

Signal Pairing Algorithm for XENONnT Event Selection

REU Program at Columbia University - Nevis Labs

Shriya Nagulapally¹

¹University of California, Davis

August 3, 2024

Abstract

Over the past century, astronomical observations have motivated the search for a non-baryonic, non-luminous particle that only interacts with the gravitational force. This elusive matter, known as dark matter, accounts for 26% of the Universe, yet the particle candidate remains unknown. The XENON Dark Matter Project aims to study dark matter by focusing on directly detecting Weakly Interacting Massive Particles (WIMPs), a well-accepted candidate for dark matter due to its agreement with Beyond-Standard-Model theories. Scintillation light signals produced from an incoming particle's collision with xenon atoms are recorded and used to reconstruct the collision event and analyze the particle's three-dimensional position and energy. An in-depth look at the collision event reconstruction from signal to event level process reveals a poor reconstruction of some events, leading to a misidentification of signal and noise events. An updated algorithm may be useful in identifying signal events and noise events with higher accuracy. Data selections corresponding to the area of the signal light, how photoelectrons are dispersed, and the depth of the interaction site can be correlated in pairs and utilized to produce best-fit functions. This study analyzes low-energy calibration data to motivate and specify parameters for a best-fit event pairing algorithm derived from these best-fit functions.



Contents

1	Introduction	3
1.1	Dark Matter Motivation	3
1.2	XENONnT Time Projection Chamber	3
2	Signal and Position Reconstruction	5
2.1	Event Building Process	5
2.2	Background and Noise Reduction	7
2.3	Data Selection Acceptance and Cuts	7
3	Development of Event Builder Pairing Algorithm	8
3.1	Analysis of Current Event Builder	8
3.2	P-Value Based Quartile Fit	9
3.3	Density of Saved Mismatched Events	11
4	Results and Discussion	12
5	Next Steps	13
6	Acknowledgements	15

1 Introduction

1.1 Dark Matter Motivation

The concept of dark matter was proposed as astronomical observations over the past century highlighted a gap in our current understanding of physics. In the 1930s, Dr. Fritz Zwicky observed that galaxy speeds within the Coma galaxy cluster varied from the projected speeds calculated using the visible mass of the cluster. In the 1970s, Dr. Vera Rubin noticed that objects toward the edge of galaxies rotated faster than hypothesized by galactic mass distribution. Other evidence such as density fluctuations in the Cosmic Microwave Background and distortions in gravitational lensing support these prior discoveries in uncovering inconsistencies in the Universe. A proposed explanation for the hypothesized increased mass is explained by a nearly undetectable particle: dark matter. Dark matter is a beyond-standard-model, non-baryonic particle that does not emit light and only interacts with the gravitational force.

Based on theoretical and observational evidence, the dark matter candidate XENONnT aims to observe is Weakly Interacting Massive Particles, or WIMPs for short. WIMPs are theorized by beyond-standard-model theories such as supersymmetry, which theorize that WIMPs were thermally produced in the early universe similar to Standard Model (SM) particles and decoupled from thermal plasma earlier than SM particles, leading to density fluctuations in the Cosmic Microwave Background [6]. A crucial characteristics of WIMPs are their theorized interactions not just with the gravitational force, but also with the weak scale force. Currently, evidence of dark matter is only given via interactions with gravitation. However, due to the especially weak nature of gravitational force and the small size of the particle, dark matter interactions with the gravitational field are too small to be detected in any capacity. Accordingly, a stronger signal is needed to confirm the identity of dark matter. Based on observational measurements of galaxy rotation curves and the mass distribution of galaxy clusters, we conclude that dark matter cannot have a strong nuclear force nor an electric charge. As such, forces dark matter may interact with in addition to gravity include the weak scale force, a very weak photon coupling force, or an unknown force. Dark matter interactions with the weak scale force can produce recoils strong enough to be detected by ultra-sensitive detectors. WIMPs are said to interact with the weak-scale force, and as such, direct detection of WIMPs relies on the assumption that dark matter interacts with the weak scale force. More information regarding WIMP selection can be found at [7].

The approach to dark matter detection relies on observing dark matter interactions with itself or standard model particles. Interactions with standard model particles can provide insight into the size, energy, and behavior of dark matter particles through detectable matter. The XENON Dark Matter Project focuses on directly detecting dark matter interactions with standard model particles via interactions with xenon atoms.

1.2 XENONnT Time Projection Chamber

The XENONnT detector, housed at the INFN Laboratori Nazionali del Gran Sasso (LNGS) underground laboratory, is a 8.6-tonne dual-phase Time Project Chamber (TPC) made up of liquid xenon (LXe) and a thin gaseous xenon layer (GXe). Due to the dual-phase nature of the TPC, events that occur within the tank can be reconstructed to determine the incoming particle's size, path, and energy [5].

The time projection chamber utilizes xenon as its target material due to its many favorable qualities. Xenon is a rare noble gas; noble gases respond to radiation with a high level of electron and photon scintillation light. Additionally, xenon has a low ionization energy due to its large atomic number, and thus electrons are easily stripped from its shell; the number of ionized electrons depends on the energy and size of the incoming particle. The high atomic number alongside xenon's high liquid density contribute to it having the largest stopping power for radiation of all noble gases. Finally, xenon can be cleaned to extreme purity from electronegative and radioactive contaminants, which is key in reducing signal interference as photoelectrons drift away from the interaction site to the detection sites[2]. These favorable properties of xenon make it an ideal material for particle detection.

The dual-phase construction of the XENONnT TPC produces two different signals that are necessary for a three-dimensional event reconstruction. The first signal is released upon particle collision with xenon as a photonic light signal called an S1 (light) signal. Upon collision, the particle interacts with xenon's electron shell and/or its nucleus, which ionizes some electrons in the xenon electron shell. These ionized electrons drift upwards towards the gaseous xenon layer, and upon interaction with the gaseous xenon, release a second charge scintillation signal called an S2 (charge) signal. To collect the photoelectrons produced from the S1 and S2 signals, 494 photomultiplier tubes (PMTs) are placed in two arrays at the top and bottom of the TPC. To induce an upward drift of ionized electrons for S2 signal production an electric field, located at the boundary between liquid and gaseous xenon at the top of the detector, is applied across the full liquid volume of xenon. As a method of removing similar signals to WIMPs, a large water tank encloses the detector to reject gamma muons and neutron events [4].

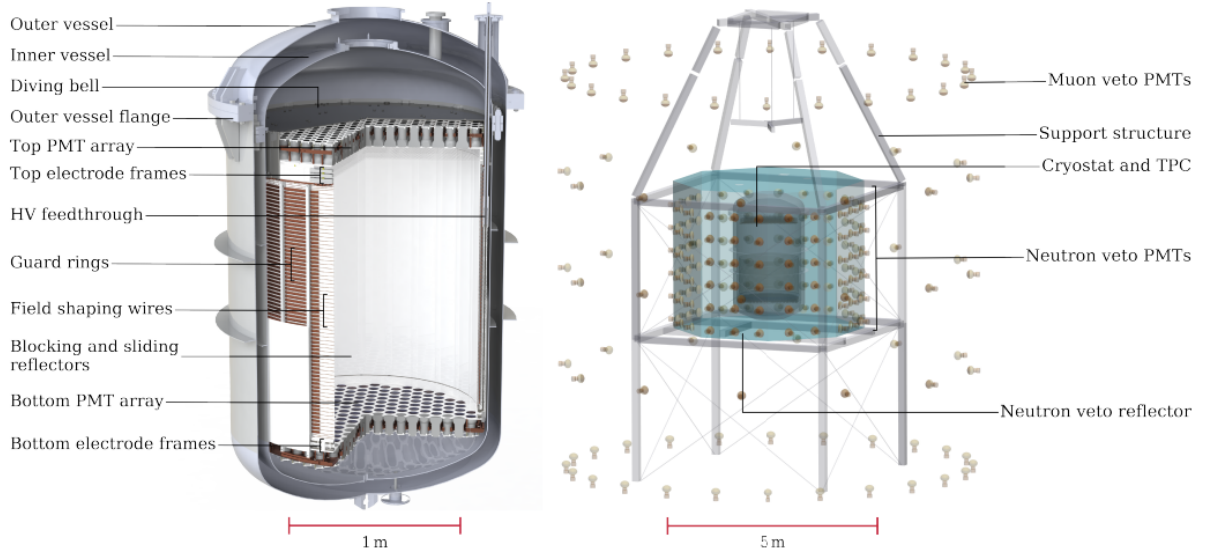


Figure 1: XENONnT Time Projection Chamber

With the collection of S1 and S2 signals, the interaction between xenon atoms and incoming particles can be classified into electronic recoil (ER) or nuclear recoil (NR) events. Electronic recoil events occur when particles with less energy interact mainly with xenon's atomic shell, lacking the energy needed to interact with the nucleus significantly. An event classified as nuclear recoil occurs when the particle has enough energy to not only penetrate xenon's atomic shell but also interact with its nucleus. Most background events such as neutrinos, beta decays, and gamma particles are expected

to be ER events, while WIMPs and neutrons would typically be detected in NR event regions. These signals produce an ER-NR band of signal events which help distinguish signal from background and place upper and lower limits on WIMP detection regions. The neutron water tank veto addresses neutron events that lie in WIMP energy ranges, thus distinguishing WIMP and neutron events [3].

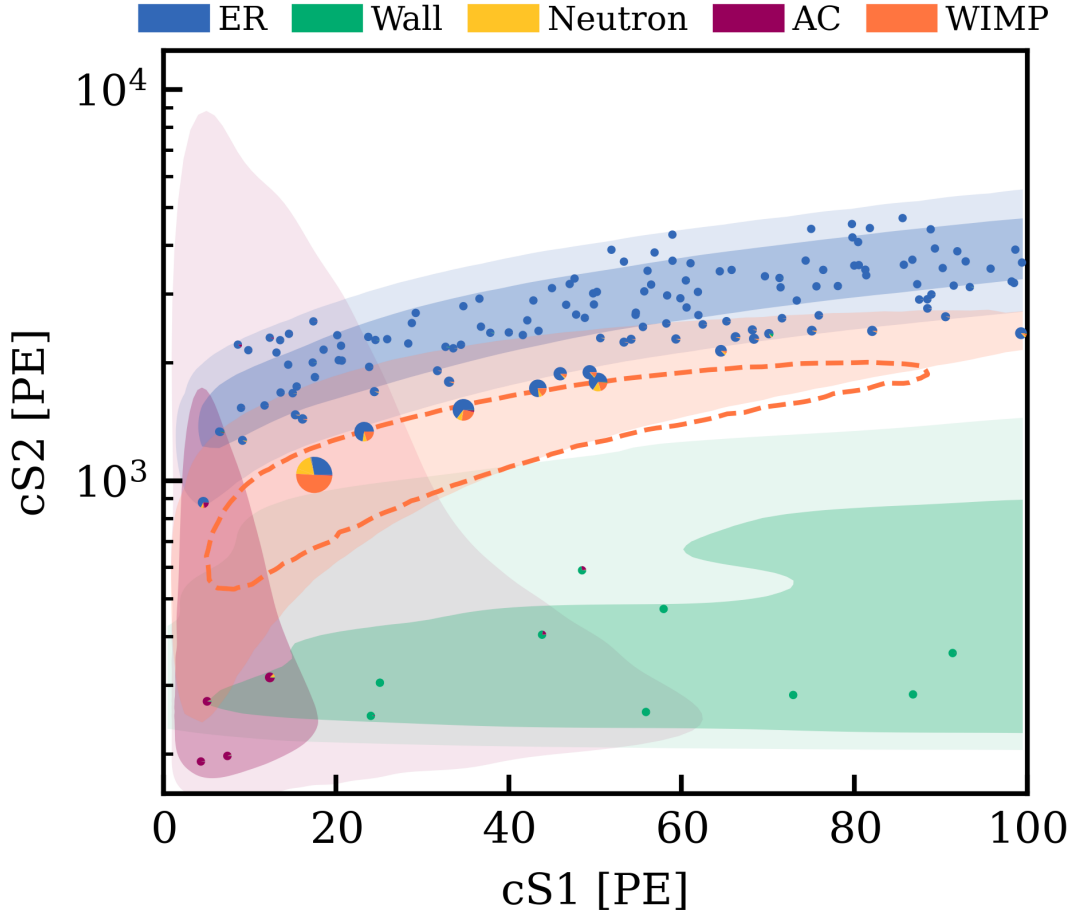


Figure 2: ER-NR band and sources of background.

2 Signal and Position Reconstruction

The reconstruction of light and charge signals is done via *straXEN*, software designed by XENON that classifies and groups S1-S2 peaks given the time of signal collection, width, and the area of photoelectrons received by the PMTs[8]. At the event level, data selection cuts are applied to remove background events that do not fit within the ER-NR region [5].

2.1 Event Building Process

The event construction process produces a three-dimensional position and energy reconstruction by initially analyzing data at the peak level. Since raw data is received at the peak level, light and charge peak-level data must be paired properly to ensure correct event data.

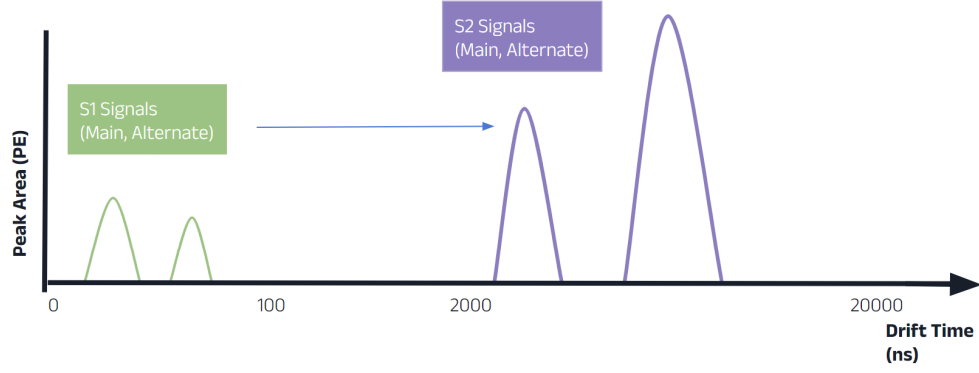


Figure 3: Visualization of *S1* and *S2* main and alternate peaks. Main peaks are used to reconstruct events.

Grouping data into 700 nanosecond time intervals, data from peaks are recorded and sorted into their respective *S1* or *S2* categories. Data recorded includes number of photoelectrons, electron drift time, and number photomultiplier tubes that collected photoelectrons. StraXEN uses the photoelectron count to pair *S1* and *S2* peaks with the largest photoelectron (PE) area and suitable fraction of photoelectrons collected in the top PMTs to form an event.

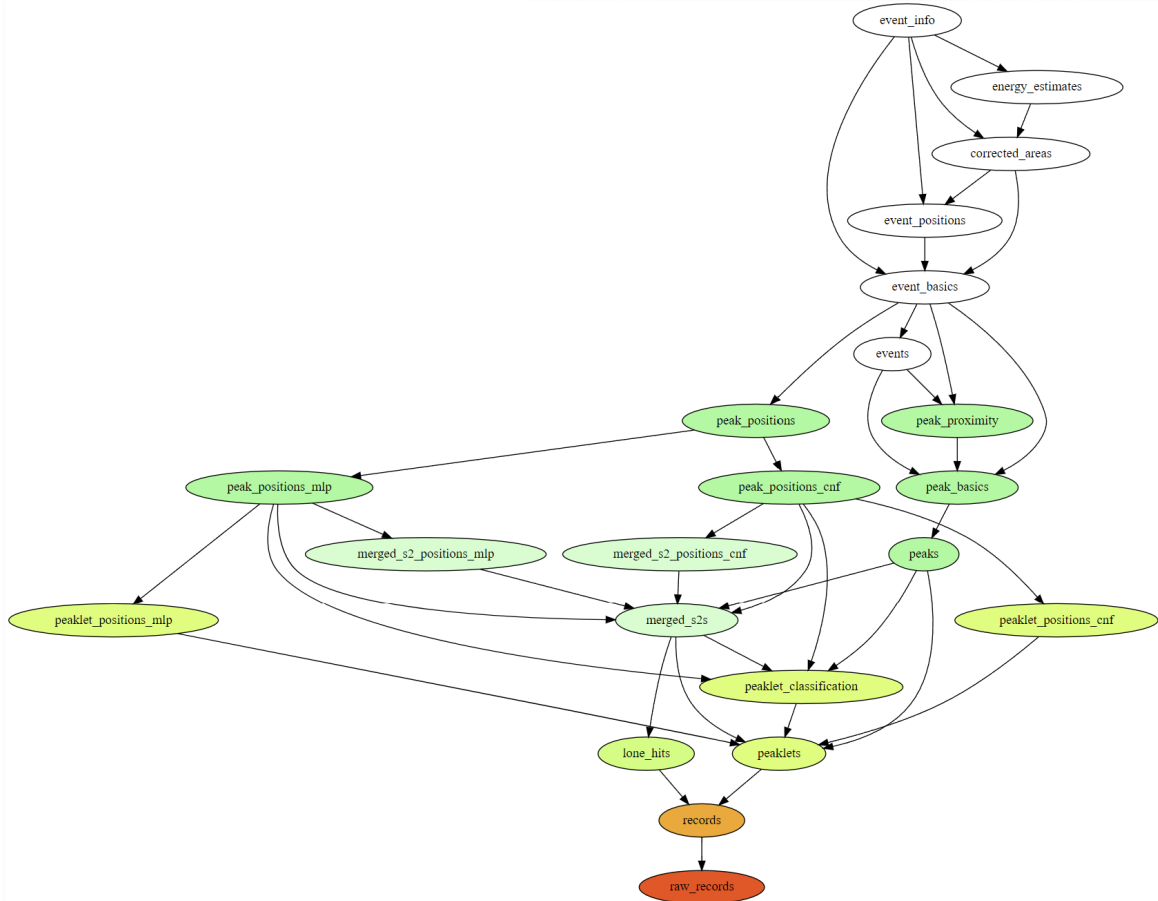


Figure 4: Flowchart of data construction. Peaks are paired to form event-level data.

Given that both peaks have sufficient area and that S2 charge signal peaks occur within a physical time range, the event is reconstructed given the peaks with the largest number of photoelectrons captured. This robust process parses the large amount of data collected to rapidly produce events from peak-level data[5].

2.2 Background and Noise Reduction

At any given time the XENONnT TPC may receive signals from many interactions. The Cherenkov tank muon veto and neutron veto shield the detector from some background events. Other background events beyond this originate from noise, multiple interactions, or electronic interference. These interactions are not usable as the peaks are produced from background events, have experienced interference, or interact outside the fiducial volume of xenon. Even signals from physical events may coincide with noise events or other isolated peaks; such cases may lead to a mispaired event composed of peaks from two different events that do not reconstruct a real event[1]. When such interactions occur and are reconstructed, the resulting false event likely will lie beyond the ER-NR band, and thus as a background signal, it is not considered for further analysis. To differentiate signals, data selections are utilized to source background to analyze signal events.

Data received by Photomultiplier tubes can additionally be affected by quantum fluctuations near the PMTs, radiation in the detector walls, and electronegative impurities in the detector. Electronegative impurities are caused by decays or applied electric fields from the detector walls and electrode gates, and this may affect the drift and number of ionized electrons. Such artifacts also produce accidental coincidence (AC) background signals. AC signals are events that the detector reconstructs that do not represent a real interaction. This is due to poor event reconstruction caused by the mispairing of isolated or mislabeled S1 and S2 signals[3]. The result is an incorrect combination of S1 and S2 signals that creates a nonphysical event.

2.3 Data Selection Acceptance and Cuts

Particle interactions and detector variables are evaluated by twenty-six data selections to source data and reject non-signal-like events or events we cannot yet model. The fiducial volume data selection accepts events within the innermost 4.2 tonnes of LXe to account for impurities near detector walls. S1 width data measures the spread of S1 photoelectrons across the PMT arrays. This spread is cross-referenced with S1 area to identify false S1 peaks. S2 width data measures the spread of the photoelectrons to ensure that the drift time of photoelectrons is physically possible. S1 and S2 single scatter data selection relies on photoelectron drift time and data acquisition time to ensure that the event only underwent a single scattering upon its collision with xenon. The area fraction top (AFT) of S2 charge signals validates that the signal is an S2 peak. Each data selection has a certain acceptance threshold that statistically accepts most events that represent a physical interaction and are not affected by interference. The sourcing and rejection of background events place stronger limits on electronic recoil (ER) and nuclear recoil (NR) events.

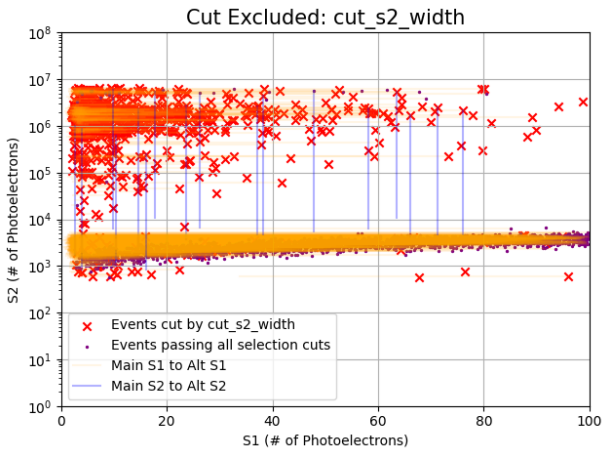
When peaks do not represent physical interactions, they are removed via data selection cuts. This is done by eliminating events that lie outside the threshold for any of the modeled data selections. These removed events typically do not match any genuine energy depositions that physically could occur in the XENONnT TPC, and as such are not considered in the final stage of event-level data analysis. In some cases, the removed events are genuine, but the probability of such rejections is quantified with the corresponding cut acceptance threshold.

3 Development of Event Builder Pairing Algorithm

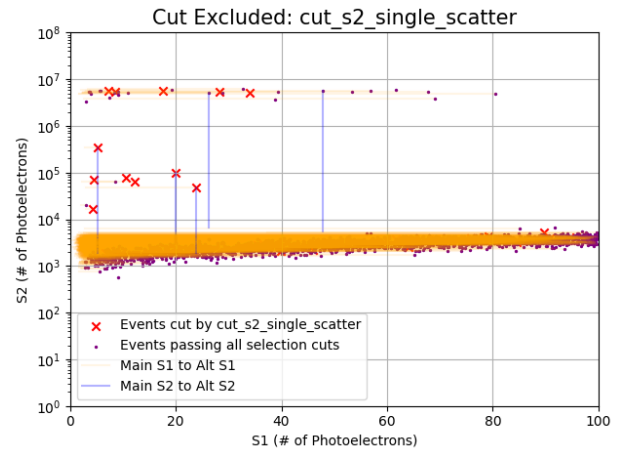
This project focuses on developing a high-performance event builder that takes into account parameters beyond the photoelectron area. An analysis was done to understand the current performance of the event builder and to optimize its selection process via a goodness-of-fit-based algorithm. Radon-220 calibration data was used to inform this study. ^{220}Rn produces low-energy electronic recoil events which is useful in identifying the region of ER background events and differentiating between NR signal events, ER background events, and non-physical AC and wall events.

3.1 Analysis of Current Event Builder

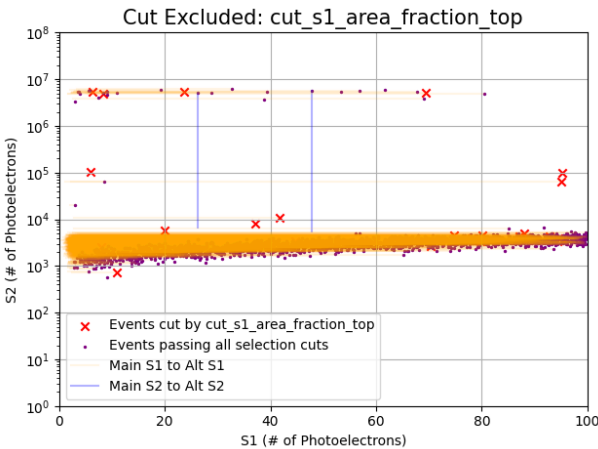
The current event reconstruction model generates events by pairing S1 and S2 peaks with the largest peak area. Primary and secondary peak data are recorded during event construction and these peaks are paired to form event-level data. After pairings form events, data selection cuts are applied to values that do not match a physical event. Additionally, event-level data that uses alternate S1 / S2 peaks during the peak pairing process is not utilized in event data analysis. An initial study of the event builder was conducted to understand the current success rate of event pairing.



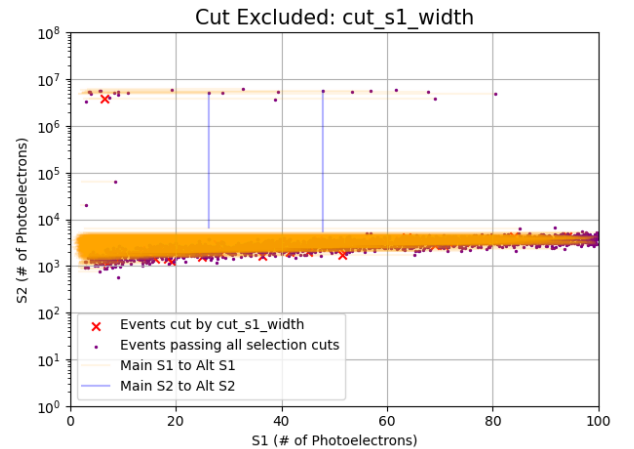
(a) All Cuts applied except S2 Width



(b) All Cuts applied except S2 Single Scatter



(c) All Cuts applied except S1 Area Fraction Top



(d) All Cuts applied except S1 Width

Figure 5: Data sets with primary and secondary S1 and S2 peaks shown. Each graph applies all data selection cuts but one to isolate the performance of the excluded selection.

This analysis aimed to graphically represent primary and secondary S1-S2 peaks and determine if and when alternate peak data is a better match for event reconstruction. Each graph removed a singular data selection cut to reveal the events the selection cut eliminated. Data sets with the S2 width and S2 single scatter selection cuts removed revealed mispaired background events in the AC background region with an alternate S2 pairing within the ER-NR band. The movement of such events into the band indicates that since some alternate S2 pairings seem to produce physical events upon reconstruction, such alternate S2 peaks are physical. Data sets where S1 data selection cuts such as S1 width and S1 area fraction top were removed did not yield significant results. This indicates that the source of alternate S1 peaks for most events comes from small, unphysical events.

3.2 P-Value Based Quartile Fit

To reduce the occurrence of mispaired events, we decided to create an algorithm that would reference all relevant data selections to improve the chance of finding the correct S1-S2 event peaks. This would optimize peak selection by searching for the peak with the best fit to the expected ranges of multiple data types. Implementing a goodness test that integrates data selections beyond the peak area can provide more precise evaluations of peaklet data and recover some mispaired and AC events.

The initial step to producing this algorithm was to find and graph data sets that were correlated. Various data selections may depend on the same observed quantities; due to this physical dependency, we can correlate data types that share the same dependencies. Such correlations model behaviors in the detector and allow us to generate a polynomial graph that rejects values outside 1 sigma of the median. We analyzed data from twenty runs of the SR0 ^{220}Rn calibration and correlated related data selections to produce quartile fit models. Graphs with strong correlations were further analyzed to understand the physical process behind the correlation and how it contributes to saving physical events.

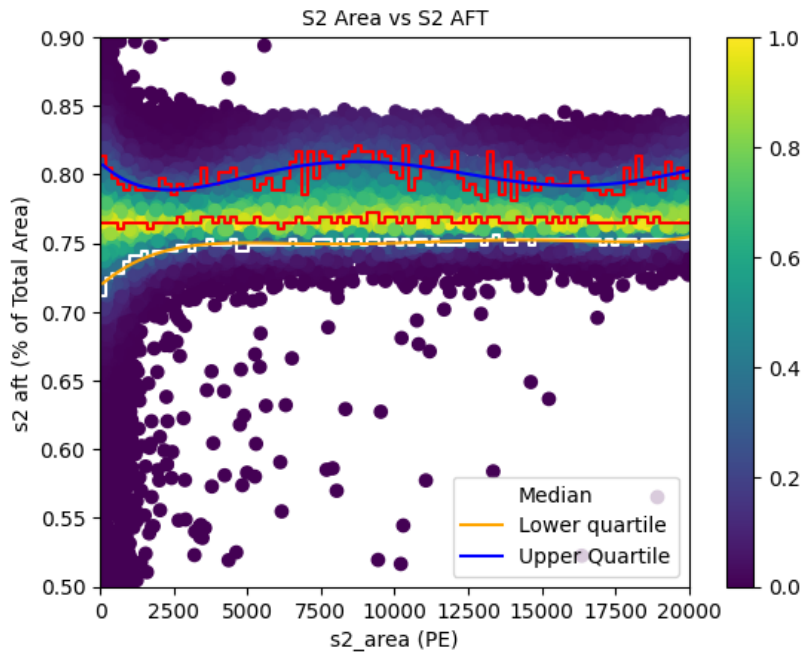


Figure 6: Polynomial graph of S2 area versus S2 AFT. Accepted events should exhibit S2 peaklet data that lies within $\pm 1\sigma$ of the expected value.

A notable correlation occurs between the S2 peak area and the percentage of S2 photoelectrons hitting the top PMT array (S2 area fraction top). At very small S2 areas, there is a large range of S2 AFT areas, which aligns with quantum fluctuations observed in smaller energy ranges. As the photoelectron area increases, the S2 AFT value remains steady, indicating that there is a relatively stable median percentage of photoelectrons hitting the top PMT array for larger photoelectron areas. There exists a median percentage for all energies, but at lower energies, the scatter around the median is large and thereby not as strongly correlated. This correlation places stronger limits on S2 peaklet identification and rejects abnormal events such as gas events or events with misidentified S1-S2 peaks.

Electron drift time and fifty percentile area range of S2 PE also exhibit a strong correlation. Interactions closer to the top PMT array have a shorter drift time. Given that the interaction electron cloud has less time to spread, the width, represented by the 50p area range of S2 photoelectrons, will have a lower value. Interactions that occur towards the bottom of the TPC will experience a greater drift time, and thus more spread in the electron cloud as it drifts upwards. Modeled by a polynomial graph, this correlation places limits on the spread of photoelectrons given the drift time of the electron cloud. This constraint on interaction width removes gas events, AC background, and events with unphysical drift times.

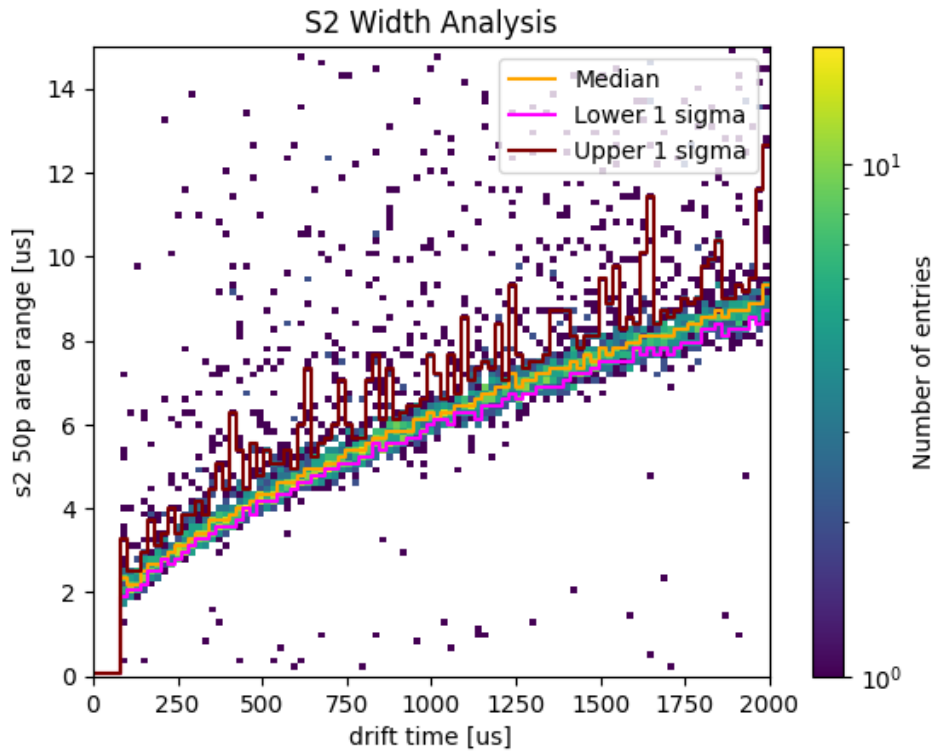
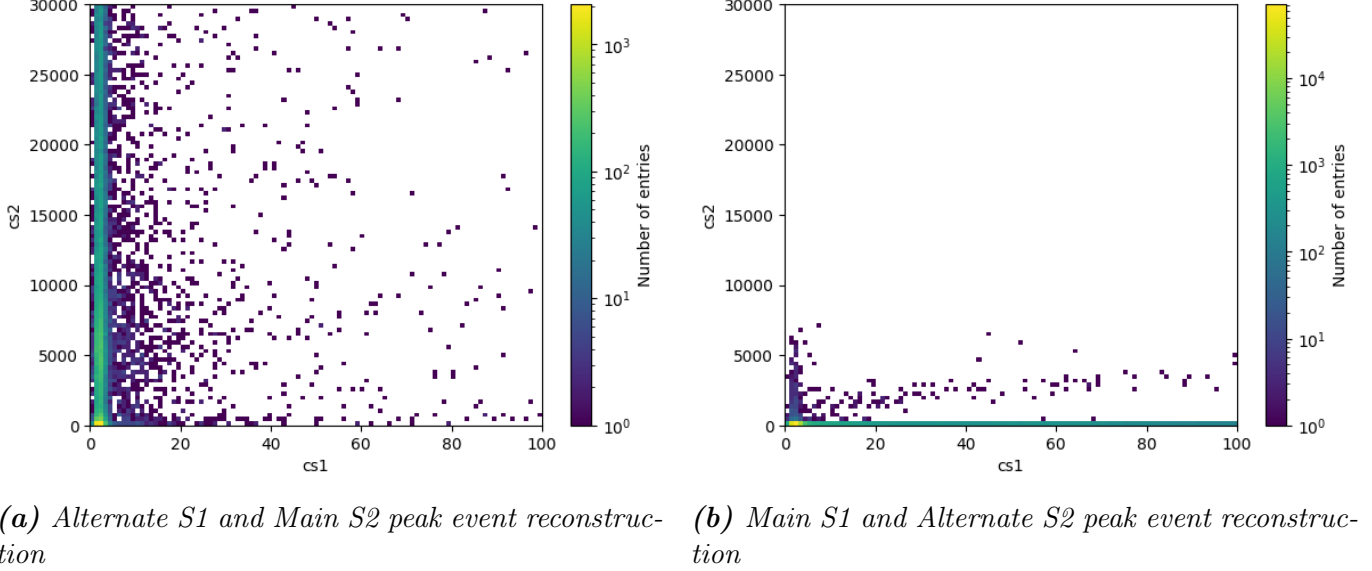


Figure 7: Polynomial graph of drift time versus S2 50-percentile area range to model width of interaction as a function of time. Accepted events must contain width values within $\pm 1\sigma$ of the expected value given the drift time of event interaction.

3.3 Density of Saved Mismatched Events

A test was conducted to observe the number of mismatched events that are reconstructed into physical events when alternate peak data is matched with main peak data. The purpose of this test was to identify where mismatches occur most frequently and observe the success of alternate peak data. For this test, a large sample set of 173 runs were utilized each containing 300 events. All available variations of peak pairing were represented graphically.



Main S1 peak and alternate S2 peak event data yielded some reconstructed events within the ER-NR band at all energy ranges. A sufficient number of events with main S1-alternate S2 pairings are properly constructed. Main S2 peak and alternate S1 peak event data output a small density of events in the ER-NR band at low energies. There is little statistical significance of this density due to the large signal-to-noise ratio of events reconstructed at low PE areas.

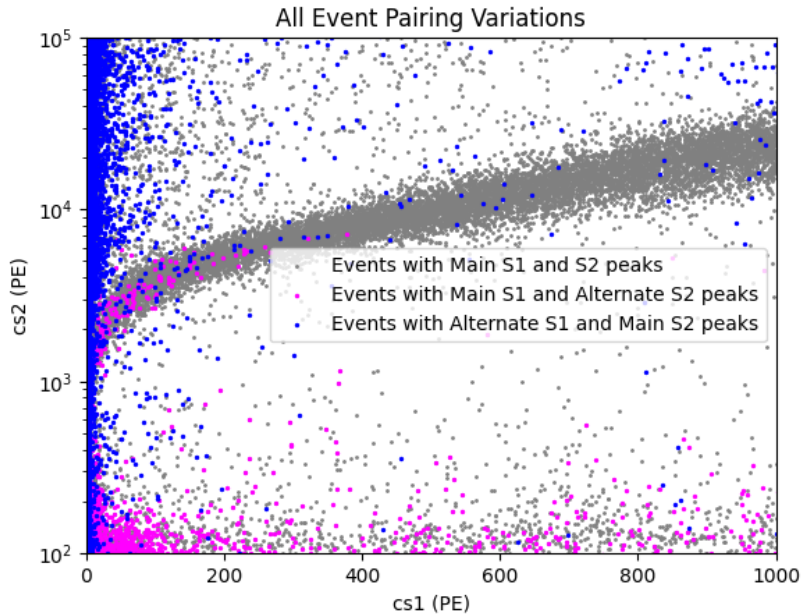


Figure 9: All Main and Alternate Pairing Variations to form Events

Condensing all data into one graph, the most dense regions of events with alternate pairings occur in regions corresponding to accidental coincidence events.

4 Results and Discussion

The analysis of the current event-building model found that some events with alternate S2 peaks at lower energies were a better match to the primary S1 peak than the main S2 peak with a higher energy. Data from the S2 width and S2 single scatter graphs suggests that some events that were removed given an unfit main S2 peak area contained a secondary S2 peak signal that matched the expected values of a physical event. Such mispairings may have been caused by the interference of simultaneous high-energy events like high-energy gamma particles. The interference by a higher-energy event obstructs the lower-energy interaction by producing a higher-energy peak that leads to a false mispairing of peak data. This indicates that peak area is not a comprehensive criterion for event selection and that a stronger algorithm is needed to account for interference caused by simultaneous higher-energy events. This algorithm should analyze multiple parameters to find the S1 and S2 peaks with the best fit to the expected data.

Initial results show that a p-value-based algorithm would be an effective method to accept signals with a close fit to expected data. Strong correlations between data selections are useful in developing a p-value-based algorithm. The next stage of the project focused on finding correlations that would prove most beneficial in the algorithm. Comparing S2 AFT vs S2 area provided a strong correlation and would therefore be a good function to include in the final algorithm. The constant expectation value of 0.78 is acceptable since we should expect most of the S2 photoelectrons to drift upwards given the induced electric field. S2 peaks that fall between the 16th percentile and 84th percentile are analyzed and the peak with the best goodness of fit to the expected value of 0.78 will receive the lowest p-value for this function. Thus, for a physical event, the S2 peak with a large number of photoelectrons must also closely match expected S2 AFT values to move forward in event reconstruction. The direct proportionality of drift time to the 50 percentile area range for S2 peaks also displayed a strong correlation. Peaks that align with the expected values of this graph fit the behavior of physical events in the detector, and the S2 peak with the closest value to expectation will receive the lowest p-value.

Measurement data of event densities given primary and secondary peak pairings provided statistically significant results for alternate S2 peak data. The event data yielded a low density of saved events, but this could be attributed to the usage of event data from ^{220}Rn calibration runs, which produces low-energy ER events. Even with this, a sufficient number of events that were reconstructed with main S1 and alternate S2 peaks fell in the ER-NR band. This data suggests that the second largest area peak for such events may have been the true physical peak of the event and that the primary peak was misidentified from another interaction. S2 peaks with high energies may originate from high-energy events that coincide and interfere with the low-energy event corresponding to the main S1 peak. This mispairing would be reduced with the updated algorithm as the algorithm would also consider goodness of fit to drift time and electron dispersion. This reduces the chance of a misidentified S2 peak being chosen as the main S2 peak for event reconstruction. The algorithm may also save more mispaired events from other data types not utilized in this study such as NR and high-energy ER events. This finding also suggests that the pairing algorithm may aid in reducing accidental coincidence events. AC background comprises falsely formed events of mispaired S1 and S2 peaks. Often the misidentified S1 or S2 peak causing the false event is a noise peak. As such, noise

peaks that have an alternate signal peak move events from the AC background to the ER-NR event region. Implementation of the updated pairing algorithm has the potential to reduce the pairing of signal and noise peaks, which would overall decrease the likelihood of noise from wall events and multiple scatters.

The algorithm can be designed using best fit polynomial functions that model the distribution of each correlated pair. The goodness of fit is given by calculating the proximity of measured values to the expectation value of the function, and this is achieved by setting the p-value of the expectation to 0 and recording the deviation from the p-value. Multiple best fit functions, including S2 area-S2 AFT and drift time-S2 50p area, can be compounded to generate a goodness of fit algorithm that determines the best S1-S2 peak pairing. Individual functions output p-values for all S1 and S2 peaks, and the algorithm outputs the peak with the lowest p-value as the best fit peak. The best S1 and S2 peaks are paired to form an event that has underwent a comprehensive selection criterion and thereby has a lower change of containing mispaired peak data. This updated event builder pairing algorithm eliminates the need for data selection cuts as the algorithm already performs this as part of the optimal peak selection process.

5 Next Steps

The next immediate step would be to develop functions based on the p-value fitting that will be used by the algorithm. The final step in the process is to compound multiple p-value functions to output a best fit S1 and S2 peak.

An important test to generate more correlated data selections would be to test event-level data taken from Krypton-83m calibration data. ^{83m}Kr models detector stability and drift length corrections. Analyzing S1 and S2 peaks from this data will introduce events that undergo a higher rate of electron cloud diffusion at deeper interactions. The resulting high event rate leads to a large pile-up of event data, and a physical pairing of events may resolve this issue.

The final study showed that alternate peak data may prove useful in reducing the AC background. Following the algorithm development for AC reduction will improve the selectivity of nuclear recoil events. Additionally, reducing the AC background with the algorithm will allow for higher signal acceptance for ER events as more events will be sourced as real background events instead of incorrectly paired false background events.

Future implementation of this algorithm requires understanding potential biases that might arise with the selection criteria acceptance. The goal of this algorithm is to optimize the process of proper event reconstruction, but this runs a potential risk of selection bias towards peaks of expected values. This is because a strict constraint of reduction of bad data at the peak level may remove bad S1-S2 peak data caused by detector issues or fluctuations. If another peak with a better fit occurs within the same time frame, the algorithm will favor the best-fit peak, and issues with the detector may go unnoticed. To mitigate this issue, the algorithm should be well-modeled and unbiased, which can be accomplished by adding a relevant correlation function or modifying the algorithm acceptance limit.

To model the event reconstruction of main and alternate signal pairings, I modified the code designed to reconstruct peak-level data, `event_basics.py`. `EventBasicsModifiedS1` and `EventBasicsModifiedS2` are modified versions of `event_basics` that model main/alternate signal pairings. `EventBasicsModified` contains a method to develop the event pairing algorithm. These programs alongside this report can be accessed via GitHub by clicking **this link**.

6 Acknowledgements

I would like to thank Dr. Knut Morå, Shenyang Shi, Dr. Michael Murra, Professor Elena Aprile, and the XENON group for their wonderful guidance and support this summer. Thank you further to Professor Georgia Karagiorgi, Professor Reshmi Mukherjee, and Amy Garwood for hosting the REU program. This material is based upon work supported by the National Science Foundation under Grant No. PHY-2349438.

References

- [1] Projected wimp sensitivity of the xenonnt dark matter experiment. *Journal of Cosmology and Astroparticle Physics*, 2020(11):031, nov 2020.
- [2] E. Aprile and T. Doke. Liquid xenon detectors for particle physics and astrophysics. *Reviews of Modern Physics*, 82(3):2053–2097, July 2010.
- [3] E. Aprile et al. Xenon1t dark matter data analysis: Signal reconstruction, calibration, and event selection. *Physical Review D*, 100(5), September 2019.
- [4] E. Aprile et al. First dark matter search with nuclear recoils from the xenonnt experiment. *Phys. Rev. Lett.*, 131:041003, Jul 2023.
- [5] XENON Collaboration and E. Aprile et al. Xenonnt analysis: Signal reconstruction, calibration and event selection, 2024.
- [6] G. G. Raffelt. Dark matter: Motivation, candidates and searches, 1997.
- [7] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. Wimp dark matter candidates and searches—current status and future prospects. *Reports on Progress in Physics*, 81(6):066201, May 2018.
- [8] XENON. Straxen datastructure.