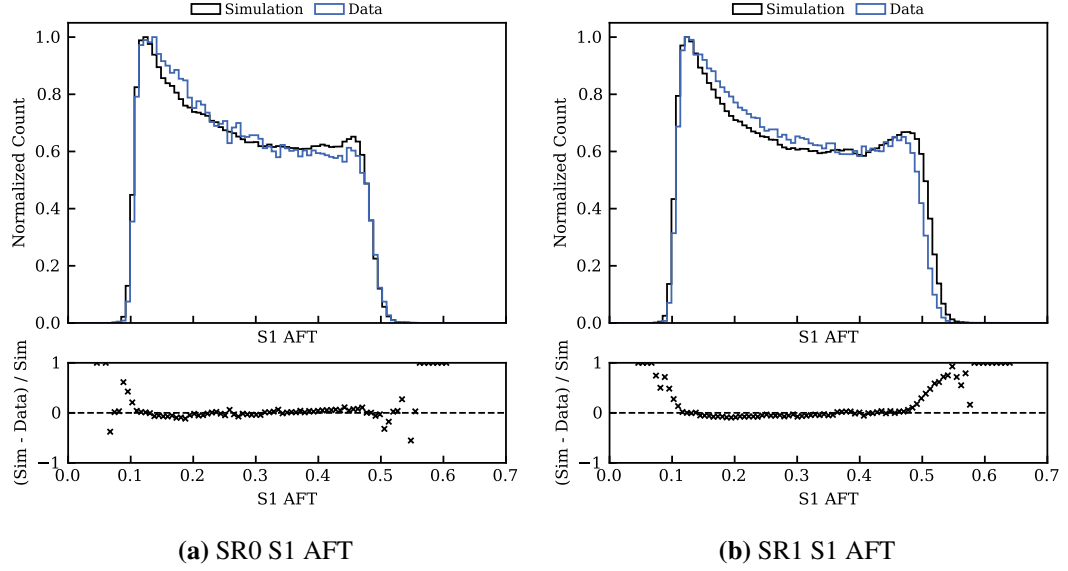


Figure D.2: A comparison of the S1 AFT of the ^{212}Pb full chain simulation and ^{220}Rn calibration data.

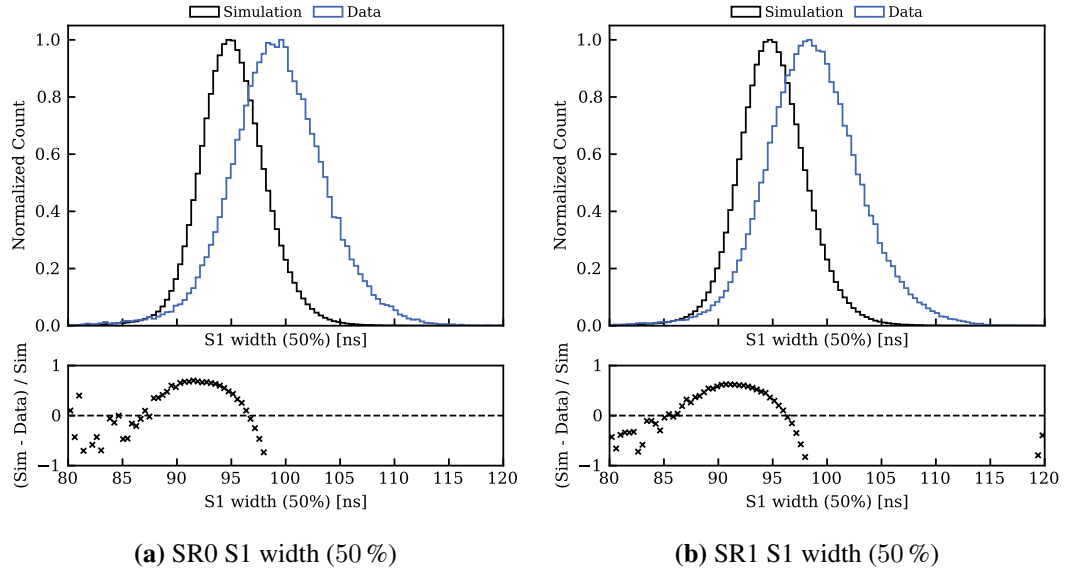


Figure D.3: A comparison of the S1 width (50 %) of the ^{212}Pb full chain simulation and ^{220}Rn calibration data.

D.3 S2 Parameters

The S2 AFT, shown in Figure D.4, shows minor discrepancies between simulation and calibration data, which is why the cS2 AFT cut was only evaluated on calibration data. The S2 width distribution is shown in Figure D.5. Up to approximately ~ 8000 ns, the width in simulations is modeled very well. However, the high width tail, which originates from events below the perpendicular wires, shows a slight discrepancy between simulations and calibration data. However, the S2 width of ^{212}Pb events is not representative of that of SS events, because of the β - γ topology of the decay. For this reason, and because the width matches for the vast majority of events, it was decided to evaluate the acceptance on simulated data.

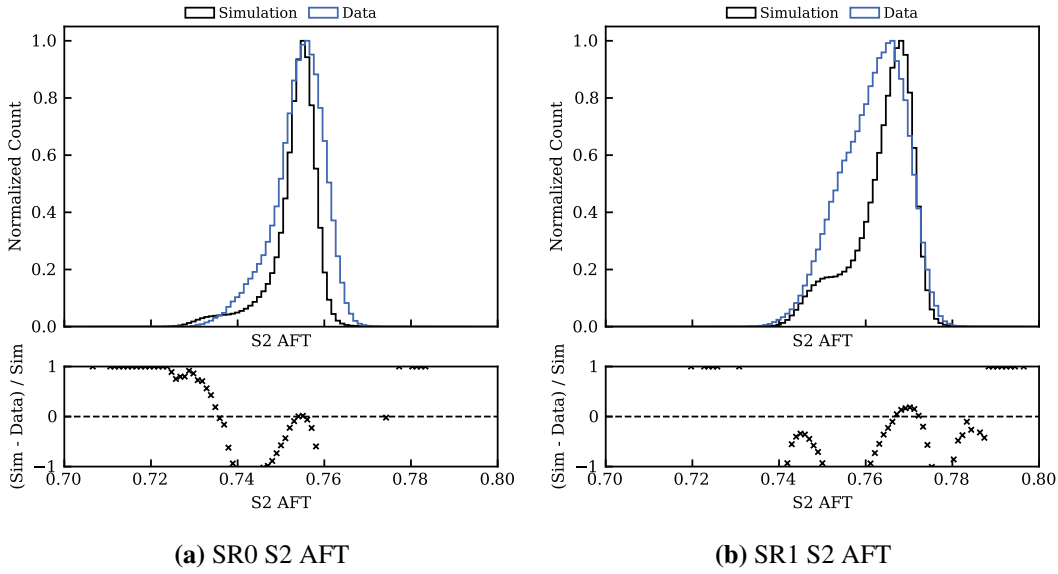


Figure D.4: A comparison of the S2 AFT of the ^{212}Pb full chain simulation and ^{220}Rn calibration data.

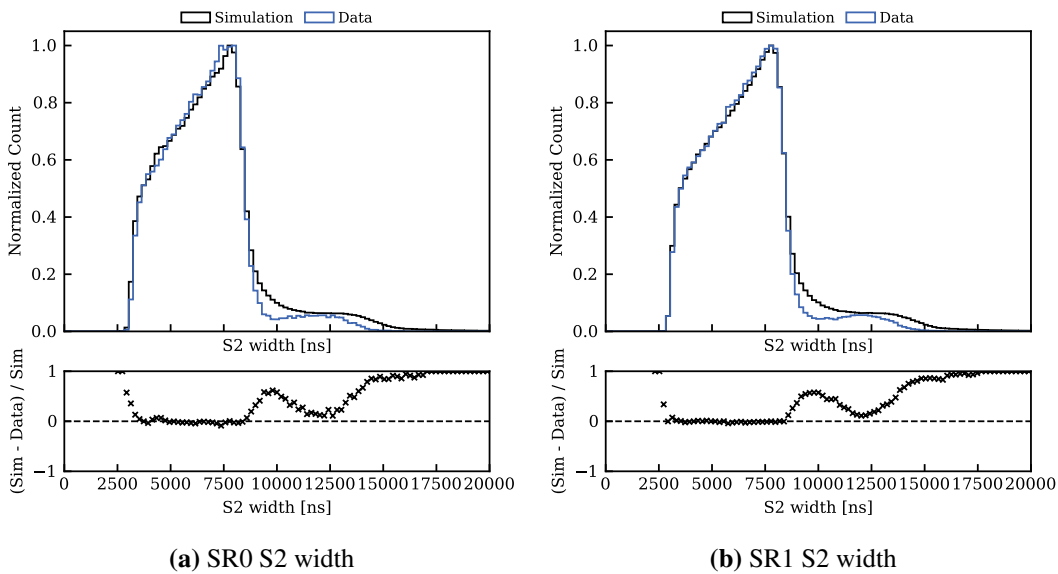
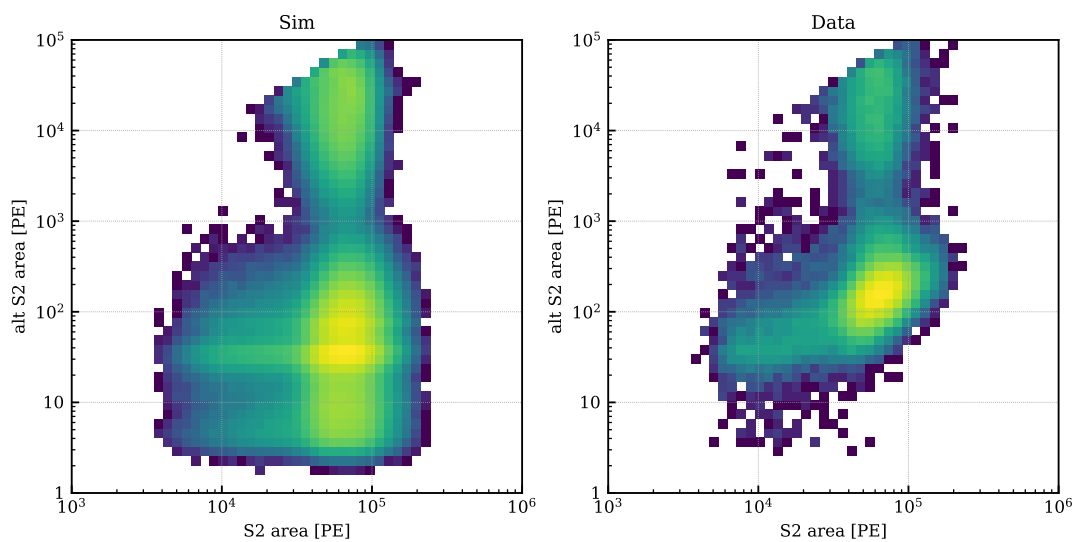


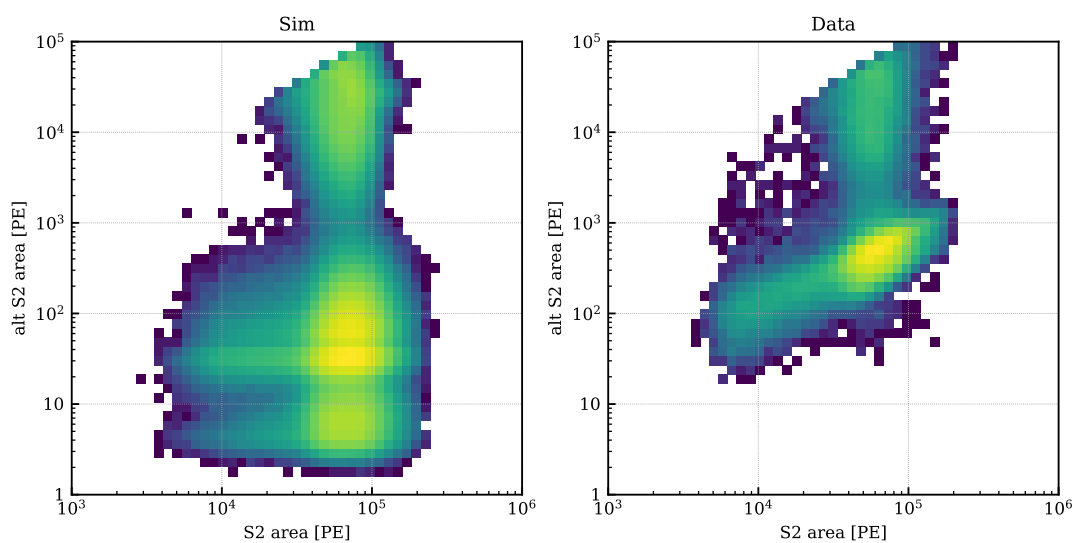
Figure D.5: A comparison of the S2 width of the ^{212}Pb full chain simulation and ^{220}Rn calibration data.

D.4 Alternative S2s and SS Selection

One of the most important reasons for generating full-chain simulations is the estimation of the SS-fraction f of the lead-originating decays. The SS-fraction characterizes the ratio of lead-originating events that are reconstructed as SS, instead of MS. To validate that this ratio can be estimated from simulated events, the SS selection parameter space, the distribution of alternative and main S2 areas, are shown in [Figure D.6](#). For both science runs, the simulated distribution of alternative S2 area extends much further down towards lower values. In addition, in the calibration data there is a well visible positive correlation between the main S2 area and the alternative S2 area. This correlation seems to be missing from the simulated events. However, as demonstrated in [Figure 5.27](#), the overall ratio between the number of SS and MS events does agree to a certain extent. The mismatch between the simulation derived and data derived values for f is included as an additional uncertainty in the lead-originating background rates, as described in [subsection 5.4.4](#).



(a) SR0 alternative S2 and S2 area distribution.



(b) SR1 alternative S2 and S2 area distribution.

Figure D.6: A comparison of the distributions of alternative S2 and main S2 areas of the ^{212}Pb full chain simulation and ^{220}Rn calibration data.

Fits of the Rates of Neutron Activated Isotopes

Below are the fits of the activities of the remaining neutron activated isotopes, described in [section 5.5](#). For each isotopes, the count of events, separated into bins of 12 h (SR0) or 24 h (SR1) is shown as black dots. The uncertainty in rate is derived from the Poissonian uncertainty of the count, and the uncertainty in time is the weighted standard deviation of the center times of all runs included in the respective bin. The blue dashed lines show the end of the AmBe calibrations. The red line shows the best-fit decay function, and the ed shaded band shows the 16 % and 84 % quantiles of fits derived via bootstrapping. In SR0, the AmBe calibration was followed immediately by a ^{83m}Kr calibration, which greatly limited the amount of available data. For this reason, all SR0 fits show a great deviation from literature values, and should be used carefully to constrain the expected rates of the neutron activated backgrounds.

The rate of ^{133}Xe events is shown in [Figure E.1](#) for SR0 and [Figure E.2](#) for SR1. Since the decay of ^{133}Xe is a β -decay, the resulting energy spectrum is continuous. The selection range to count the ^{133}Xe events was chosen as $[80, 140]\text{keV}$, to avoid overlap with other time dependent backgrounds. The SR0 fit does not model the data very well, which is why the upper limit value is taken to derive the nominal expectation of ^{133}Xe events in SR0. The difference between upper limit and best-fit is added to the uncertainty of the expected rate.

[Figure E.3](#) and [Figure E.4](#) depict the rate of ^{131m}Xe events for SR0 and SR1 respectively. To select ^{131m}Xe events, a 2σ energy interval around the literature value of the energy was used, as is explained in [section 5.5](#). Once more, the SR0 fit is off from the underlying data. Similar to the ^{133}Xe case, the upper limit value is taken to derive the nominal expectation of events, and the difference between upper limit and best-fit is propagated as additional uncertainty. The exact same logic applies to the ^{129m}Xe rates, visualized in [Figure E.5](#) and [Figure E.6](#).

The energy range to select ^{127}Xe events is the 2σ interval around the peak at 408.2 keV,

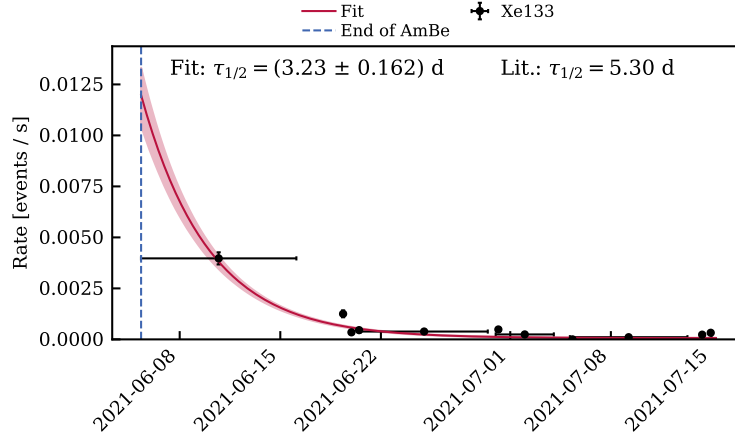


Figure E.1: The rate of ^{133}Xe events in SR0. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

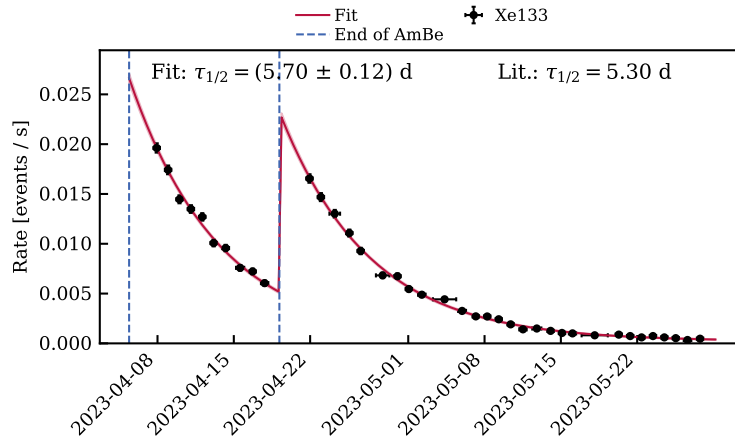


Figure E.2: The rate of ^{133}Xe events in SR1. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

which was chosen to avoid overlap with other time dependent backgrounds. The expected rate has to be corrected for the BR of 39.5 % of this decay branch. The ^{127}Xe fits, shown in [Figure E.7](#) and [Figure E.8](#), catch the overall trend of the data, but the derived half lives are far off from the literature values. To compute the expected rates of ^{127}Xe , the greater literature value for the half life is taken, resulting in a greater number of expected events.

Finally, [Figure E.9](#) and [Figure E.10](#) depict the rate of ^{125}Xe . The event selection for ^{125}Xe was performed in the 2σ energy interval around the peak at 276.6 keV, which has the greatest BR. However, because of the energy resolution, the event in this selection overlap slightly with those of the ^{129m}Xe decay, further worsening the data quality. To remain conservative, the best-fit value of the half-life was used to compute the expected number of events, which is greater than the literature value and therefore overestimates the total, rather than underestimate it.

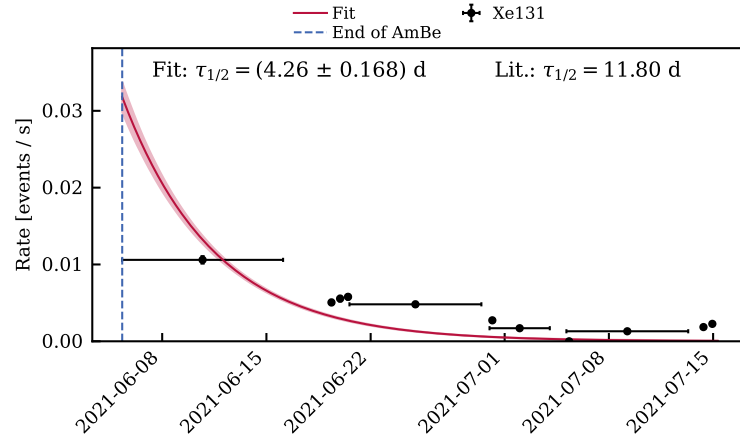


Figure E.3: The rate of ^{131m}Xe events in SR0. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

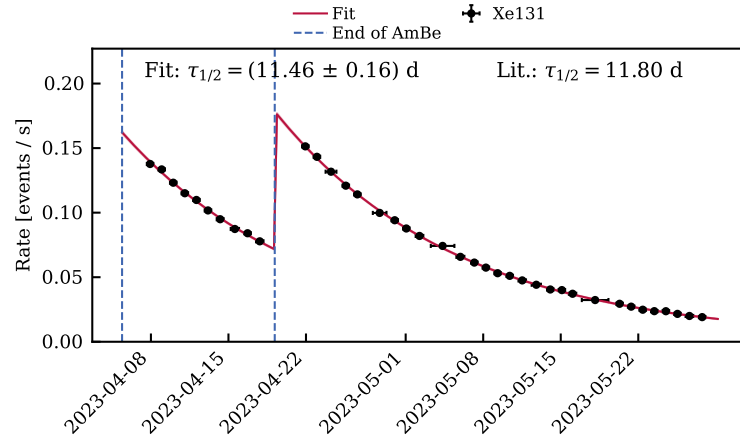


Figure E.4: The rate of ^{131m}Xe events in SR1. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

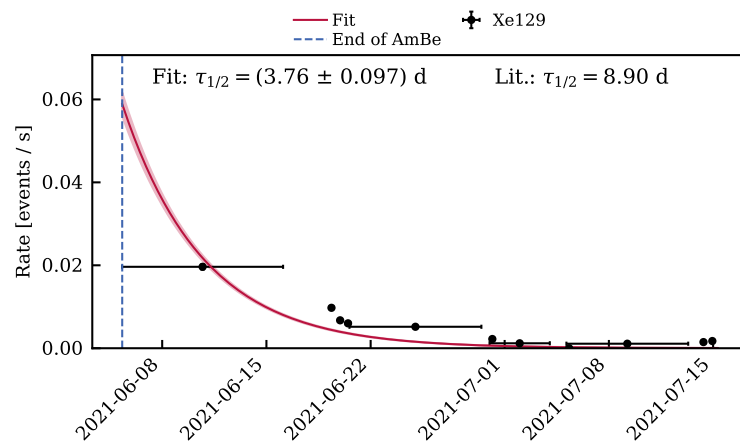


Figure E.5: The rate of ^{129m}Xe events in SR0. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

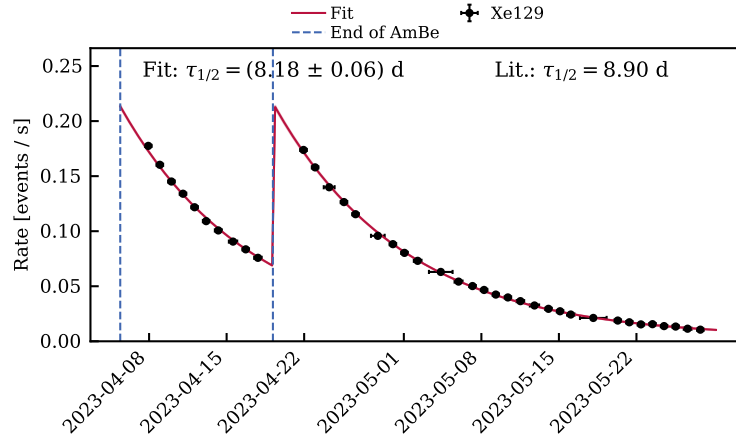


Figure E.6: The rate of ^{129m}Xe events in SR1. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

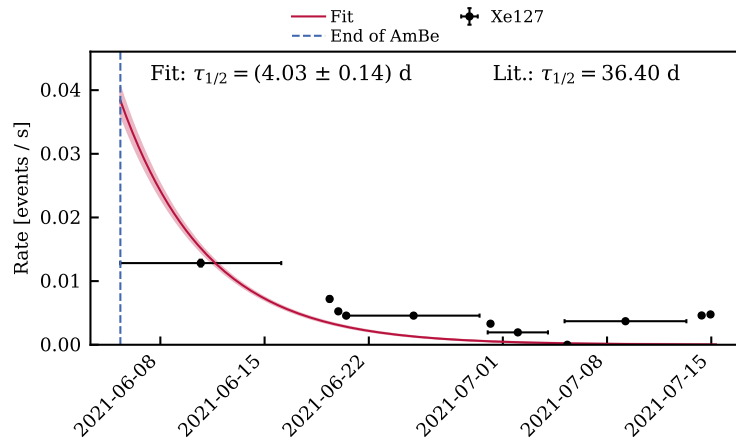


Figure E.7: The rate of ^{127}Xe events in SR0. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

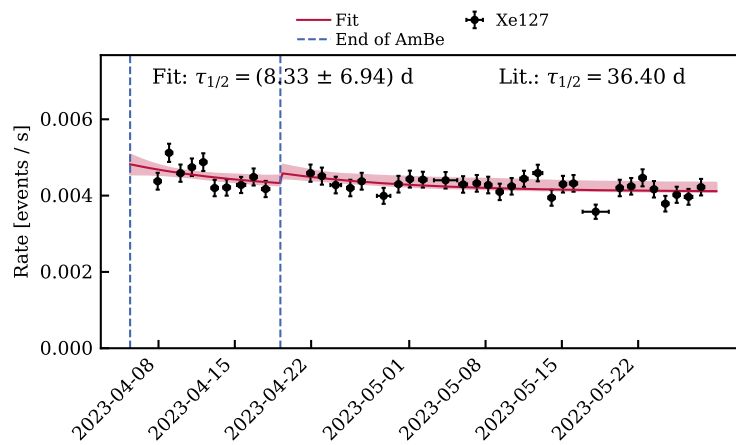


Figure E.8: The rate of ^{127}Xe events in SR1. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

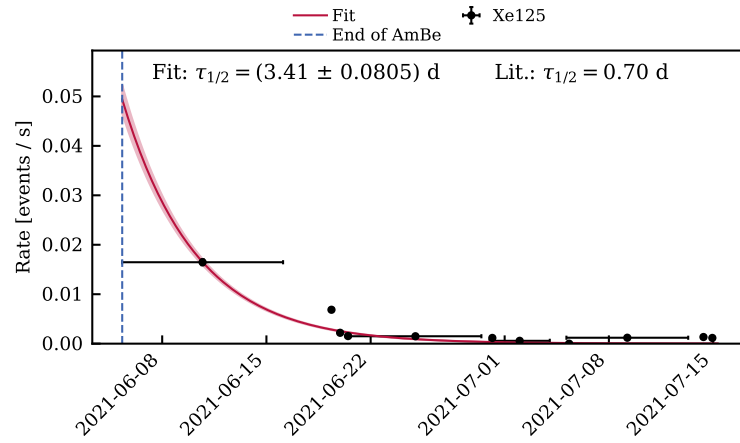


Figure E.9: The rate of ^{125}Xe events in SR0. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

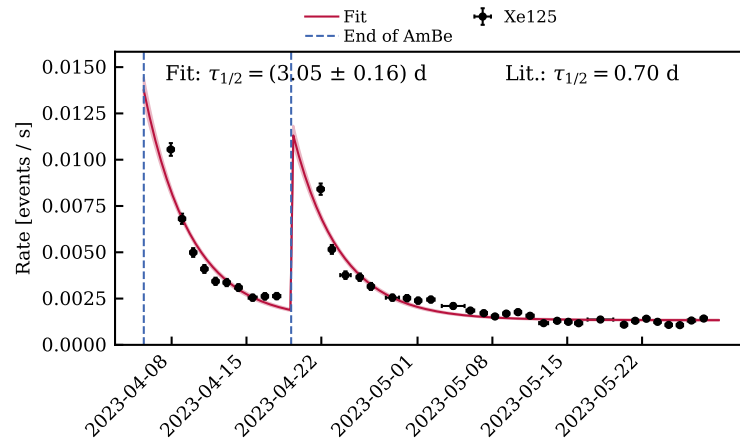


Figure E.10: The rate of ^{125}Xe events in SR1. The best-fit and uncertainty is shown in red. The blue line indicates the end of the AmBe calibration.

Derivation of the Critical Region

The sensitivity projection derived in [chapter 5](#) relies on the assumption, that the PLR test statistic $q(N)$ is near the asymptotic limit from Wilk's theorem [124], meaning that $q(N)$ follows a χ^2 distribution. In general, a hypothesis H_N is rejected at 90 % CL, if its test statistic, q_N , is greater than 90 % of the distribution of $q(N)$. Under the asymptotic assumption of Wilk's theorem, this is at $q_N > 2.71$, the 90 % threshold of the χ^2 distribution. However, the asymptotic assumption does not always hold. In this case, the 90 % threshold can be derived via toy MC data, by looking at the 90 % percentile of computed $q(N)$ values. The *critical region* is the region of test statistic values greater than this threshold. More details on how to identify non-asymptotic behavior in test statistics, and alleviate such situations, can be found for example in [230], on which the here shown analysis is based.

Instead of relying on assumptions about the distribution $q(N)$, a number of toy MC datasets were simulated using *alea*, that include N injected signals of a DM particle with mass m_a . One such example, using $m_a = 100 \text{ keV}$ and $N = 100 \text{ t}^{-1}\text{y}^{-1}$ is shown in [Figure F.1](#). In total, 12500 simulations were generated per mass and per signal strength.

For each simulation, the best-fit likelihood $\mathcal{L}_{\text{best}} = \mathcal{L}(x|\hat{\theta}, \hat{N})$ and the *true signal* likelihood $\mathcal{L}_{\text{true}} = \mathcal{L}(x|\theta_{\text{max}}, N = N)$ were computed. The distribution of the PLR, $q(N) = \mathcal{L}_{\text{true}}/\mathcal{L}_{\text{best}}$, of a number simulations provides insight into the threshold of the 90 % CL. A likelihood $\mathcal{L}_N$, that leads to a higher value of $q_N = \mathcal{L}_N/\mathcal{L}_{\text{best}}$ than the threshold of the $q(N)$ distribution, has a chance less than 10 % to represent the true signal strength N . The distribution of $q(N)$, for different simulated signal strengths N , is shown in [Figure F.2](#), with the vertical lines visualizing the 90 % percentile of the respective distribution.

The critical region is defined by the 90 % percentiles derived for each value of N , as shown in [Figure F.3](#) for DM particles of masses 10 keV, 100 keV and 500 keV. The critical region would replace the red line in [Figure 5.39](#). The χ^2 -threshold, shown in red, is in general greater than the MC derived 90 % threshold, which means that

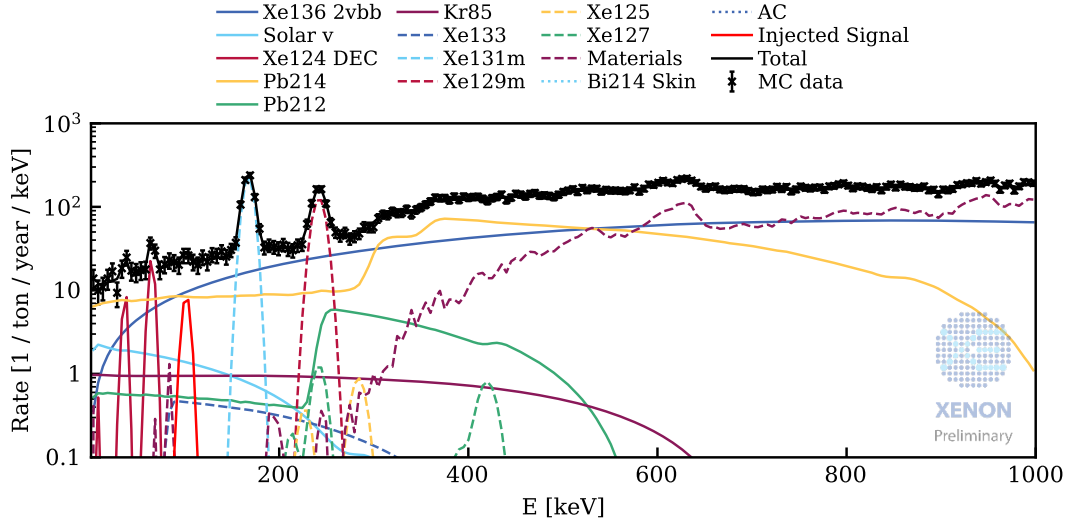


Figure F.1: SR0 toy MC data, sampled from a model including a signal from a DM particle with mass $m_a = 100$ keV and an expected rate of $N = 100 \text{ t}^{-1} \text{ y}^{-1}$.

using the asymptotic approximation *overestimates* the upper limit of the exclusion sensitivity. In this sense, the limits presented in [chapter 5](#) are conservative, and can be expected to be improved, when the derivation of the critical region is performed for every tested value of m_a .

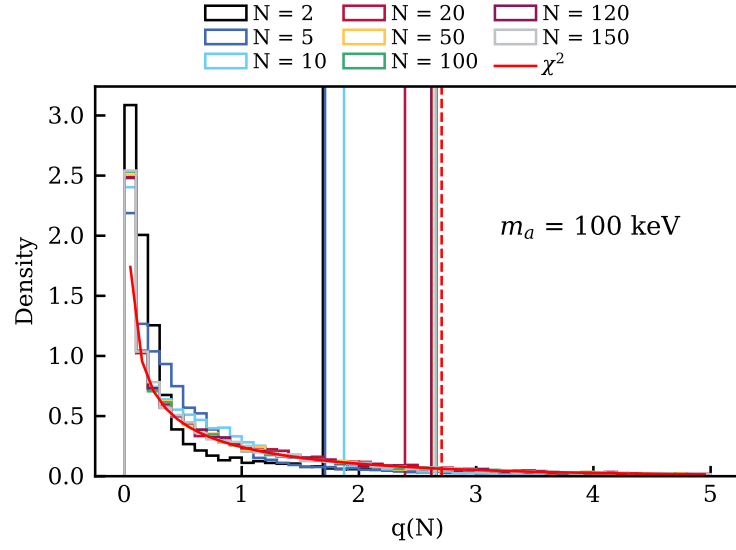


Figure F.2: The distribution of the test statistic $q(N)$, for different injected signal strengths N from the absorption of a DM particle of mass $m_a = 100$ keV. In red is the theoretical χ^2 distribution. The vertical lines visualize the 90 % percentile of the distributions.

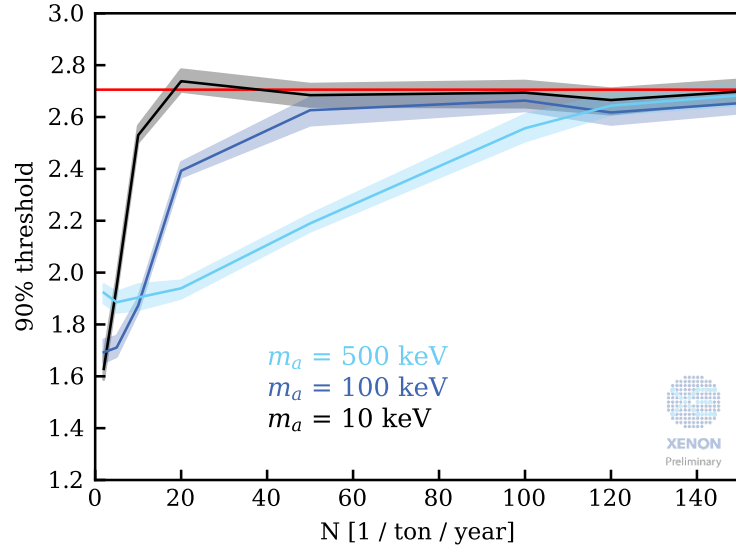


Figure F.3: The 90 % threshold as a function of signal strengths N from the absorption of a DM particle of different masses. The shaded regions indicate the interval covering a 68 % confidence interval of the 90 % threshold, derived via quantile test. The red line shows the 90 % threshold of a χ^2 distribution, which is taken in the asymptotic approximation.

Bibliography

- [1] J. Qin and S. Vetter. *Normalising Flows for Probabilistic Position Reconstruction in Dual-phase Noble Element Time Projection Chambers*. In preparation. 2026.
- [2] A. Elykov et al. “Development and characterization of electrodes for large-scale Xenon Time Projection Chambers”. In: *Journal of Instrumentation* 21.02 (Feb. 2026), P02035. DOI: [10.1088/1748-0221/21/02/p02035](https://doi.org/10.1088/1748-0221/21/02/p02035).
- [3] E. Aprile et al. *Measurement of the solar pp neutrino spectrum in XENONnT*. In preparation. 2026.
- [4] G. Box. *Empirical model-building and response surfaces*. New York: Wiley, 1987.
- [5] A. Einstein. “Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie”. In: vol. 47. 1915, pp. 78–87. DOI: [10.1002/3527608958.ch4](https://doi.org/10.1002/3527608958.ch4).
- [6] B. P. Abbott et al. “Observation of Gravitational Waves from a Binary Black Hole Merger”. In: *Physical Review Letters* 116.6 (Feb. 2016). DOI: [10.1103/physrevlett.116.061102](https://doi.org/10.1103/physrevlett.116.061102).
- [7] M. Schumann. “Direct detection of WIMP dark matter: concepts and status”. In: *Journal of Physics G: Nuclear and Particle Physics* 46.10 (Aug. 2019), p. 103003. DOI: [10.1088/1361-6471/ab2ea5](https://doi.org/10.1088/1361-6471/ab2ea5).
- [8] T. M. Undagoitia and L. Rauch. “Dark matter direct-detection experiments”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.1 (Dec. 2015), p. 013001. DOI: [10.1088/0954-3899/43/1/013001](https://doi.org/10.1088/0954-3899/43/1/013001).
- [9] G. Jungman, M. Kamionkowski, and K. Griest. “Supersymmetric dark matter”. In: *Physics Reports* 267.5-6 (Mar. 1996), pp. 195–373. DOI: [10.1016/0370-1573\(95\)00058-5](https://doi.org/10.1016/0370-1573(95)00058-5).
- [10] F. Wilczek. “Problem of Strong P and T Invariance in the Presence of Instantons”. In: *Phys. Rev. Lett.* 40 (5 Jan. 1978), pp. 279–282. DOI: [10.1103/PhysRevLett.40.279](https://doi.org/10.1103/PhysRevLett.40.279).
- [11] S. Weinberg. “A New Light Boson?” In: *Phys. Rev. Lett.* 40 (4 Jan. 1978), pp. 223–226. DOI: [10.1103/PhysRevLett.40.223](https://doi.org/10.1103/PhysRevLett.40.223).
- [12] B. Holdom. “Two $U(1)$ ’s and charge shifts”. In: *Physics Letters B* 166.2 (1986), pp. 196–198. DOI: [https://doi.org/10.1016/0370-2693\(86\)91377-8](https://doi.org/10.1016/0370-2693(86)91377-8).

- [13] M. Feickert and B. Nachman. *A Living Review of Machine Learning for Particle Physics*. 2021. URL: <https://arxiv.org/abs/2102.02770>.
- [14] F. Zwicky. “Die Rotverschiebung von extragalaktischen Nebeln”. In: *Helv. Phys. Acta* 6 (1933), pp. 110–127. DOI: [10.1007/s10714-008-0707-4](https://doi.org/10.1007/s10714-008-0707-4).
- [15] V. C. Rubin and W. K. Ford Jr. “Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions”. In: *Astrophys. J.* 159 (Feb. 1970), p. 379. DOI: [10.1086/150317](https://doi.org/10.1086/150317).
- [16] F. Lelli, S. S. McGaugh, and J. M. Schombert. “SPARC: MASS MODELS FOR 175 DISK GALAXIES WITH SPITZER PHOTOMETRY AND ACCURATE ROTATION CURVES”. In: *The Astronomical Journal* 152.6 (Nov. 2016), p. 157. DOI: [10.3847/0004-6256/152/6/157](https://doi.org/10.3847/0004-6256/152/6/157).
- [17] M. W. Werner et al. “The Spitzer Space Telescope Mission”. In: *The Astrophysical Journal Supplement Series* 154.1 (Sept. 2004), pp. 1–9. DOI: [10.1086/422992](https://doi.org/10.1086/422992).
- [18] J. F. Navarro, C. S. Frenk, and S. D. M. White. “A Universal Density Profile from Hierarchical Clustering”. In: *The Astrophysical Journal* 490.2 (Dec. 1997), pp. 493–508. DOI: [10.1086/304888](https://doi.org/10.1086/304888).
- [19] D. Clowe, A. Gonzalez, and M. Markevitch. “Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657–558: Direct Evidence for the Existence of Dark Matter*.”. In: *The Astrophysical Journal* 604.2 (Apr. 2004), p. 596. DOI: [10.1086/381970](https://doi.org/10.1086/381970).
- [20] M. Markevitch et al. “Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657–56”. In: *The Astrophysical Journal* 606.2 (May 2004), p. 819. DOI: [10.1086/383178](https://doi.org/10.1086/383178).
- [21] R. Durrer. “The cosmic microwave background: the history of its experimental investigation and its significance for cosmology”. In: *Classical and Quantum Gravity* 32.12 (June 2015), p. 124007. DOI: [10.1088/0264-9381/32/12/124007](https://doi.org/10.1088/0264-9381/32/12/124007).
- [22] N. Aghanim et al. “Planck2018 results: I. Overview and the cosmological legacy of Planck”. In: *Astronomy and Astrophysics* 641 (Sept. 2020), A1. DOI: [10.1051/0004-6361/201833880](https://doi.org/10.1051/0004-6361/201833880).
- [23] D. J. Fixsen. “THE TEMPERATURE OF THE COSMIC MICROWAVE BACKGROUND”. In: *The Astrophysical Journal* 707.2 (Nov. 2009), p. 916. DOI: [10.1088/0004-637X/707/2/916](https://doi.org/10.1088/0004-637X/707/2/916).
- [24] Planck Collaboration. *The 2018 Planck map of the temperature anisotropies of the CMB*. July 2018. URL: <https://www.cosmos.esa.int/web/planck/picture-gallery>.
- [25] W. Hu and S. Dodelson. “Cosmic Microwave Background Anisotropies”. In: *Annual Review of Astronomy and Astrophysics* 40.1 (Sept. 2002), pp. 171–216. DOI: [10.1146/annurev.astro.40.060401.093926](https://doi.org/10.1146/annurev.astro.40.060401.093926).
- [26] N. Aghanim et al. “Planck 2018 results: VI. Cosmological parameters”. In: *Astronomy and Astrophysics* 641 (Sept. 2020), A6. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910).
- [27] M. Kuhlen, M. Vogelsberger, and R. Angulo. “Numerical simulations of the dark universe: State of the art and the next decade”. In: *Physics of the Dark Universe* 1.1 (2012). Next Decade in Dark Matter and Dark Energy, pp. 50–93. DOI: <https://doi.org/10.1016/j.dark.2012.10.002>.

- [28] D. J. Eisenstein et al. “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies”. In: *The Astrophysical Journal* 633.2 (Nov. 2005), pp. 560–574. DOI: [10.1086/466512](https://doi.org/10.1086/466512).
 - [29] S. Cole et al. “The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications”. In: *Monthly Notices of the Royal Astronomical Society* 362.2 (2005), pp. 505–534. DOI: [10.1111/j.1365-2966.2005.09318.x](https://doi.org/10.1111/j.1365-2966.2005.09318.x).
 - [30] C. Frenk and S. White. “Dark matter and cosmic structure”. In: *Annalen der Physik* 524.9-10 (2012), pp. 507–534. DOI: <https://doi.org/10.1002/andp.201200212>.
 - [31] L. Söding, R. L. Bartel, and P. Mertsch. “Local dark matter density from Gaia DR3 K-dwarfs using Gaussian processes”. In: *Monthly Notices of the Royal Astronomical Society* 542.4 (Oct. 2025), pp. 2987–2997. DOI: [10.1093/mnras/staf1391](https://doi.org/10.1093/mnras/staf1391).
 - [32] A. Vallenari et al. “Gaia Data Release 3: Summary of the content and survey properties”. In: *Astronomy & Astrophysics* 674 (2023), A1. DOI: [10.1051/0004-6361/202243940](https://doi.org/10.1051/0004-6361/202243940).
 - [33] D. Baxter et al. “Recommended conventions for reporting results from direct dark matter searches”. In: *The European Physical Journal C* 81.10 (Oct. 2021). DOI: [10.1140/epjc/s10052-021-09655-y](https://doi.org/10.1140/epjc/s10052-021-09655-y).
 - [34] G. Arcadi et al. “The waning of the WIMP: endgame?” In: *The European Physical Journal C* 85.2 (Feb. 2025). DOI: [10.1140/epjc/s10052-024-13672-y](https://doi.org/10.1140/epjc/s10052-024-13672-y).
 - [35] G. Steigman, B. Dasgupta, and J. F. Beacom. “Precise relic WIMP abundance and its impact on searches for dark matter annihilation”. In: *Physical Review D* 86.2 (July 2012). DOI: [10.1103/physrevd.86.023506](https://doi.org/10.1103/physrevd.86.023506).
 - [36] L. Roszkowski, E. M. Sessolo, and S. Trojanowski. “WIMP dark matter candidates and searches—current status and future prospects”. In: *Reports on Progress in Physics* 81.6 (May 2018), p. 066201. DOI: [10.1088/1361-6633/aab913](https://doi.org/10.1088/1361-6633/aab913).
 - [37] H. P. Nilles. “Supersymmetry, Supergravity and Particle Physics”. In: *Phys. Rept.* 110 (1984), pp. 1–162. DOI: [10.1016/0370-1573\(84\)90008-5](https://doi.org/10.1016/0370-1573(84)90008-5).
 - [38] M. E. Peskin. “Supersymmetric dark matter in the harsh light of the Large Hadron Collider”. In: *Proceedings of the National Academy of Sciences* 112.40 (2015), pp. 12256–12263. DOI: [10.1073/pnas.1308787111](https://doi.org/10.1073/pnas.1308787111).
 - [39] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: [10.1103/PhysRevLett.13.138](https://doi.org/10.1103/PhysRevLett.13.138).
 - [40] D. J. Marsh. “Axion cosmology”. In: *Physics Reports* 643 (2016). Axion cosmology, pp. 1–79. DOI: <https://doi.org/10.1016/j.physrep.2016.06.005>.
 - [41] J. M. Pendlebury et al. “Revised experimental upper limit on the electric dipole moment of the neutron”. In: *Physical Review D* 92.9 (Nov. 2015). DOI: [10.1103/physrevd.92.092003](https://doi.org/10.1103/physrevd.92.092003).
 - [42] R. D. Peccei and H. R. Quinn. “CP Conservation in the Presence of Pseudoparticles”. In: *Phys. Rev. Lett.* 38 (25 June 1977), pp. 1440–1443. DOI: [10.1103/PhysRevLett.38.1440](https://doi.org/10.1103/PhysRevLett.38.1440).
-

-
- [43] P. Sikivie. “Axion Cosmology”. In: *Axions: Theory, Cosmology, and Experimental Searches*. Ed. by M. Kuster, G. Raffelt, and B. Beltrán. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 19–50. DOI: [10.1007/978-3-540-73518-2_2](https://doi.org/10.1007/978-3-540-73518-2_2).
 - [44] M. Shifman, A. Vainshtein, and V. Zakharov. “Can confinement ensure natural CP invariance of strong interactions?” In: *Nuclear Physics B* 166.3 (1980), pp. 493–506. DOI: [https://doi.org/10.1016/0550-3213\(80\)90209-6](https://doi.org/10.1016/0550-3213(80)90209-6).
 - [45] M. Dine, W. Fischler, and M. Srednicki. “A simple solution to the strong CP problem with a harmless axion”. In: *Physics Letters B* 104.3 (1981), pp. 199–202. DOI: [https://doi.org/10.1016/0370-2693\(81\)90590-6](https://doi.org/10.1016/0370-2693(81)90590-6).
 - [46] E. Armengaud et al. “Conceptual design of the International Axion Observatory (IAXO)”. In: *Journal of Instrumentation* 9.05 (May 2014), T05002–T05002. DOI: [10.1088/1748-0221/9/05/t05002](https://doi.org/10.1088/1748-0221/9/05/t05002).
 - [47] R. Khatriwada et al. “Axion Dark Matter Experiment: Detailed design and operations”. In: *Review of Scientific Instruments* 92.12 (Dec. 2021). DOI: [10.1063/5.0037857](https://doi.org/10.1063/5.0037857).
 - [48] S. Dimopoulos, G. D. Starkman, and B. W. Lynn. “Atomic enhancements in the detection of weakly interacting particles”. In: *Physics Letters B* 168.1 (1986), pp. 145–150. DOI: [https://doi.org/10.1016/0370-2693\(86\)91477-2](https://doi.org/10.1016/0370-2693(86)91477-2).
 - [49] J. Ye. “Searches for WIMPs and Axions with the XENON1T Experiment”. PhD thesis. UC, San Diego (main), UC, San Diego (main), 2020.
 - [50] M. Pospelov, A. Ritz, and M. Voloshin. “Bosonic super-WIMPs as keV-scale dark matter”. In: *Physical Review D* 78.11 (Dec. 2008). DOI: [10.1103/physrevd.78.115012](https://doi.org/10.1103/physrevd.78.115012).
 - [51] A. Filippi and M. De Napoli. “Searching in the dark: the hunt for the dark photon”. In: *Reviews in Physics* 5 (2020), p. 100042. DOI: <https://doi.org/10.1016/j.revip.2020.100042>.
 - [52] D. Hooper and T. Linden. “Origin of the gamma rays from the Galactic Center”. In: *Phys. Rev. D* 84 (12 Dec. 2011), p. 123005. DOI: [10.1103/PhysRevD.84.123005](https://doi.org/10.1103/PhysRevD.84.123005).
 - [53] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, and S. W. Randall. “Detection of an unidentified emission line in the stacked x-ray spectrum of galaxy clusters”. In: *Astrophysical Journal* 789.1 (2014). Cited by: 654; All Open Access, Bronze Open Access, Green Open Access. DOI: [10.1088/0004-637X/789/1/13](https://doi.org/10.1088/0004-637X/789/1/13).
 - [54] E. Aprile et al. “Search for New Physics in Electronic Recoil Data from XENONnT”. In: *Physical Review Letters* 129.16 (Oct. 2022). DOI: [10.1103/physrevlett.129.161805](https://doi.org/10.1103/physrevlett.129.161805).
 - [55] H. An, M. Pospelov, J. Pradler, and A. Ritz. “Direct detection constraints on dark photon dark matter”. In: *Physics Letters B* 747 (July 2015), pp. 331–338. DOI: [10.1016/j.physletb.2015.06.018](https://doi.org/10.1016/j.physletb.2015.06.018).
 - [56] A. De Roeck. “Dark matter searches at accelerators”. In: *Nuclear Physics B* 1003 (2024). Special Issue of Nobel Symposium 182 on Dark Matter, p. 116480. DOI: <https://doi.org/10.1016/j.nuclphysb.2024.116480>.
-

- [57] C. Pérez de los Heros. “Status, Challenges and Directions in Indirect Dark Matter Searches”. In: *Symmetry* 12.10 (Oct. 2020), p. 1648. DOI: [10.3390/sym12101648](https://doi.org/10.3390/sym12101648).
 - [58] J. Billard, E. Figueroa-Feliciano, and L. Strigari. “Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments”. In: *Physical Review D* 89.2 (Jan. 2014). DOI: [10.1103/physrevd.89.023524](https://doi.org/10.1103/physrevd.89.023524).
 - [59] C. A. J. O’Hare. “New Definition of the Neutrino Floor for Direct Dark Matter Searches”. In: *Physical Review Letters* 127.25 (Dec. 2021). DOI: [10.1103/physrevlett.127.251802](https://doi.org/10.1103/physrevlett.127.251802).
 - [60] C. O’Hare. *DirectDetectionPlots*. 2026. URL: <https://github.com/cajohare/DirectDetectionPlots/tree/main>.
 - [61] P. Agnes et al. “Low-Mass Dark Matter Search with the DarkSide-50 Experiment”. In: *Physical Review Letters* 121.8 (Aug. 2018). DOI: [10.1103/physrevlett.121.081307](https://doi.org/10.1103/physrevlett.121.081307).
 - [62] Z. Bo et al. “Dark Matter Search Results from 1.54 Tonne Year Exposure of PandaX-4T”. In: *Physical Review Letters* 134.1 (Jan. 2025). DOI: [10.1103/physrevlett.134.011805](https://doi.org/10.1103/physrevlett.134.011805).
 - [63] S. Li et al. “Search for Light Dark Matter with Ionization Signals in the PandaX-4T Experiment”. In: *Physical Review Letters* 130.26 (2023). DOI: [10.1103/physrevlett.130.261001](https://doi.org/10.1103/physrevlett.130.261001).
 - [64] J. Aalbers et al. “Dark Matter Search Results from 4.2 Tonne Years of Exposure of the LUX-ZEPLIN (LZ) Experiment”. In: *Physical Review Letters* 135.1 (July 2025). DOI: [10.1103/4dyc-z8zf](https://doi.org/10.1103/4dyc-z8zf).
 - [65] E. Aprile et al. *WIMP Dark Matter Search using a 3.1 Tonne-Year Exposure of the XENONnT Experiment*. 2025. DOI: <https://doi.org/10.1103/msw4-t342>.
 - [66] E. Aprile et al. *Light Dark Matter Search with 7.8 Tonne-Year of Ionization-Only Data in XENONnT*. 2026. URL: <https://arxiv.org/abs/2601.11296>.
 - [67] I. M. Bloch, R. Essig, K. Tobioka, T. Volansky, and T.-T. Yu. “Searching for dark absorption with direct detection experiments”. In: *Journal of High Energy Physics* 2017.6 (June 2017). DOI: [10.1007/jhep06\(2017\)087](https://doi.org/10.1007/jhep06(2017)087).
 - [68] T. Aralis et al. “Constraints on dark photons and axionlike particles from the SuperCDMS Soudan experiment”. In: *Phys. Rev. D* 101 (5 Mar. 2020), p. 052008. DOI: [10.1103/PhysRevD.101.052008](https://doi.org/10.1103/PhysRevD.101.052008).
 - [69] E. Armengaud et al. “Axion searches with the EDELWEISS-II experiment”. In: *Journal of Cosmology and Astroparticle Physics* 2013.11 (Nov. 2013), pp. 067–067. DOI: [10.1088/1475-7516/2013/11/067](https://doi.org/10.1088/1475-7516/2013/11/067).
 - [70] P. Agnes et al. “Search for Dark Matter Particle Interactions with Electron Final States with DarkSide-50”. In: *Physical Review Letters* 130.10 (Mar. 2023). DOI: [10.1103/physrevlett.130.101002](https://doi.org/10.1103/physrevlett.130.101002).
 - [71] M. Agostini et al. “First Search for Bosonic Superweakly Interacting Massive Particles with Masses up to 1 MeV with GERDA”. In: *Physical Review Letters* 125.1 (2020). DOI: [10.1103/physrevlett.125.011801](https://doi.org/10.1103/physrevlett.125.011801).
 - [72] T. Li et al. “Search for MeV-Scale Axionlike Particles and Dark Photons with PandaX-4T”. In: *Physical Review Letters* 134.7 (Feb. 2025). DOI: [10.1103/physrevlett.134.071004](https://doi.org/10.1103/physrevlett.134.071004).
-

-
- [73] E. Aprile et al. “First Indication of Solar ^8B Neutrinos via Coherent Elastic Neutrino-Nucleus Scattering with XENONnT”. In: *Physical Review Letters* 133.19 (Nov. 2024). DOI: [10.1103/physrevlett.133.191002](https://doi.org/10.1103/physrevlett.133.191002).
 - [74] E. Aprile et al. “First Search for Light Dark Matter in the Neutrino Fog with XENONnT”. In: *Physical Review Letters* 134.11 (Mar. 2025). DOI: [10.1103/physrevlett.134.111802](https://doi.org/10.1103/physrevlett.134.111802).
 - [75] XENON Collaboration et al. *Constraints on Solar Reflected Dark Matter from a combined analysis of XENON1T and XENONnT data*. 2025. URL: <https://arxiv.org/abs/2512.19592>.
 - [76] E. Aprile et al. “Challenging Spontaneous Quantum Collapse with the XENONnT Dark Matter Detector”. In: *Physical Review Letters* 136.12 (Mar. 2026). DOI: [10.1103/2jm3-4976](https://doi.org/10.1103/2jm3-4976).
 - [77] E. Aprile et al. *Enhancing Neutrinoless Double-Beta Decay Sensitivity of Liquid-Xenon Time Projection Chamber with Augmented Convolutional Neural Network*. 2026. URL: <https://arxiv.org/abs/2603.23549>.
 - [78] M. Kara et al. *The SNEWS 2.0 Alert Software for the Coincident Detection of Neutrinos from Core-Collapse Supernovae*. 2024. URL: <https://arxiv.org/abs/2406.17743>.
 - [79] E. Aprile et al. “The XENONnT dark matter experiment”. In: *The European Physical Journal C* 84.8 (Aug. 2024). DOI: [10.1140/epjc/s10052-024-12982-5](https://doi.org/10.1140/epjc/s10052-024-12982-5).
 - [80] D. Akerib et al. “The LUX-ZEPLIN (LZ) experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 953 (Feb. 2020), p. 163047. DOI: [10.1016/j.nima.2019.163047](https://doi.org/10.1016/j.nima.2019.163047).
 - [81] H. Zhang et al. “Dark matter direct search sensitivity of the PandaX-4T experiment”. In: *Science China Physics, Mechanics and Astronomy* 62.3 (Aug. 2018). DOI: [10.1007/s11433-018-9259-0](https://doi.org/10.1007/s11433-018-9259-0).
 - [82] C. E. Aalseth et al. “DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS”. In: *Eur. Phys. J. Plus* 133 (2018), p. 131. DOI: [10.1140/epjp/i2018-11973-4](https://doi.org/10.1140/epjp/i2018-11973-4).
 - [83] M. Auger et al. “The EXO-200 detector, part I: detector design and construction”. In: *Journal of Instrumentation* 7.05 (May 2012), P05010–P05010. DOI: [10.1088/1748-0221/7/05/p05010](https://doi.org/10.1088/1748-0221/7/05/p05010).
 - [84] J. Aalbers et al. “The XLZD Design Book: towards the next-generation liquid xenon observatory for dark matter and neutrino physics”. In: *The European Physical Journal C* 85.10 (Oct. 2025). DOI: [10.1140/epjc/s10052-025-14810-w](https://doi.org/10.1140/epjc/s10052-025-14810-w).
 - [85] M. Berglund and M. E. Wieser. “Isotopic compositions of the elements 2009 (IUPAC Technical Report)”. In: *Pure and Applied Chemistry* 83.2 (2011), pp. 397–410. DOI: [doi:10.1351/PAC-REP-10-06-02](https://doi.org/doi:10.1351/PAC-REP-10-06-02).
 - [86] E. Aprile et al. “Double-weak decays of ^{124}Xe ^{136}Xe and in the XENON1T and XENONnT experiments”. In: *Physical Review C* 106.2 (Aug. 2022). DOI: [10.1103/physrevc.106.024328](https://doi.org/10.1103/physrevc.106.024328).
 - [87] J. Aalbers et al. “Two-neutrino double electron capture of ^{124}Xe in the first LUX-ZEPLIN exposure”. In: *Journal of Physics G: Nuclear and Particle Physics* 52.1 (Dec. 2024), p. 015103. DOI: [10.1088/1361-6471/ad9039](https://doi.org/10.1088/1361-6471/ad9039).
-

- [88] S. A. Kharusi et al. *Search for Double Beta Decays of ^{134}Xe with EXO-200 Phase II*. 2025. URL: <https://arxiv.org/abs/2511.13990>.
 - [89] N. I. of Standards and Technology. *XCOM: Photon Cross Sections Database*. Tech. rep. Washington, D.C.: U.S. Department of Commerce, 2010. DOI: <https://dx.doi.org/10.18434/T48G6X>.
 - [90] N. I. of Standards and Technology. *Stopping-Power Range Tables for Electrons, Protons, and Helium Ions*. Tech. rep. Washington, D.C.: U.S. Department of Commerce, 2017. DOI: <https://dx.doi.org/10.18434/T4NC7P>.
 - [91] N. I. of Standards and Technology. *NIST Chemistry WebBook*. Tech. rep. Washington, D.C.: U.S. Department of Commerce, 2025. DOI: <https://doi.org/10.18434/T4D303>.
 - [92] E. Aprile and T. Doke. “Liquid xenon detectors for particle physics and astrophysics”. In: *Rev. Mod. Phys.* 82 (3 July 2010), pp. 2053–2097. DOI: [10.1103/RevModPhys.82.2053](https://doi.org/10.1103/RevModPhys.82.2053).
 - [93] T. Doke, A. Hitachi, J. Kikuchi, K. Masuda, H. Okada, and E. Shibamura. “Absolute Scintillation Yields in Liquid Argon and Xenon for Various Particles”. In: *Japanese Journal of Applied Physics* 41.3R (Mar. 2002), p. 1538. DOI: [10.1143/JJAP.41.1538](https://doi.org/10.1143/JJAP.41.1538).
 - [94] D. R. Nygren. “Proposal to investigate the feasibility of a novel concept in particle detection”. In: (1974).
 - [95] E. Aprile et al. “Design and performance of the XENON10 dark matter experiment”. In: *Astroparticle Physics* 34.9 (Apr. 2011), pp. 679–698. DOI: [10.1016/j.astropartphys.2011.01.006](https://doi.org/10.1016/j.astropartphys.2011.01.006).
 - [96] E. Aprile et al. “The triggerless data acquisition system of the XENONnT experiment”. In: *Journal of Instrumentation* 18.07 (July 2023), P07054. DOI: [10.1088/1748-0221/18/07/p07054](https://doi.org/10.1088/1748-0221/18/07/p07054).
 - [97] E. Aprile et al. *Radon Removal in XENONnT down to the Solar Neutrino Level*. 2025. URL: <https://arxiv.org/abs/2502.04209>.
 - [98] E. Aprile et al. “Conceptual design and simulation of a water Cherenkov muon veto for the XENON1T experiment”. In: *Journal of Instrumentation* 9.11 (Nov. 2014), P11006. DOI: [10.1088/1748-0221/9/11/P11006](https://doi.org/10.1088/1748-0221/9/11/P11006).
 - [99] E. Aprile et al. “The neutron veto of the XENONnT experiment: results with demineralized water”. In: *The European Physical Journal C* 85.6 (June 2025). DOI: [10.1140/epjc/s10052-025-14105-0](https://doi.org/10.1140/epjc/s10052-025-14105-0).
 - [100] E. Aprile et al. “Projected WIMP sensitivity of the XENONnT dark matter experiment”. In: *Journal of Cosmology and Astroparticle Physics* 2020.11 (Nov. 2020), pp. 031–031. DOI: [10.1088/1475-7516/2020/11/031](https://doi.org/10.1088/1475-7516/2020/11/031).
 - [101] E. Aprile et al. “Material radiopurity control in the XENONnT experiment”. In: *The European Physical Journal C* 82.7 (July 2022). DOI: [10.1140/epjc/s10052-022-10345-6](https://doi.org/10.1140/epjc/s10052-022-10345-6).
 - [102] E. Aprile et al. “Design and performance of the field cage for the XENONnT experiment”. In: *The European Physical Journal C* 84.2 (Feb. 2024). DOI: [10.1140/epjc/s10052-023-12296-y](https://doi.org/10.1140/epjc/s10052-023-12296-y).
 - [103] V. Chepel and H. Araújo. “Liquid noble gas detectors for low energy particle physics”. In: *Journal of Instrumentation* 8.04 (Apr. 2013), R04001–R04001. DOI: [10.1088/1748-0221/8/04/r04001](https://doi.org/10.1088/1748-0221/8/04/r04001).
-

-
- [104] J. Lindhard, M. Scharff, and H. E. Schiøtt. “RANGE CONCEPTS AND HEAVY ION RANGES (NOTES ON ATOMIC COLLISIONS, II)”. In: 1963. URL: <https://api.semanticscholar.org/CorpusID:94740919>.
 - [105] M. Szydagis, A. Fyhrie, D. Thorngren, and M. Tripathi. “Enhancement of NEST capabilities for simulating low-energy recoils in liquid xenon”. In: *Journal of Instrumentation* 8.10 (Oct. 2013), p. C10003. DOI: [10.1088/1748-0221/8/10/C10003](https://doi.org/10.1088/1748-0221/8/10/C10003).
 - [106] B. Lenardo, K. Kazkaz, A. Manalaysay, J. Mock, M. Szydagis, and M. Tripathi. “A Global Analysis of Light and Charge Yields in Liquid Xenon”. In: *IEEE Transactions on Nuclear Science* 62.6 (2015), pp. 3387–3396. DOI: [10.1109/TNS.2015.2481322](https://doi.org/10.1109/TNS.2015.2481322).
 - [107] G. Plante. “The XENON100 Dark Matter Experiment: Design, Construction, Calibration and 2010 Search Results with Improved Measurement of the Scintillation Response of Liquid Xenon to Low-Energy Nuclear Recoils”. PhD thesis. Jan. 2012.
 - [108] G. Volta. “Characterization and Monitoring of XENONnT Photosensors and Search for New Physics with the First XENONnT Science Data”. PhD thesis. UNIVERSITÄT ZÜRICH, 2023.
 - [109] J. Thomas and D. A. Imel. “Recombination of electron-ion pairs in liquid argon and liquid xenon”. In: *Phys. Rev. A* 36 (2 July 1987), pp. 614–616. DOI: [10.1103/PhysRevA.36.614](https://doi.org/10.1103/PhysRevA.36.614).
 - [110] M. Szydagis et al. *Noble Element Simulation Technique*. Version v2.4.5beta. Dec. 2025. DOI: [10.5281/zenodo.17851406](https://doi.org/10.5281/zenodo.17851406).
 - [111] XENON collaboration. *XENONnT blessed plots*. XENONnT internal resource.
 - [112] XENON Collaboration et al. *XENONnT Analysis: Signal Reconstruction, Calibration and Event Selection*. 2024. URL: <https://arxiv.org/abs/2409.08778>.
 - [113] E. Aprile et al. “XENON1T dark matter data analysis: Signal and background models and statistical inference”. In: *Physical Review D* 99.11 (June 2019). DOI: [10.1103/physrevd.99.112009](https://doi.org/10.1103/physrevd.99.112009).
 - [114] E. Aprile et al. “Observation and applications of single-electron charge signals in the XENON100 experiment”. In: *Journal of Physics G: Nuclear and Particle Physics* 41.3 (Feb. 2014), p. 035201. DOI: [10.1088/0954-3899/41/3/035201](https://doi.org/10.1088/0954-3899/41/3/035201).
 - [115] D. S. Akerib et al. *Study of few-electron backgrounds in the LUX-ZEPLIN detector*. 2026. URL: <https://arxiv.org/abs/2510.06500>.
 - [116] J. Xu et al. “Electron extraction efficiency study for dual-phase xenon dark matter experiments”. In: *Phys. Rev. D* 99 (10 May 2019), p. 103024. DOI: [10.1103/PhysRevD.99.103024](https://doi.org/10.1103/PhysRevD.99.103024).
 - [117] J. R. Angevaere et al. *XENONnT/straxen*: v3.2.8. Version v3.2.8. Mar. 2026. DOI: [10.5281/zenodo.18841409](https://doi.org/10.5281/zenodo.18841409).
 - [118] J. Aalbers et al. *AxFoundation/strax*: v2.2.2. Version v2.2.2. Dec. 2025. DOI: [10.5281/zenodo.18727696](https://doi.org/10.5281/zenodo.18727696).
 - [119] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, De-*
-

- tectors and Associated Equipment* 506.3 (2003), pp. 250–303. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [120] H. S. Eißing et al. *XENONnT/fuse: 1.6.3*. Version 1.6.3. Jan. 2026. DOI: [10.5281/zenodo.18418865](https://doi.org/10.5281/zenodo.18418865).
 - [121] D. Xu et al. *XENONnT/alea: v0.3.5*. Version v0.3.5. Mar. 2026. DOI: [10.5281/zenodo.18840642](https://doi.org/10.5281/zenodo.18840642).
 - [122] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011). DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0).
 - [123] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. *Power-Constrained Limits*. 2011. URL: <https://arxiv.org/abs/1105.3166>.
 - [124] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: [10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360).
 - [125] R. Hammann. “Probing the Invisible: From Dark Matter and Neutrinos in XENONnT to Infrared Scintillation for Future Xenon Detectors”. PhD thesis. Heidelberg University, 2025.
 - [126] D. Akerib et al. “Position reconstruction in LUX”. In: *Journal of Instrumentation* 13.02 (Feb. 2018), P02001–P02001. DOI: [10.1088/1748-0221/13/02/p02001](https://doi.org/10.1088/1748-0221/13/02/p02001).
 - [127] J. Neyman and E. S. Pearson. “IX. On the problem of the most efficient tests of statistical hypotheses”. In: *Philosophical Transactions of the Royal Society of London, Series A: Containing Papers of a Mathematical or Physical Character* 231.694-706 (Feb. 1933), pp. 289–337. DOI: [10.1098/rsta.1933.0009](https://doi.org/10.1098/rsta.1933.0009).
 - [128] *Merriam Webster: machine learning*. <https://www.merriam-webster.com/dictionary/machinelearning>. Accessed: 2026-05-04.
 - [129] OpenAI et al. *GPT-4 Technical Report*. 2024. URL: <https://arxiv.org/abs/2303.08774>.
 - [130] F. Rosenblatt. “The perceptron: A probabilistic model for information storage and organization in the brain”. In: *Psychological Review* (1958). URL: <https://doi.org/10.1037/h0042519>.
 - [131] A. Jung. “Machine Learning: The Basics”. In: Springer, 2022. URL: <https://arxiv.org/abs/1805.05052>.
 - [132] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. 2016. URL: <https://arxiv.org/abs/1511.07289>.
 - [133] A. F. Agarap. *Deep Learning using Rectified Linear Units (ReLU)*. 2019. URL: <https://arxiv.org/abs/1803.08375>.
 - [134] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri. “Activation functions in deep learning: A comprehensive survey and benchmark”. In: *Neurocomputing* 503 (2022), pp. 92–108. DOI: <https://doi.org/10.1016/j.neucom.2022.06.111>.
 - [135] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. URL: <https://arxiv.org/abs/1412.6980>.
 - [136] I. Loshchilov and F. Hutter. *Decoupled Weight Decay Regularization*. 2019. URL: <https://arxiv.org/abs/1711.05101>.
-

-
- [137] D. Wolpert and W. Macready. “No free lunch theorems for optimization”. In: *IEEE Transactions on Evolutionary Computation* 1.1 (1997), pp. 67–82. DOI: [10.1109/4235.585893](https://doi.org/10.1109/4235.585893).
 - [138] V. Lakshmanan, S. Robinson, and M. Munn. “Machine Learning Design Patterns”. In: O’Reilly Media, 2020. Chap. 4. URL: <https://books.google.de/books?id=djwDEAAQBAJ>.
 - [139] A. Alwosheel, S. van Cranenburgh, and C. G. Chorus. “Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis”. In: *Journal of Choice Modelling* 28 (2018), pp. 167–182. DOI: <https://doi.org/10.1016/j.jocm.2018.07.002>.
 - [140] S. Liang et al. “Domain-Informed Neural Networks for Interaction Localization Within Astroparticle Experiments”. In: *Front. Artif. Intell.* 5 (2022), p. 832909. DOI: [10.3389/frai.2022.832909](https://doi.org/10.3389/frai.2022.832909).
 - [141] D. Harris and S. Harris. “Digital Design and Computer Architecture”. In: Morgan Kaufmann, 2012, pp. 123–125.
 - [142] A. Azzalini and A. D. Valle. “The Multivariate Skew-Normal Distribution”. In: *Biometrika* 83.4 (1996), pp. 715–726. URL: <http://www.jstor.org/stable/2337278> (visited on Apr. 2, 2026).
 - [143] L. Dinh, J. Sohl-Dickstein, and S. Bengio. *Density estimation using Real NVP*. 2017. URL: <https://arxiv.org/abs/1605.08803>.
 - [144] D. Rezende and S. Mohamed. “Variational Inference with Normalizing Flows”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by F. Bach and D. Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, July 2015, pp. 1530–1538. URL: <https://proceedings.mlr.press/v37/rezende15.html>.
 - [145] E. G. Tabak and C. V. Turner. “A Family of Nonparametric Density Estimation Algorithms”. In: *Communications on Pure and Applied Mathematics* 66.2 (2013), pp. 145–164. DOI: <https://doi.org/10.1002/cpa.21423>.
 - [146] I. Kobyzev, S. J. Prince, and M. A. Brubaker. “Normalizing Flows: An Introduction and Review of Current Methods”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.11 (2021), pp. 3964–3979. DOI: [10.1109/TPAMI.2020.2992934](https://doi.org/10.1109/TPAMI.2020.2992934).
 - [147] J. Riebesell and S. Bringuier. *Collection of scientific diagrams*. Version 0.2.0. 10.5281/zenodo.7486911 - <https://github.com/janosh/diagrams>. Aug. 9, 2020. DOI: [10.5281/zenodo.7486911](https://doi.org/10.5281/zenodo.7486911).
 - [148] C. Winkler, D. Worrall, E. Hoogeboom, and M. Welling. *Learning Likelihoods with Conditional Normalizing Flows*. 2023. URL: <https://arxiv.org/abs/1912.00042>.
 - [149] D. Ward. *FlowJAX: Distributions and Normalizing Flows in Jax*. Version 12.2.0. 2024. DOI: [10.5281/zenodo.10402073](https://doi.org/10.5281/zenodo.10402073).
 - [150] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. *Neural Spline Flows*. 2019. URL: <https://arxiv.org/abs/1906.04032>.
 - [151] P. Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
 - [152] K. Cranmer, J. Brehmer, and G. Louppe. “The frontier of simulation-based inference”. In: *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30055–30062. DOI: [10.1073/pnas.1912789117](https://doi.org/10.1073/pnas.1912789117).
-

- [153] H. Dembinski and P. O. et al. “scikit-hep/iminuit”. In: (Dec. 2020). DOI: [10.5281/zenodo.3949207](https://doi.org/10.5281/zenodo.3949207).
 - [154] F. Toschi. “Design of the field cage and charge response of the XENONnT dark matter experiment”. PhD thesis. Albert-Ludwigs-Universität Freiburg im Breisgau, 2022.
 - [155] G. F. Jenks. “The Data Model Concept in Statistical Mapping”. In: 1967. URL: <https://api.semanticscholar.org/CorpusID:215850874>.
 - [156] J. B. Albert et al. “Measurement of the drift velocity and transverse diffusion of electrons in liquid xenon with the EXO-200 detector”. In: *Physical Review C* 95.2 (Feb. 2017). DOI: [10.1103/physrevc.95.025502](https://doi.org/10.1103/physrevc.95.025502).
 - [157] T. Doke. “Fundamental Properties of Liquid Argon, Krypton and Xenon as Radiation Detector Media”. In: *Portugal. Phys.* 12 (1981), pp. 9–48.
 - [158] Y. Biondi. “Sensitivity of DARWIN to Rare Events and the Purity Monitor for Xenoscope”. PhD thesis. 2022.
 - [159] E. Aprile, K. L. Giboni, P. Majewski, K. Ni, and M. Yamashita. “Proportional Light in a Dual-Phase Xenon Chamber”. In: *IEEE Transactions on Nuclear Science* 51.5 (Jan. 2004), pp. 1986–1990. DOI: [10.1109/TNS.2004.832690](https://doi.org/10.1109/TNS.2004.832690).
 - [160] V. H.-S. Wu. “Electrode Realisation and SiPM Characterisation for Dark Matter Search with XENONnT and XLZD”. PhD thesis. Feb. 2026.
 - [161] E. Aprile et al. “The XENON100 Dark Matter Experiment”. In: *Astropart. Phys.* 35 (2012), pp. 573–590. DOI: [10.1016/j.astropartphys.2012.01.003](https://doi.org/10.1016/j.astropartphys.2012.01.003).
 - [162] D. Akerib et al. “The Large Underground Xenon (LUX) experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 704 (Mar. 2013), pp. 111–126. DOI: [10.1016/j.nima.2012.11.135](https://doi.org/10.1016/j.nima.2012.11.135).
 - [163] D. S. Akerib et al. “First Results from the LUX Dark Matter Experiment at the Sanford Underground Research Facility”. In: *Physical Review Letters* 112.9 (Mar. 2014). DOI: [10.1103/physrevlett.112.091303](https://doi.org/10.1103/physrevlett.112.091303).
 - [164] E. Aprile et al. “Physics reach of the XENON1T dark matter experiment.” In: *Journal of Cosmology and Astroparticle Physics* 2016.04 (Apr. 2016), pp. 027–027. DOI: [10.1088/1475-7516/2016/04/027](https://doi.org/10.1088/1475-7516/2016/04/027).
 - [165] E. Aprile et al. “The XENON1T dark matter experiment”. In: *Eur. Phys. J. C* 77.12 (2017), p. 881. DOI: [10.1140/epjc/s10052-017-5326-3](https://doi.org/10.1140/epjc/s10052-017-5326-3).
 - [166] K. Mistry et al. “Design, characterization and installation of the NEXT-100 cathode and electroluminescence regions”. In: *Journal of Instrumentation* 19.02 (Feb. 2024), P02007. DOI: [10.1088/1748-0221/19/02/p02007](https://doi.org/10.1088/1748-0221/19/02/p02007).
 - [167] PCM Products Inc. *Photo chemical etching and machining*. 2025.
 - [168] D. S. Akerib et al. “Investigation of background electron emission in the LUX detector”. In: *Physical Review D* 102.9 (Nov. 2020). DOI: [10.1103/physrevd.102.092004](https://doi.org/10.1103/physrevd.102.092004).
 - [169] A. Tomás et al. “Study and mitigation of spurious electron emission from cathodic wires in noble liquid time projection chambers”. In: *Astroparticle Physics* 103 (Dec. 2018), pp. 49–61. DOI: [10.1016/j.astropartphys.2018.07.001](https://doi.org/10.1016/j.astropartphys.2018.07.001).
-

- [170] Y. Xia, X. Cao, F. Wen, G. Hua, and J. Sun. “Learning Discriminative Reconstructions for Unsupervised Outlier Removal”. In: *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1511–1519. URL: <https://api.semanticscholar.org/CorpusID:13982294>.
- [171] J. An and S. Cho. “Variational Autoencoder based Anomaly Detection using Reconstruction Probability”. In: 2015. URL: <https://api.semanticscholar.org/CorpusID:36663713>.
- [172] X. Tao, D. Zhang, W. Ma, X. Liu, and D. Xu. “Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks”. In: *Applied Sciences* 8,9 (2018). DOI: [10.3390/app8091575](https://doi.org/10.3390/app8091575).
- [173] D. Gong et al. “Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection”. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, Nov. 2019, pp. 1705–1714. DOI: [10.1109/ICCV.2019.00179](https://doi.org/10.1109/ICCV.2019.00179).
- [174] Z. Chen, X. Feng, L. Liu, and Z. Jia. “Surface defect detection of industrial components based on vision”. In: *Nature Scientific Reports* (2023). DOI: [10.1038/s41598-023-49359-9](https://doi.org/10.1038/s41598-023-49359-9).
- [175] H. Hojjati, T. K. K. Ho, and N. Armanfard. “Self-supervised anomaly detection in computer vision and beyond: A survey and outlook”. In: *Neural Networks* 172 (2024), p. 106106. DOI: <https://doi.org/10.1016/j.neunet.2024.106106>.
- [176] I. O. for Standardization. *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*. 1994.
- [177] J. Beyerer, F. Puente León, and C. Frese. “Preprocessing and Image Enhancement”. In: *Machine Vision: Automated Visual Inspection: Theory, Practice and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 465–519. DOI: [10.1007/978-3-662-47794-6_9](https://doi.org/10.1007/978-3-662-47794-6_9).
- [178] L. Shapiro and G. Stockman. *Computer Vision*. Prentice Hall, 2001. URL: <https://books.google.de/books?id=FftDAQAAIAAJ>.
- [179] S. van der Walt et al. “scikit-image: image processing in Python”. In: *PeerJ* 2 (June 2014), e453. DOI: [10.7717/peerj.453](https://doi.org/10.7717/peerj.453).
- [180] J. Beyerer, F. Puente León, and C. Frese. “Morphological Image Processing”. In: *Machine Vision: Automated Visual Inspection: Theory, Practice and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 607–647. DOI: [10.1007/978-3-662-47794-6_12](https://doi.org/10.1007/978-3-662-47794-6_12).
- [181] G. Bradski. “The OpenCV Library”. In: *Dr. Dobb’s Journal of Software Tools* (2000).
- [182] J. Beyerer, F. Puente León, and C. Frese. “Segmentation”. In: *Machine Vision: Automated Visual Inspection: Theory, Practice and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 553–605. DOI: [10.1007/978-3-662-47794-6_11](https://doi.org/10.1007/978-3-662-47794-6_11).
- [183] K. Wada et al. *wkentaro/labelme: v4.6.0*. Version v4.6.0. Nov. 2021. DOI: [10.5281/zenodo.5711226](https://doi.org/10.5281/zenodo.5711226).
- [184] T. K. Ho. “Random decision forests”. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 1995, 278–282 vol.1. DOI: [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994).

- [185] P. Mehta et al. “A high-bias, low-variance introduction to Machine Learning for physicists”. In: *Physics Reports* 810 (May 2019), pp. 1–124. DOI: [10.1016/j.physrep.2019.03.001](https://doi.org/10.1016/j.physrep.2019.03.001).
 - [186] L. Grinsztajn, E. Oyallon, and G. Varoquaux. *Why do tree-based models still outperform deep learning on tabular data?* 2022. URL: <https://arxiv.org/abs/2207.08815>.
 - [187] C. G. Northcutt, A. Athalye, and J. Mueller. *Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks*. 2021. URL: <https://arxiv.org/abs/2103.14749>.
 - [188] D. Wenz. “Commissioning of the world’s first water Cherenkov neutron veto and first WIMP dark matter search results of the XENONnT experiment”. PhD thesis. Johannes Gutenberg-Universität in Mainz, 2023.
 - [189] D. Rumelhart, G. Hinton, and R. Williams. “Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations”. In: MIT Press, Cambridge, MA, 1986. Chap. 8.
 - [190] D. Bank, N. Koenigstein, and R. Giryes. *Autoencoders*. 2021. URL: <https://arxiv.org/abs/2003.05991>.
 - [191] G. W. Cottrell, P. Munro, and D. Zipser. “Learning Internal Representation From Gray-Scale Images: An Example of Extensional Programming”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. 1987.
 - [192] D. Harrison. *A Connectionist Framework for Continuous Speech Recognition*. University of Cambridge PhD Dissertation, January 1988. 1988. URL: <https://books.google.de/books?id=FjG30QEACAAJ>.
 - [193] L. Deng. “The mnist database of handwritten digit images for machine learning research”. In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 141–142.
 - [194] G. E. Hinton and R. R. Salakhutdinov. “Reducing the Dimensionality of Data with Neural Networks”. In: *Science* 313.5786 (2006), pp. 504–507. DOI: [10.1126/science.1127647](https://doi.org/10.1126/science.1127647).
 - [195] D. P. Kingma and M. Welling. “Auto-Encoding Variational Bayes”. In: (2022). URL: <https://arxiv.org/abs/1312.6114>.
 - [196] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. URL: <http://www.jstor.org/stable/2236703> (visited on Apr. 11, 2026).
 - [197] D. J. Rezende, S. Mohamed, and D. Wierstra. *Stochastic Backpropagation and Approximate Inference in Deep Generative Models*. 2014. URL: <https://arxiv.org/abs/1401.4082>.
 - [198] D. P. Kingma and M. Welling. “An Introduction to Variational Autoencoders”. In: *Foundations and Trends® in Machine Learning* 12.4 (Nov. 2019), pp. 307–392. DOI: [10.1561/22000000056](https://doi.org/10.1561/22000000056).
 - [199] Y. Lecun, P. Haffner, Y. Rachmad, and L. Bottou. “Gradient-Based Learning Applied to Document Recognition”. In: *Proceedings of the IEEE* 86 (Dec. 1998), pp. 2278–2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
 - [200] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. “Deconvolutional networks”. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 2528–2535. DOI: [10.1109/CVPR.2010.5539957](https://doi.org/10.1109/CVPR.2010.5539957).
-

-
- [201] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio. *Generating Sentences from a Continuous Space*. 2016. URL: <https://arxiv.org/abs/1511.06349>.
 - [202] O. Ronneberger, P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. URL: <https://arxiv.org/abs/1505.04597>.
 - [203] M. Reid. *MatthewReid854/reliability: v0.8.16*. Version v0.8.16. Dec. 2023. DOI: [10.5281/zenodo.10433917](https://doi.org/10.5281/zenodo.10433917).
 - [204] S. Mitra et al. “GRANITE: high-resolution imaging and electrical qualification of large-area TPC electrodes”. In: *Journal of Instrumentation* 21.01 (Jan. 2026), P01034. DOI: [10.1088/1748-0221/21/01/p01034](https://doi.org/10.1088/1748-0221/21/01/p01034).
 - [205] A. Götz. “Development of a Prototype for an Automated HV Scanning System for XLZD Electrodes”. M.Sc. thesis. Karlsruhe Institute of Technology, 2025.
 - [206] Y. Hochberg, B. von Krosigk, E. Kuflik, and T. C. Yu. “Impact of Dark Compton Scattering on Direct Dark Matter Absorption Searches”. In: *Phys. Rev. Lett.* 128 (19 May 2022), p. 191801. DOI: [10.1103/PhysRevLett.128.191801](https://doi.org/10.1103/PhysRevLett.128.191801).
 - [207] E. Aprile et al. *Spectral Measurement of the ^{214}Bi beta-decay to the ^{214}Po Ground State with XENONnT*. 2025. URL: <https://arxiv.org/abs/2510.04846>.
 - [208] M. Guida. “Low-energy electronic recoils in XENONnT: new physics searches, first sub-keV calibration, and improved krypton assay”. PhD thesis. Heidelberg University, 2025.
 - [209] M. P. Wand. “Data-Based Choice of Histogram Bin Width”. In: *The American Statistician* 51.1 (1997), pp. 59–64. DOI: [10.1080/00031305.1997.10473591](https://doi.org/10.1080/00031305.1997.10473591).
 - [210] C. J. Clopper and E. S. Pearson. “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial”. In: *Biometrika* 26.4 (1934), pp. 404–413.
 - [211] P. Gaemers et al. *XENONnT/WFSim: v1.0.2*. Version v1.0.2. Oct. 2022. DOI: [10.5281/zenodo.7216324](https://doi.org/10.5281/zenodo.7216324).
 - [212] S. Schoene and M. Leuermann. *Multi-dimensional Kernel Density Estimation (KDE) including adaptive bandwidths*. Apr. 23, 2019. URL: <https://github.com/icecube/kde/tree/master> (visited on Apr. 29, 2026).
 - [213] E. and Aprile et al. “ ^{222}Rn emanation measurements for the XENON1T experiment”. In: *The European Physical Journal C* 81.4 (Apr. 2021). DOI: [10.1140/epjc/s10052-020-08777-z](https://doi.org/10.1140/epjc/s10052-020-08777-z).
 - [214] C. Ferrari. “Radon-Induced Lead Isotopes in XENONnT: Studies on Branching Ratios and Surface Background”. PhD thesis. Gran Sasso Science Institute, 2025.
 - [215] Y. Khazov, A. Rodionov, and F. Kondev. “Nuclear Data Sheets for A = 133”. In: *Nuclear Data Sheets* 112.4 (2011), pp. 855–1113. DOI: <https://doi.org/10.1016/j.nds.2011.03.001>.
 - [216] J. Timar, Z. Elekes, and B. Singh. “Nuclear Data Sheets for A = 129”. In: *Nucl. Data Sheets* 121 (2014), pp. 143–394. DOI: [10.1016/j.nds.2014.09.002](https://doi.org/10.1016/j.nds.2014.09.002).
-

- [217] “Nuclear Data Sheets for A = 131”. In: *Nuclear Data Sheets* 107.11 (2006), pp. 2715–2930. DOI: <https://doi.org/10.1016/j.nds.2006.10.001>.
 - [218] R. Auble. “Nuclear data sheets for A=127”. In: *Nuclear Data Sheets* 8.2 (1972), pp. 77–121. DOI: [https://doi.org/10.1016/S0090-3752\(72\)80017-6](https://doi.org/10.1016/S0090-3752(72)80017-6).
 - [219] J. Katakura. “Nuclear Data Sheets for A = 125”. In: *Nuclear Data Sheets* 112.3 (2011), pp. 495–705. DOI: <https://doi.org/10.1016/j.nds.2011.02.001>.
 - [220] H. Bateman. “The solution of a system of differential equations occurring in the theory of radioactive transformations”. In: *Cambridge Philosophical Society*. Vol. 15. 1910.
 - [221] M. Murra. *Estimate of purification lifetime of oxygen contaminants*. Internal communication. 2025.
 - [222] M. Guida, Y.-T. Lin, and H. Simgen. “Improved and automated krypton assay for low-background xenon detectors with Auto-RGMS”. In: *The European Physical Journal C* 85.5 (May 2025). DOI: [10.1140/epjc/s10052-025-14262-2](https://doi.org/10.1140/epjc/s10052-025-14262-2).
 - [223] J. B. Albert et al. “Improved measurement of the $2\nu\beta\beta$ half-life of ^{136}Xe with the EXO-200 detector”. In: *Phys. Rev. C* 89 (1 Jan. 2014), p. 015502. DOI: [10.1103/PhysRevC.89.015502](https://doi.org/10.1103/PhysRevC.89.015502).
 - [224] L. W. Kastens, S. B. Cahn, A. Manzur, and D. N. McKinsey. “Calibration of a liquid xenon detector with Kr83m”. In: *Physical Review C* 80.4 (Oct. 2009). DOI: [10.1103/physrevc.80.045809](https://doi.org/10.1103/physrevc.80.045809).
 - [225] E. Aprile et al. “Excess electronic recoil events in XENON1T”. In: *Physical Review D* 102.7 (Oct. 2020). DOI: [10.1103/physrevd.102.072004](https://doi.org/10.1103/physrevd.102.072004).
 - [226] J.-W. Chen, H.-C. Chi, C.-P. Liu, and C.-P. Wu. “Low-energy electronic recoil in xenon detectors by solar neutrinos”. In: *Physics Letters B* 774 (2017), pp. 656–661. DOI: <https://doi.org/10.1016/j.physletb.2017.10.029>.
 - [227] M. Pallavicini et al. “First real-time detection of solar pp neutrinos by Borexino”. In: *EPJ Web Conf.* 121 (2016). Ed. by P. Piattelli et al., p. 01001. DOI: [10.1051/epjconf/201612101001](https://doi.org/10.1051/epjconf/201612101001).
 - [228] S. Kumaran, L. Ludhova, Ö. Penek, and G. Settanta. “Borexino Results on Neutrinos from the Sun and Earth”. In: *Universe* 7.7 (2021). DOI: [10.3390/universe7070231](https://doi.org/10.3390/universe7070231).
 - [229] A. Bernard, O. Dadoun, and L. Althüser. *Light Collection Efficiency Map for S2 Simulation*. XENONnT internal resource.
 - [230] S. Algeri, J. Aalbers, K. D. Morå, and J. Conrad. “Searching for new phenomena with profile likelihood ratio tests”. In: *Nature Reviews Physics* 2.5 (Apr. 2020), pp. 245–252. DOI: [10.1038/s42254-020-0169-5](https://doi.org/10.1038/s42254-020-0169-5).
-