university of groningen

faculty of science and engineering

Modelling the delayed single and few electron background in XENONnT

Author:
Keerthana UMESH
(s4767969)

Supervisor:
Dr. Jelle AALBERS
Second examiner :
Dr. Mick MULDER

Bachelor's Thesis
To fulfill the requirements for the degree of
Bachelor of Science in Physics
at the University of Groningen

July 12, 2024

Abstract

The search for dark matter has become a central focus in the field of particle physics and cosmology, driven by a wealth of evidence that dark matter constitutes a significant portion of the universe's mass-energy content. Traditional dark matter detection experiments, designed to identify Weakly Interacting Massive Particles (WIMPs), have yet to yield definitive results, prompting the exploration of alternative candidates in the low mass regime. The XENONnT experiment aims to detect dark matter particles in this regime using a liquid xenon target within a dual-phase time projection chamber that is capable of reconstructing the 3D position and energy of particle interactions. However, a persistent challenge in this search is the precise characterization and mitigation of background signals that could obscure potential dark matter interactions. Dual-phase time projection chambers exhibit a persistent rate of signals that correspond to single and few electron emissions in the detector. Lacking a thorough understanding of their phenomenology, these limit our ability to distinguish between a small signal background vs. a small signal from low mass dark matter. In this thesis, some preliminary steps are taken to model these single and few electron signals using previous characterization studies in similar detectors. A novel approach is proposed, where these emissions are modelled as a function of all preceding energetic interactions in the detector. This model extends to longer (i.e. delayed) time ranges than previous attempts, and shows promise in being an effective background model with some further development, thereby enhancing the sensitivity of low mass dark matter searches with XENONnT.

Contents

List of Figures

List of Abbreviations

CMB Cosmic Microwave Background. 9

DM Dark Matter. 9

DS Double scatter. 17

ER Electronic Recoils. 15

GXe Gaseous Xenon. 16

LXe Liquid Xenon. 15

MLE Maximum Likelihood estimation. 22

NR Nuclear Recoils. 15

PDF Probability Density function. 22

PMT Photomultiplier Tube. 16

S1 Primary scintillation signal in XENONnT. 16

S2 Secondary scintillation signal from ionized electrons in XENONnT. 16

SE Refers jointly to small S2 signals that corresponds to single and few electrons. 18

SM Standard Model. 12

TPC (Dual-phase) Time Projection Chamber. 15

WIMP Weakly Interacting Massive Particle. 12

1 Introduction

The Standard Model of particle physics is a remarkably successful framework that can accurately describe the fundamental particles and forces governing our Universe. It has withstood rigorous experimental testing and has been instrumental in explaining a wide range of phenomena. There is, however, an overwhelming body of astrophysical evidence that up to 85% of the matter that makes up the Universe is some unknown, unseen particle from beyond the Standard Model. The existence of this "dark matter" can only be inferred through gravitational effects, but it has been instrumental in the large-scale evolution in our Universe [1].

Despite its abundance, dark matter interacts so weakly with "ordinary matter" that we are yet to have an unambiguous detection of a dark matter particle. Earth-based direct detection searches involve studying the scatter of the potential dark matter candidates(s) off some target nuclei and transfer some energy which is recorded as a scintillation, ionization or heat signal [2]. From these signals, expected scatters from known particles (i.e. the background in the search) can be subtracted. Direct detection experiments look for some excess of leftover signals that would point to scatters from some beyond the Standard Model particles. These can then be compared against predictions from different theorized dark matter masses and interaction cross-sections. Dark matter interactions with nuclei are incredible rare, and thus, direct detection require extensive background shielding and removal methods to identify any excess [3].

Of the direct detection strategies that currently exist, dual-phase time projection chambers that use liquid xenon as the scattering medium have emerged as leading candidates in the search for dark matter [4]. Incident particles recoil off of the xenon nuclei, exciting and ionizing them. From these scintillation and ionization signals, the 3D position and energy of the particle can be reconstructed and used to discriminate between known backgrounds and unknown potential dark matter signals. Such time projection chambers are primarily sensitive to searches for dark matter with masses in the $>$GeV/c^2 range. Liquid xenon time projection chambers have been operational in this range for three decades already, and have managed set stringent limits that exclude large regions in parameter space for such masses [3][5].

In the past few years, efforts have instead shifted towards light dark mattter in the sub-GeV range [6]. Light dark matter interactions with liquid xenon would result in very small energy transfers, corresponding to only a single or few ionized electrons. Dual-phase time projection chambers however, are capable of sensing signals down to this level- i.e. they are theoretically sensitive enough to potentially find low mass dark matter interactions [4].

The challenge with probing lower mass regimes is that the number of instrumental and radioactive backgrounds increases at small energies [6]. In particular, signals corresponding to single and few electron emissions are observed to be especially elevated after a large energy deposition in the detector and persist for O(100 ms) after. These emissions seem to be ubiquitous in all liquid xenon time project chambers. While several hypotheses exist, we are lacking an understanding of the physical mechanism(s) that drive these emissions at time scales beyond a few milliseconds after an energetic interaction in the detector. The emissions seemingly have no characteristic timescale, which means that we cannot distinguish between the "delayed" single and few electron signals that *may* still be some background from the previous energetic interaction, and potential low mass dark matter signatures [5][7][9].

Over the past few years, significant effort has been dedicated to characterizing the dependence of these delayed emissions on preceding energetic interactions in an effort to understand their phenomenology. The underlying still mechanisms remain inconclusive, but they have found certain correlations between the rates of these small signals and the size, time and position of large interactions in the detector directly preceding them [7][8][9]. These were then used these to identify and remove background single and few electron emissions. This approach is limited to specific time scales after each large interaction, and cannot be extended to distinguish single and few electrons occurring several seconds after. That is, a comprehensive background model that can capture the full range of these emissions has not yet been developed, which means that any excess observed in this low energy limit cannot be correctly distinguished from potential background emissions. This constrains low mass dark matter searches to only being able to set upper limits, rather than being able to identifying definitive dark matter signals [8].

In this thesis, these characterization studies are leveraged and some preliminary steps are taken towards modelling this delayed single and few electron background for data in XENONnT, the latest liquid xenon time projection chamber to join the hunt. XENONnT is a promising player in the low mass dark matter frontier as one of the largest liquid xenon detectors in the world with a background five times lower than its predecessor [10]. A novel approach where the single and few electron emissions are modelled as a function of *all* preceding large energy depositions in the detector is presented. This background model shows some promise in being able to capture the full extent of the delayed electron emissions and thereby improve our ability to probe lower mass limits of dark matter.

Moving forwards, several of the terms mentioned above are abbreviated- a reference of these can be found on page 3. The sections become increasingly narrower in scope- Chapter 2 begins with a general introduction to the evidence for dark matter, some salient dark matter models for this investigation, and the premise behind direct detection. Chapter 3 talks specifically of direct detection with XENONnT, and discusses how the experiment mitigates its backgrounds and conducts low mass dark matter. Chapter 4 summarizes the single and few electron backgrounds limiting these searches and raises some key findings from the studies characterizing them. In Chapter 5, a general approach on how these findings drove the modelling process is outlined. In Chapter 6, three potential models are presented in increasing order of complexity, with the final one showing a new approach that solves previous limitations. Chapter 7 summarizes these findings and provides some further ways to improve the model.

2 Dark matter

2.1 Evidence for dark matter

The evidence pointing to some volume of unseen matter in our Universe is wide and varied, extending from the galactic scale (primarily from the mass-to-light ratios of various astronomical objects) to the cosmological scale (primarily from the Cosmic Microwave Background (CMB). Here, they are briefly summarized- a full review can be found in [1] and [11].

One of the first observations came in 1933, when Fritz Zwicky applied the virial theorem to the Coma Cluster and concluded that the galaxies within it were moving too fast to be gravitationally bound by their visible matter alone. He termed this extra mass holding them together as "dunkle Materie", i.e. dark matter (DM) [12]. This mass defect continues to be observed even under more recent X-ray imaging and modelling that better accounts for the hot gas within the cluster, with DM needing to make up some 85% of the total mass [1].

A mismatch between motion and gravitational bounds can also be found at the galactic level. Spiral galaxies are stable, bound systems, with young stars and gas mostly rotating in a plane around the center. Assuming no external perturbations, the gravitational pull should equal the centripetal force, from which we get Kepler's Law [11]:

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{1}$$

where v(r) is the circular velocity profile, r the radius, i.e. the distance from the center, M(r) the mass profile w.r.t. r and G is the gravitational constant. At radii enclosing the bulk of the mass content of the galaxy, we expect to see v(r) fall off $\propto r^{-\frac{1}{2}}$. From radio observations of the 21 cm spectral line (i.e. the 1s hyperfine transition of neutral hydrogen), we can probe rotational velocities to large radii beyond the optical extent of the galaxy [1]. Surprisingly, the rotational velocity curve at these large radii flatten out and stay constant in nearly all observed spiral galaxies, counter to the expected Keplerian fall. Fig 1 shows this observed rotation curve for M33. That is, the mass from the stellar and gas bulk in the center is not the only contributing mass. These flat rotation curves can be explained by accounting for some "halo" of DM dominating the outskirts of spiral galaxies [1].

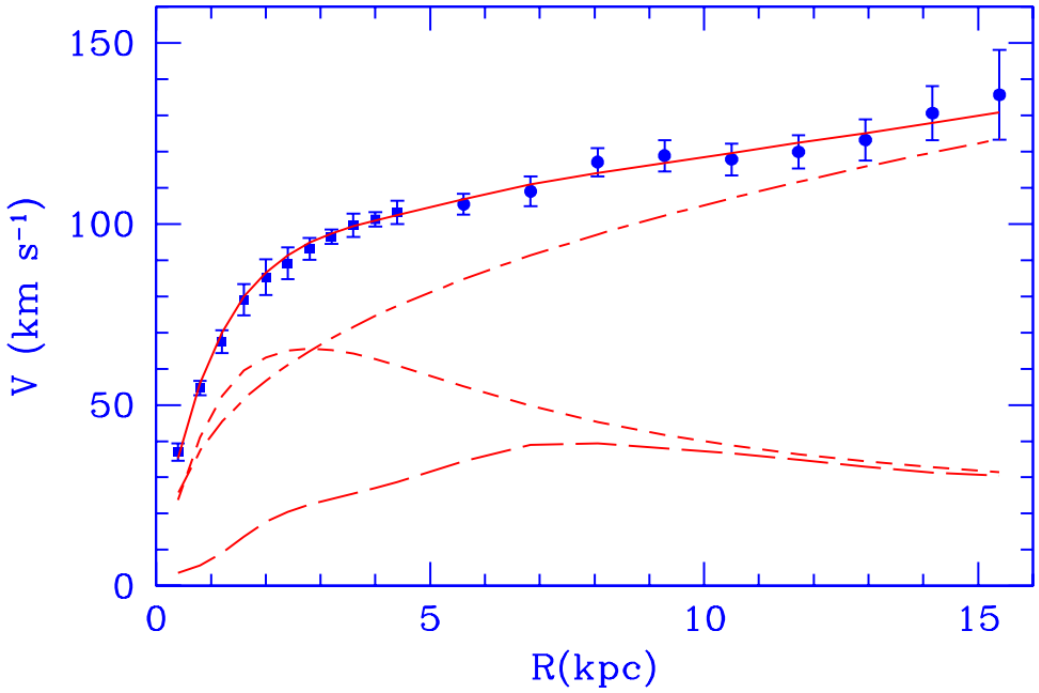

Figure 1: Scatter points showing the observed rotation curve of the M33 galaxy. The DM halo contribution is given by the dashed dotted line, the stellar disk (i.e. the visible disk expectation) by the short dashed line and the gas contribution is indicated by the long dashed line. Measurements <5kpc come from starlight, while longer radii come from 21 cm hydrogen measurements. From [13]

This mass defect can also be observed independently of the proper motion of astronomical bodies, by looking instead at its effects on light. When light from a distance astronomical sources passes near a massive object, its trajectory is bent by the curvature of spacetime, similar to a lens. The degree of the focus/dispersion of light depends (alongside others) on the gravitational potential from by the lens. Thus from observations of this light distortion and the geometry of the system, we can unravel the mass distribution of the lensing object [1][14].

Fig 2 shows the Bullet Cluster, one of the most well-studied interacting systems where two galaxy clusters collided, passed through each other and are currently moving apart. By studying the gravitational lensing around this system, the gravitational potentials (and therefore the center of mass) is shown by the white contours. This is in clear offset from the center of the hot gas (in pink), which is thought to make up the bulk of the visible baryonic matter. The putative DM component (in purple) simply passed through the interacting clusters along with the diffuse stellar components, while the hot gas interacting electromagnetically was halted in its path [11].

Figure 2: A composite image of the Bullet cluster from NASA's Chandra Observatory [15]. The white contours show the gravitational potential and center of mass of the system observed through gravitational lensing. X-ray imaging shows the hot gas in pink. Purple indicates where most of the matter is found. Stellar components from optical images are superimposed.

Finally, the CMB is the oldest radition we can measure, emitted after the Big Bang as the Universe transitioned from a hot plasma to being transparent to photons, i.e. it gives a picture of their last scattering. The CMB is almost perfectly isotropic, but measurements from the COBE, WMAP and PLANCK satellites show small non-uniformities of $\Delta T/T \sim 10^{-5}$K [11]. They show small regions of under and overdensities in the primordial Universe that were then magnified by its expansion. These primordial density fluctuations are thought be responsible for the "web-like" cosmic structure we observe today, of galaxy clusters and superclusters interspaced by voids as matter gravitated to regions of overdensity. Dark matter is largely pressureless due to its weak self-interaction, and thus was able to accumulate and "clump" together to amplify these density perturbations and provide gravitational wells where galaxies could form. A similar sort of evolution from a baryonic matter only model would require much larger density perturbations than currently measured in the CMB due to its repulsive pressure [1].

The CMB anisotropies are in excellent agreement with a Universe containing a "cold" (i.e. non-relativistic) DM component that enhances structure formation, known as the Λ_{CDM} model [2]. Under this paradigm, $\sim$26.6% of the energy content of the Universe is tied up in matter interacting only through the weak-scale or lower [11]. The observed abundances of primordial elements today provide an estimate of the baryonic density through the concept of Big Bang nucleosynthesis in the early Universe. These results align with estimates derived from the Cosmic Microwave Background (CMB), but both point to the fact that baryonic matter constitutes only 1/5th of the Universe's total matter content [14]. Though more in the theoretical regime, the anisotropies in the CMB is one of the most compelling pieces of evidence for the existence of dark matter.

2.2 Dark matter models

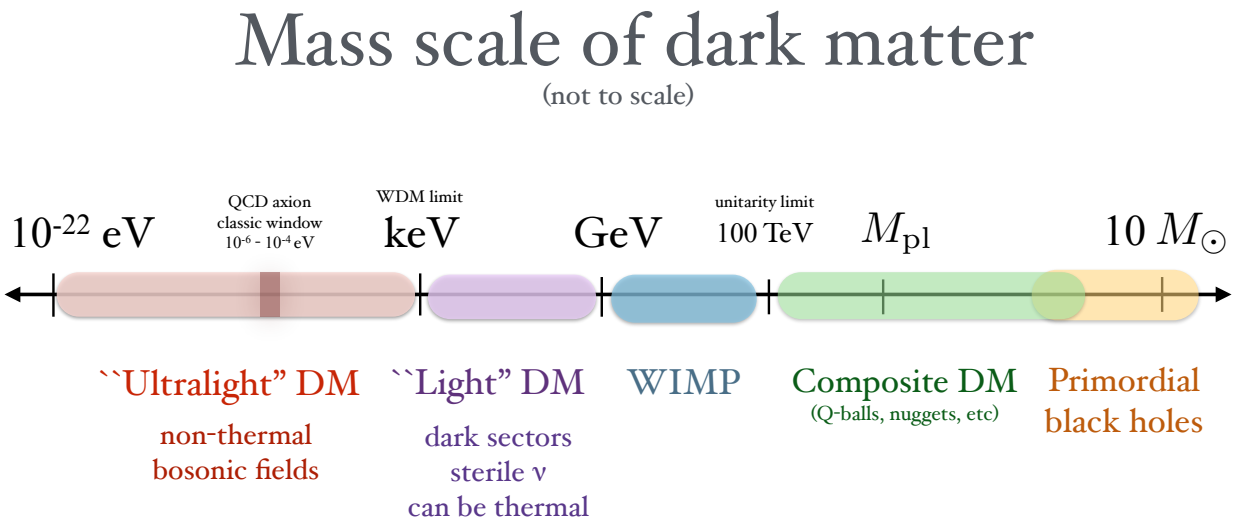

Figure 3: Mass range of potential dark matter models including limits from theoretical considerations. Includes not just particle candidates, but also bosonic and primordial black holes. Figure from [16]

We can summarize some key properties of DM from this evidence. It has to have some mass, but no charge or colour charge since it does not interact electromagnetically and is essentially collisionless. Its key role in structure formation and presence in the early Universe imply that it is non-baryonic, non-relativistic (to prevent large structures from "washing out" under fast-moving particles) and stable with timescales comparable to the age of the Universe [2].

No Standard Model (SM) particle fits these requirements. A number of theoretical models have been proposed- Fig 3 summarizes their mass regimes. Despite theoretical and observational constraints, they range across some 90 orders of magnitude in mass, with bosonic proposals on one end to macroscopic models (MACHOS) on the other that may account for the missing mass. Their general features and motivations are summarized in [16]. Several of these models are based on the idea of a "hidden" or dark weak sector of yet unobserved quantum fields that mediate DM and SM particle interactions in addition to gravity. Accordingly, their interaction cross section with "normal" matter also ranges across 20 orders of magnitude [3].

Much research and interest until now as been focused on Weakly Interacting Massive Particles (WIMPs), which lie in the region of ∼10GeV - 10TeV if we include supersymmetric extensions of the SM out of which they naturally arise [16]. WIMPs have masses and cross sections with SM particles typical for the weak interaction. A self-annhilation cross section of $\sim 3\text{x}10^{-26}$ cm^3/s and mass of $\sim$100 GeV/c^2 almost perfectly predicts the relic abundance of DM that we observe today- a coincidence that is often called the "Wimp Miracle" [2].

The focus in this report, however, is towards low mass dark matter models found in the keV-10GeV range (though typically simply referred to as "sub-GeV"). This is a natural conclusion given their weak coupling with SM particles [17]. To give the correct relic abundance of DM, they require

interaction cross-sections smaller than the weak scale, a constrain known as the Lee-Weinberg bound. Light DM interactions would therefore be mediated by some hidden sector forces [16].

2.3 Dark matter detection

To probe such a wide parameter space, there is also a diversity of detection methods. Broadly speaking, they can be summarized into three types based on the type of DM-SM interaction signatures they look for: (a) accelerator searches, (b) indirect detection and (c) direct detection [2].

In accelerator-based searches, e.g. with the ATLAS and CMS, high energy collisions of SM particles could result in the production of DM particles and their mediators. This can be inferred through some missing transverse energy and momentum "carried away" by these particles. Indirect detection- for example using the Fermi Gamma-Ray Space Telescope -instead searches for a surplus of SM particles like gamma rays and neutrinos in known regions of high dark matter density. These may be potentially caused by DM self-annihilation. The focus of this report is on earth-based direct detection which hopes to observe DM from its energy transfer upon scattering off some known target material [2].

The approach to earth-based direct detection methods can be summarized as follows. When potential DM particles recoil off of some target nucleus, they transfer some energy which can manifest as a light (i.e. scintillation), ionization or heat signal. From these signals, we can then subtract an expected background rate due to scatters from *known* particles and examine the remainder to see if it makes up some "statistically significant excess" (usually reported at the 90% confidence interval) [3]. The excess is an indicator of some new physics- i.e. potential dark matter -and can be compared against the expected rate of signals in the detector for various DM models. This is governed by the following (generalized, spin-independent) rate equation per unit energy [3]:

$$\frac{dR}{dE} = \frac{\rho_\chi A^2 \sigma F^2(E)}{2M_\chi \mu^2} \int_{v_{min}}^{v_{esc}} \frac{f(v, v_0)}{v} dv \otimes \mathscr{G}(E) \tag{2}$$

A, σ and F(E) depend on the target and are the number of nucleons, the cross section per nucleon and the nuclear form factor respectively. M_χ is the mass of the DM candidate and μ gives the reduced mass of the scattering particles. Meanwhile ρ_χ is the local dark matter density and f(v,v_0) is its (Maxwellian) velocity distribution integrated from v_{min}, the minimum detectable velocity in the experimental energy threshold to v_{esc}, the Milky Way escape velocity. $\mathscr{G}(E)$ is a token function that captures the resolution, quenching, acceptance etc. of the detector for different energy ranges and the two are convoluted together (represented by $\otimes$). The final signal rate is found by integrating this across the experimental energy window [3].

If no excess of signals is found, the results thereby exclude certain DM models in parameter space. Fig 4 summarizes these exclusion limits from a variety of different direct dark matter experiments, shown as a function of the proposed DM mass and its interaction cross section with a nucleus [18]. The curves on the graph trace out the spin-independent DM limit- everything above it has been probed with null results. The asymptotic nature of these curves come from the reduction in dark matter density at higher masses, and the region carved out in blue shows the neutrino "floor", where the rate of coherent scattering of neutrinos off of the target material becomes comparable to the DM scattering rate and limits measurements [4]. It is clear to see in Fig 4 that the sub-Gev range, i.e. low mass dark

matter is still relatively unexplored.

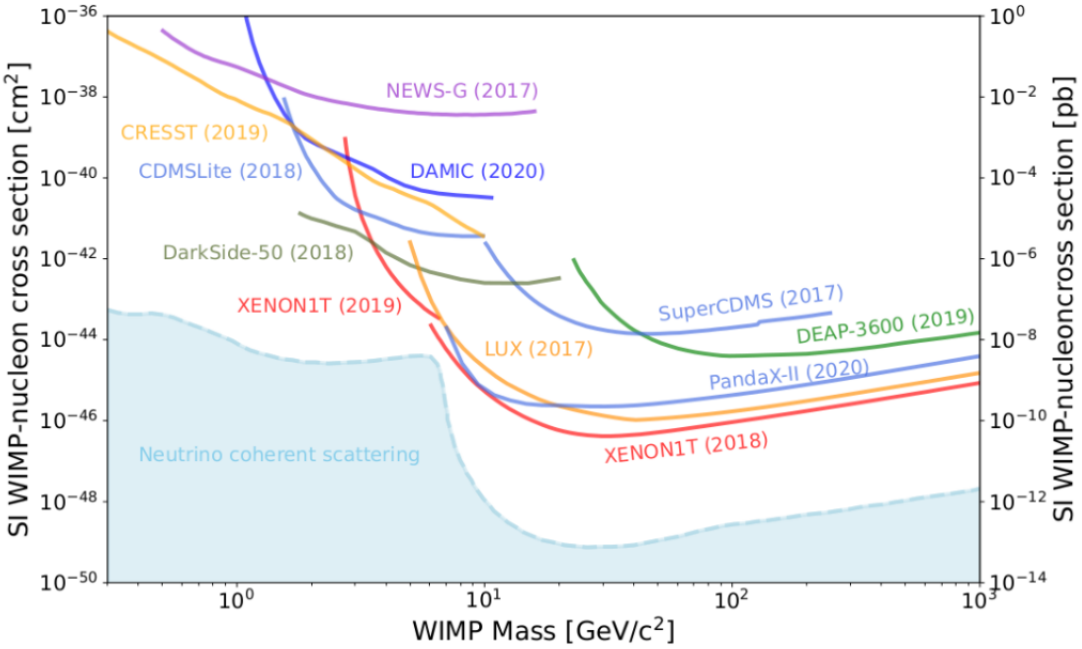

Figure 4: A summary of the exclusion limits set by various direct detection experiments as a function of WIMP mass and its interaction cross-section with a nucleon. The dotted lines are projected limits from upcoming experiments. The area obscured by neutrino scatters is indicated in light blue. From [18].

3 The XENONnT detector

3.1 The Liquid Xenon Dual-Phase Time Projection Chamber

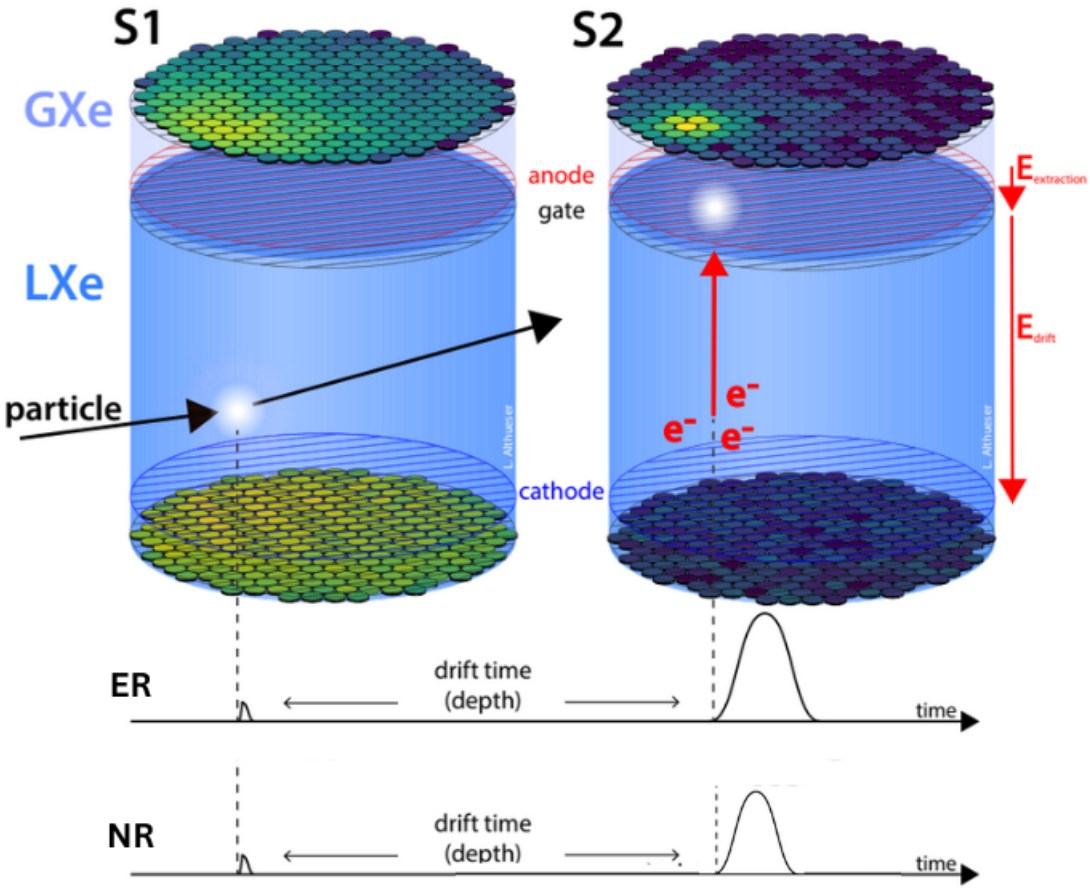

Figure 5: Working principle of the time projection chamber in XENONnT. The scintillation signal S1 is followed by an ionization signal S2 as the free electrons are extracted into the gaseous xenon. The ratio between the two is used to distinguish between an electronic recoil (ER) or nuclear recoil (NR). Figure adapted from [2], original image credits L. Althüser.

Located at the Laboratori Nazionali del Gran Sasso in Central Italy, The XENONnT experiment is one the largest earth-based direct DM detection experiments. It has 5.9 tonnes of liquid xenon (LXe) as its active target, and is a part of a class of detectors that uses a dual-phase time projection chamber (TPC). Dual-phase TPCs are currently the world's leading technology for the direct detection of WIMP dark matter [4][19]

The working principle of the TPC is visualized in Fig 5. When a particle enters the interaction medium, it will undergo either a nuclear recoil (NR) or electronic recoil (ER) interaction with the xenon atoms. The resulting transfer of energy excites and ionizes the LXe, which then combine with ground state xenon atoms to form meta-stable molecular states. When these states deexcite or re-

combine, they emit light with a mean wavelength of 178nm to which the *atomic* xenon is transparent [2]. The photons are thus able to travel through the interaction medium without interruption, and are captured by an array of photomultiplier tubes (PMTs) at the top and bottom of the detector. This produces a scintillation signal termed S1.

Meanwhile the ionized electrons that escape recombination are drifted upwards under a uniform electric field. The top of the detector contains gaseous xenon (GXe), and when they reach the liquid gas interface, the are accelerated under a stronger electric field and a second, proportional scintillation signal called S2 is produced through electroluminescence [20]. The hit pattern of the PMTs gives the x-y position of the interaction, while the z-coordinate, i.e. the depth of the interaction in the LXe column, is given by the time difference between the S1 and the S2 signal. From the size of the S1 and S2 signals, the energy deposited in the interaction can be reconstructed. A typical interaction or "event" is of an S1 followed by an S2 within one drift time (=2.3ms), i.e. the maximum time it takes for an electron to traverse the length of the detector [21].

3.2 Background mitigation

We saw in Section 2.3 that to observe any excess in a direct detection experiment, a good knowledge and removal of possible backgrounds is crucial. For the XENONnT detector, the relevant backgrounds can be split into two types: intrinsic and extrinsic.

Extrinsic backgrounds come from cosmogenic or environmental sources. This can be in the form of muons and neutrons from cosmic rays or solar neutrinos, as well as environmental sources of radioactivity that results in α, β and γ radiation that may scatter off of the LXe. These scatters can also result from intrinsic radioactive impurities such as ^{85}Kr, ^{222}Rn trace amounts of uranium, thorium or ^{60}C and their daughters mixed in within the LXe volume or emanating from the walls, electrodes or other components of the detector. Electronegative impurities may also capture electrons and attenuate observed signals [10]. Additional instrumental effects like dark counts in the PMTs can also result in a background signal.

To mitigate the effects of extrinsic backgrounds, the detector operates deep underground, and has a water shield and a neutron veto system that tags incident muons or neutrons respectively. By choosing low contamination materials and reducing the internal surface-to-volume ratio during construction, the intrinsic backgrounds from components are also reduced. The isotope ^{214}Pb resulting from the β-decay of Rn is the dominant background in the detector- a radon and krypton distillation column purifies the LXe volume. Additional purification systems also remove potential electronegative impurities as well as tritium, which proved to be a troublesome background in XENONnT's predecessor XENON1T. In the range of (1,30)keV, XENONnT has the lowest backgrounds ever achieved in a DM detector at a rate of only $\sim$15 events/(ton x year x keV) [10]. More details as well as further backgrounds and mitigation techniques can be found in found in [19] and [21]

Any leftover backgrounds from these sources are studied, modelled and cut from the dataset. This requires an accurate knowledge of their phenomenology. A common background reduction technique is to exclude electron recoil signals- background β-decays typically result in ER signals, while NR is the signature of WIMP and neutron recoils. Neutrons are much heavier than electrons, and so NR in xenon produces shorter "tracks" than ER which results in a smaller S2 signal (see Fig 5) [20]. The S1/S2 ratio therefore provides a means to distinguish between the two recoils in the detector.

The probability of DM scattering twice within the detector is incredibly low- thus events occuring within one drift time (2.3 ms), called double scatters (DS) can be removed as well. Finally, xenon as an interaction medium is also motivated by background reduction. It is a noble gas, i.e. it doesn't interact chemically and has long lived isotopes, meaning easier purification and lower background rates. Liquid xenon is also self-shielding. Any γ-ray emissions from the edges will not travel far into the detector leading to a pure central "fiducial volume" [14][22].

3.3 S2-only analysis & sub-GeV detection

There is also another motivation for choosing xenon. To result in a clear signal, the target nucleus needs to be of comparable mass as the DM particle to allow for coherent scattering. Particles with dissimilar masses would be less efficient in their momentum transfer, as most of the kinetic energy remains with the heavier particle. Too light nuclei would be prone to background interactions and result in a broad recoil spectrum that cannot effectively constrain the scattering DM mass [22]. Given its atomic mass of 131 GeV/c^2, xenon-based TPCs are a popular choice for probing WIMP masses O(100 GeV/c^2). DM in the sub-GeV range on the other hand cannot transfer enough recoil energy to the xenon nuclei excite the atoms. It can, however, result in the ionization of a single or few electrons through electronic recoils or the hypothesized Migdal effect, where a nucleus perturbed w.r.t the charge cloud from the electrons due to an nuclear scatter results in ionization [2][23]. Sub-GeV models may result in small energy deposits, but they contribute a high interaction rate in the detector and can lead to a detectable excess if sensitive enough [6].

As the energy deposited in the detector decreases, naturally so too does the size of the S1 and S2 signals. The smallest signals that can exist in detector is in the form of lone S2s, where the interaction energy is too small to result in an S1 over the threshold, but the amplification from the extraction field at the liquid-gas interface results in a small, yet detectable S2. These small S2 signals correspond to emissions of single or few ionized electrons, i.e. the expected signature of low mass DM in the detector. So despite being optimized to probe WIMP massses, LXe TPCs *are* sensitive enough to reach the low mass DM regime. In XENON1T, this was used to develop an "S2-only" analysis technique that could reach 2-3x lower energies than traditional analyses [24].

The lack of an S1 necessarily results in some loss of information- we cannot accurately reconstruct the z-position of the interaction *and* we lose the discrimination between ER and NR signals that is typically used for background removal. The biggest limiting factor in such ionization-only analysis is that the number of instrumental and radioactive backgrounds increases at the low-energy limit. Despite the background mitigation techniques described above, elevated rates of single and few electron emissions have been observed in nearly *all* liquid xenon TPCs, and we do not yet have a rigorous understanding of all of their origins [6][8]. Because of this, we are lacking a comprehensive background model, making it difficult to attribute any observed excess to light DM interactions. This means that the S2-only analysis until now has been restricted to only setting upper limits to low mass DM models by conservatively attributing *all* uncorrelated (see Section 4) small S2 signals to DM.

4 Background characterization and phenomenology

It is clear from Section 3.3 that accurate background modelling and reduction requires a thorough understanding of potential sources. For measurements of low mass DM, this requires an understanding of the single and few electron emissions in LXe TPCs. Interestingly, single and few-electron backgrounds have also been observed in other direct detection experiments [5]. There have been a few studies, including from the XENON1T collaboration, that have attempted to characterise these emissions and find their correlation with various detector conditions or large S1/S2 signals. The results were then used to explain their origins and develop methods of event selections to obtain a sample of uncorrelated signals that *could* be a low mass DM signature [9]. This section summarizes some salient findings from these papers, with a focus on identifying relevant correlations to be used for modelling.

Particle interactions in XENONnT are separated by long time windows throughout which single and few-electron signals can be observed. A key observation from nearly *all* characterization studies is the correlation of the rates of single and few-electron emissions with large S2 signals, i.e. large energy depositions in the detector. Hereafter, the S2s corresponding to single and few-electron emissions are called SEs to distinguish them from these large S2s, even though both arise from ionized electrons.

SEs that occur directly after a large main signal are called prompt electrons. The origin behind these are relatively well understood. They occur as a result of the light from the S1 and S2 signals photo-ionizing exposed metal surfaces or electronegative impurities with low work functions within the detector. In accordance with this reasoning, they are generally observed within a few drift times after an S2 signal. Another SE background is in the form of ”e-bursts”, i.e. a cluster of SE emissions trapped at the liquid-gas interface that ”burst through” to yield an uncharacteristically wide S2 signal [7]. These signals can also be identified and removed- notably, they were not observed in XENON1t which had a better extraction efficiency than previous experiments [8][9].

What *is* concerning for the low mass DM search are observed SE signals that extend to 100x the maximum drift time after a large S2. The origin behind these so-called ”delayed” electrons is not well understood, but there are two main competing hypotheses. (1) They may be due to impurities within the TPC- either in the form of negative ions that release electrons through collisional detachment, or electronegative impurities that capture electrons with a re-emission orders of magnitude later than the drift time. (2) They may be the a result of electrons losing energy at liquid-gas interface and getting trapped due to imperfect extraction efficiency. Unless captured by some impurity, they are freed after delayed timescales once they gain enough energy from the extraction field to overcome the potential barrier [9][25]. The two hypotheses are studied by understanding the dependence of the SE rate on detector purity (measured through the electron lifetime) and extraction field strength respectively [8].

Evidence for or against these two hypotheses vary depending on the detector studied- in XENON1T, there was more support for the presence of some unknown impurities. A similar sort of evaluation for XENONnT is beyond the scope of this thesis. The focus is instead on modelling the this long tail of SE signals using primarily the results from the XENON1T study. XENONnT has carried over much of the same infrastructure from this detector [21]. The two hypotheses are highlighted because in *both* cases, delayed single electron emissions should be correlated in position with their progenitor S2 pulse as the resulting electrons produced drift upwards with minimal x-y divergence. This will prove to be a useful tool for analysing the models.

The typical approach has been to try and find correlations between these delayed SEs and the large S2 signal occuring directly before it. If uncorrelated, these should follow an exponential distribution given the random, independent arrivals of particles in the detector [7]. The results, however, show that the rate of SEs falls off with a power law (i.e. of the form t^{-n}) after a large S2 pulse. The exponent value determined from different studies are summarized in Table 1. The physical reason behind this power law form is not well understood- it indicates that these emissions have no characteristic timescale. When fitted to a restricted range after each S2, the power law exponent was also found to fall very close to or even smaller than 1, which makes it unnormalizable [7][9]. That is, any cut that simply removes signals following this power law is not a technique that can be extended to long timescales.

A higher energy interaction in the detector will also result in more ionization electrons. According to the both hypotheses, this should also result in a larger delayed SE rate. Electrons from S2 signals deeper within the TPC will also diffuse and potentially encounter impurities more, so we should see a larger delayed SE component there too. Both of these were observed in XENON1T when each SE was associated with its directly preceding S2 [8]. Additional parameters like drift field dependence or LXe purity can also vary SE rates, but are not considered here since they remain mostly constant over short detector runs [21].

Detector	**Active target**	**Fitted range after S2**	**Power-law exponent**
RED-1 [25]	5 kg	0.2 - 20 ms	1.4 ± 0.2
LUX [7]	250 kg	0.325 - 1000 ms	1.1 to 1.0
ASTERiX [9]	0.150 kg	0.03 - 1 ms	1.20 ± 0.04 for 1e 1.64 ± 0.09 for 2e 2.34 ± 2.24 for <5e
XENON1T [8]	2000 kg	2 - 200 ms	1.1 for 1e 1.3 for 2e 1.4 for <5e

Table 1: A summary of the power-law decay exponents for delayed electron rate since preceeding S2 observed in different LXe TPC masses.

5 Approach

It is clear to see that creating a comprehensive, all encompassing background model is not a trivial task. The electron emissions have different phenomenology, and not only do they potentially depend on a variety of parameters, but the parameters themselves may range over several orders of magnitude. Section 4 outlined how the single electron emissions strongly depend on the properties of the progenitor large S2 signal, i.e. its time and its size. Prompt electrons can be easily removed by removing a few drift times after such a large S2 signal- the delayed electron tail requires a more complex model.

The basic idea here is to develop three models for these delayed electrons in increasing order of complexity. As a baseline, we can start by fitting a constant background rate, i.e. this takes into account neither the time nor size correlation w.r.t the S2s but instead establishes an average. The second model, driven by the approach taken in the characterization studies, gives the rate of single electrons as a function of their time since their preceeding S2- we saw in Section 4 that this should take the form of a power law decay, i.e. we have a series of truncated power laws that ”start” and ”stop” in the time intervals between each S2 signal. We will see that this approach is limited due to the varying time windows between each S2, and find that the previous S2 is not always the progenitor signal.

The third model solves this by removing this truncation of the power law and instead models each SE as a function of *all* of the S2 signals before it. Additionally, the energy and depth of the interaction are also taken into account through the size of the S2 signal. The area of the S2, i.e. the total number of photoelectrons measured in the PMT scales with interaction energy in XENONnT. The width, i.e. the spread of the peak in time as the corresponding electrons reach the PMT is a measure of the depth of the interaction. Deeper interactions correspond to wider peaks as the electron diffuse while passing through the detector [2]. If we include both these parameters, we can make sure to distinguish between elevated rates because of a larger signal vs. an interaction closer to the PMTs. This third approach has not been previously explored in literature and is the main contribution from this thesis.

5.1 Data selection

The models were tested on a three hour snapshot of data from Science Run 0, which collected data from May to December 2021. Populations corresponding to large S2 signals and the single and few (<5) electron emission signals were selected as shown in Fig 6. An algorithm trained on simulated data from XENONnT classifies the peaks into types based on its area and width. The 2D histogram in Fig 6 shows that this rudimentary selection method already includes the vast majority of SE signals in the detector. The single and few electron signals (called n-e on the graph) were combined together to form the ”SE” dataset- from the three hours dataset alone, this provided close to $2x10^6$ SE signals. The selection also includes the largest S2 areas which can correspond to muon interactions- these have been observed to result in an extend period of electron emissions and make up an important background in the detector [27]

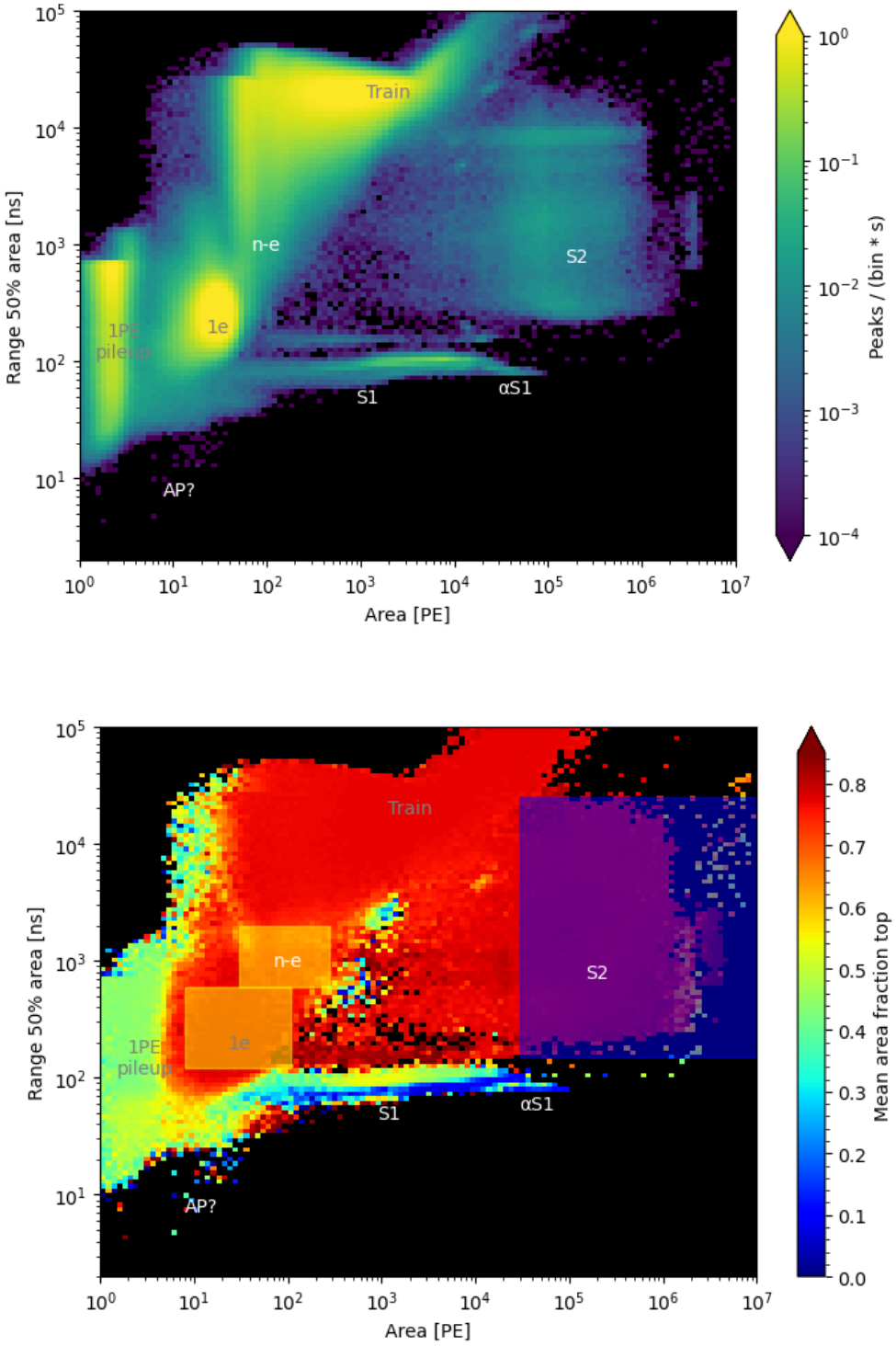

Figure 6: **Top:** A 2D histogram of *all* signals captured in a 3hr time window. The peaks are binned based on their area measured through the number of photoelectrons (PE) detected in the PMTs, and the time range that includes 50% of the signal from start to end of the peak on the y-axis. **Bottom:** A colour map of area, width and the mean area fraction top, i.e. the total light of the signal seen by the top PMT. The plots are made using STRAX, the processing framework for XENONnT data, which uses these parameters to also classifies peaks into types. The yellow box indicates the selected single and few electrons to study (SE) and green shows the chosen large S2 signals- with this selection, we make sure to also include the majority of SE signals as seen from the top histogram.

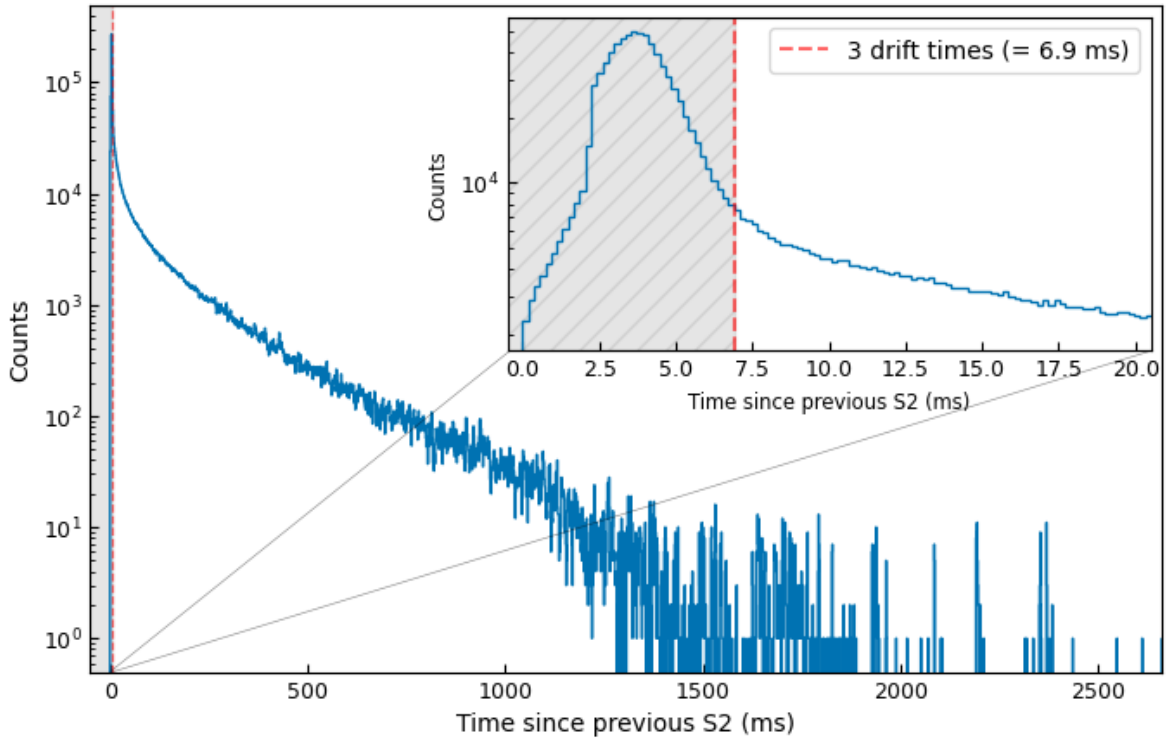

Figure 7: **Main:** A histogram of the time difference between each SE signal and the primary S2 signal that occurred directly before it. **Inset:** A zoomed-in version of the same graph for the region directly after each S2. The gray shaded region shows the cut that was applied to remove the prompt electrons.

The focus is on finding a model for the *delayed* electrons. Fig 7 shows the region of prompt photoionization electrons after each S2 that was removed from the dataset. While these should typically be restricted to one drift time after a signal, there can be significant second and even third order photoionization effects, where the scintillation from these prompt electrons themselves result in more photoionization [14]. It is clear to see that the region of interest extends to O(100x) the drift time, and the cut can therefore be liberally placed at three times the maximum drift. This factor has also been used in the past [14]- it retains $\sim$97% of the time to fit while removing $\sim$40% of all single electrons from the initial selection.

5.2 The extended unbinned MLE method

The models were fit using the iminuit package on Python, originally designed by CERN's ROOT team for optimized parameter estimation and error calculation in high-energy physics. iminuit's *extended unbinned maximum likelihood* cost function was used for all fittings. Single electron emissions persist for several orders of magnitude and arise out of a variety of different processes, i.e. they encompasses a wide, dynamic range. Binning this data, especially SEs plotted as the time since start of the run, could potentially result in the loss of smaller scale features. If not computationally intensive, data is usually fit unbinned [28].

The maximum likelihood estimation method (MLE) takes a specific probability distribution function (PDF) and fits it to the observed (in this case unbinned) data by finding the parameters for which its likelihood function is maximized [29]. Intuitively, the likelihood function $L(\theta|\mathbf{x})$ quantifies the

probability of observing the data $\mathbf{x}$, given that it arises from a distribution with parameter θ. In the case of multiple parameter models, the method involves finding the combination of parameters that can maximize this function (or more commonly, minimize its negative logarithm). iminuit does this numerically. A covariance matrix is also returned, from which the variance of each parameter around its maximum (i.e. the matrix diagonal values) is used to give a lower bound to the uncertainty in the estimated parameters [30]. This can be propagated to the values estimated from the fit from a numerically computed Jacobian using the Jacobi Python package. In the following section however, the models are visualized by binning data. The error of each bin is taken to be the square root of its number of entries, i.e. they are assumed to be Poisson distributed and so the variance is given by the square root of the mean.

The benefit of the MLE method is that it is asymptotically efficient, i.e. it achieves the lowest possible variance among estimators as the sample size increases [29]. The results can be scaled up to larger datasets with minimal discrepancy [28]. Furthermore, the models explored in this study are variants of a power law distribution, for which the MLE method is better suited for estimating parameters as opposed to alternatives like a linear regression from a log-log scale [31].

In the *extended* maximum likelihood method, the normalization of the model is left as a free parameter, i.e. both the shape and the overall sample size is estimated. The probability density function (pdf) integrated across the fitted range returns a finite value, but this value is not required be 1. The characterization studies found that the number of single electron emissions *can* vary with various detector conditions like electron lifetime, drift field etc. Choosing an extended fit takes this into account- if changing some particular parameter results also in an increase/decrease in the number of electrons, this is reflected in the covariance of the fit [32].

6 Results & Discussion

The selection and cut as described in Section 5 finally resulted in ∼1,100,000 SE signals in the dataset, visualized in Fig 8. To begin with, a uniform distribution is fit to the data to find an average rate of ∼105 delayed single and few electron signals per second. The fit does not pass through the majority of the errorbars so it is immediately clear to see the variation in events across the run. Indeed, we also saw in Section 4 that such a model is not well justified- elevated rates of SEs after a main S2 extend into the delayed region as well.

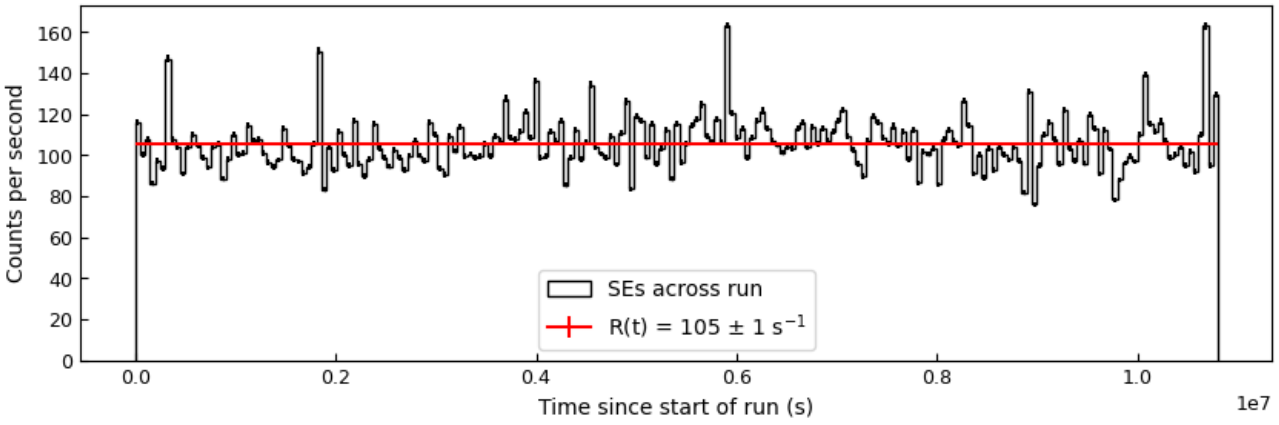

Figure 8: Rate of electrons expressed as a function of time since the start of the run. The result shows an averaged rate of ∼105 per second. For visualization, the bin widths here are 50s long, and the y-axis has been scaled accordingly.

We can now introduce our first level of complexity, namely, the rate of single electrons as a function of time from its progenitor S2 peak. On average, single electrons can only exist for some finite time (15.0±0.4 ms) in the detector before being captured by electronegative impurities [21]. The most obvious guess is that each single electron signal has been caused by the S2 peak occurring directly before it. We can therefore associate each single and few electron signal in our dataset with the most recent previous S2 and express the SE rate (R) as a function of their time differences (Δt), as opposed to the time since the start of the run. This allows us to use the characterization studies- i.e. the rate of single electrons decays in a power law fashion after a main signal. We can fit the following power law model:

$$R(\Delta t|A,n,t_{min}) = A * \begin{cases} 0 & \Delta t < t_{min} \\ \frac{n-1}{t_{min}}(\frac{\Delta t}{t_{min}})^{-n}, & otherwise \end{cases} \tag{3}$$

where the power law within the brackets has been normalized, and A is found from the extended MLE to scale the function to fit the total sample size. For our dataset, $t_{min} = 6.9ms$ because of the photoionization electron cut.

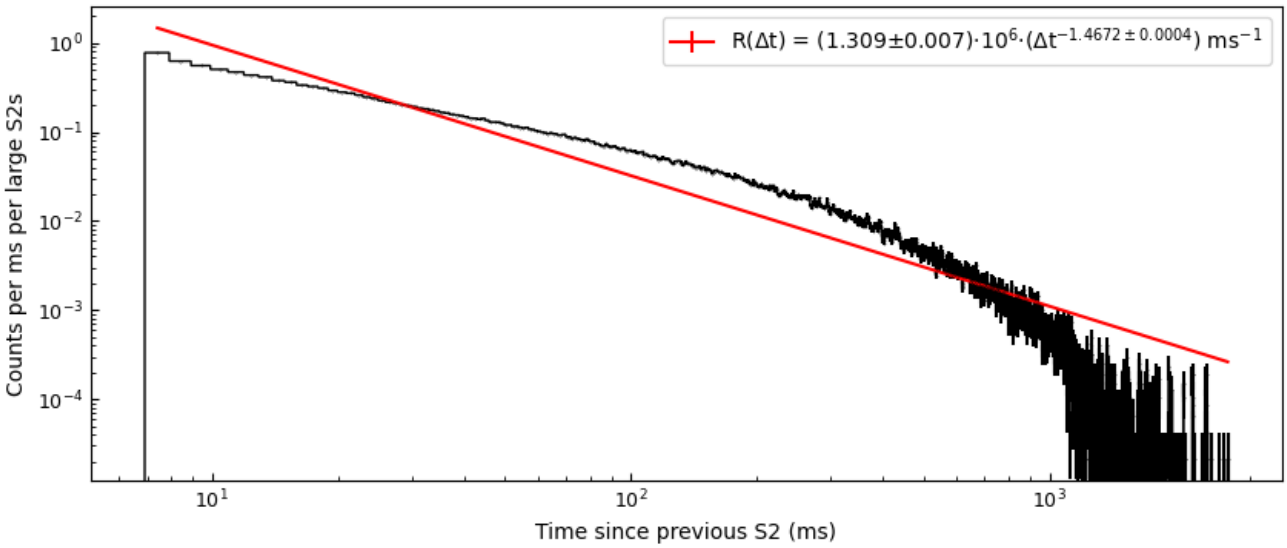

Figure 9: Rate of electrons expressed as a function of time since the previous large S2 signal. The fit here is a simple power law with t_{min}=6.9ms since the prompt electron region was removed. The y-axis has been divided by the resolution of each bin and the total number of S2s in the dataset to allow for easier comparisons between different datasets.

Fig 9 shows the results of this fit. Despite the small uncertainties in the parameters, it is also clearly inadequate. The exponent n=1.4672±0.0004 is also higher than what was found in previous studies (see Table 1). On examining the data, the power law behaviour seems to be restricted to smaller time differences, with a sharp drop in the tail end of the data which is more characteristic of an exponential distribution on logarithmic axes.

There is an immediate argument that can be made here- electrons at larger time differences from the most recent S2 are uncorrelated, since an exponential distribution is what we would expect to see in the case of independent processes with some constant rate, i.e. a Poisson process [7]. Arguably, electrons at these longer timescales may be caused by a different mechanism than earlier ones.

We can check this idea by plotting a histogram of the time vs. position differences of the SE signal associated with the most recent S2. This is shown in Fig 10(a) where the plotted data is restricted to S2s that occurred in the central region (<30cm) of the detector to avoid bias towards smaller position differences from signals too close to the edge. The results are surprising- there is a spread in the position of the single electrons w.r.t. its previous S2 across all time differences. If later electrons were the mostly uncorrelated ones, we would expect to see the yellow region along the diagonal of the axes. Instead, we observe that even electrons immediately after a large S2 signal can be significantly separated from them in position. This position non-correlation has also been observed in previous studies- in XENON1T, when the correlated electrons are defined as those <15cm away from their previous S2, the ”fraction of (position) correlated electrons represent a non-negligible proportion of the total number of delayed electrons” [8]. Given the two hypotheses of the origin of these electrons,

this should not be true if the previous S2 is almost always the progenitor signal for these single and few electrons.

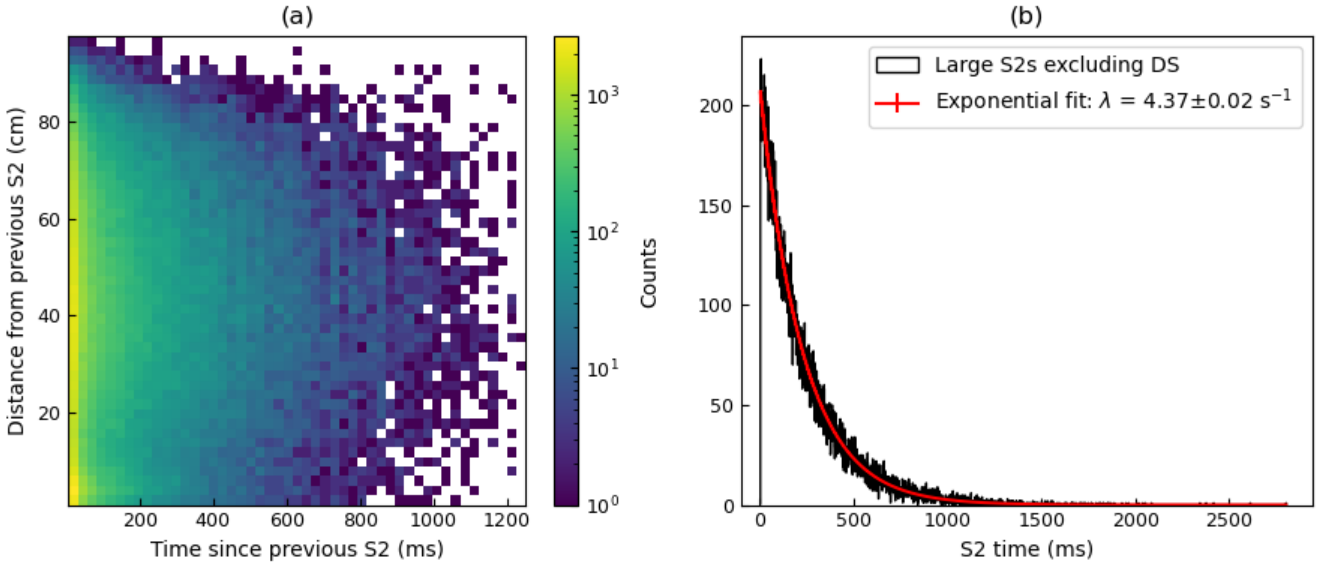

Figure 10: **(a)** A histogram of the time vs. position differences of the SEs from their previous associated S2. Central S2 peaks are chosen to avoid bias. **(b)** A histogram of the time windows in between each main S2 event fitted with an exponential decay function in accordance with their expected Poisson behaviour. Double scatters have been removed in (b) to prevent a large spike at t¡2.3ms

So, there *is* a different gradient across timescales in Fig 9, but given this evidence, we cannot confirm that the drop in the tail is indeed due to more uncorrelated electrons. Multiplying the power law with an exponential, for example, may result in a better fit, but it is not justified physically. We can try and investigate where this change in gradient occurs by fitting a (normalized) continuous broken power law of the form:

$$R(\Delta t|n_1,n_2,t_b,k,t_{min}) = A * \begin{cases} 0 & \Delta t \leq t_{min} \\ \dfrac{\Delta t^{-n_1}}{\Gamma} & \Delta t_{min} < t < t_b \\ \dfrac{t_b^{n_2-n_1}}{\Gamma}\Delta t^{-n_2} & \Delta t \geq t_b \end{cases} \tag{4}$$

where:

$$\Gamma = \frac{t_b^{1-n_1} - t_{min}^{1-n_1}}{1-n_1} + \frac{t_b^{1-n_1}}{n_2-1} \tag{5}$$

and allow the break point to be placed anywhere between the start and end of the data. The results of this are shown in Fig 11.

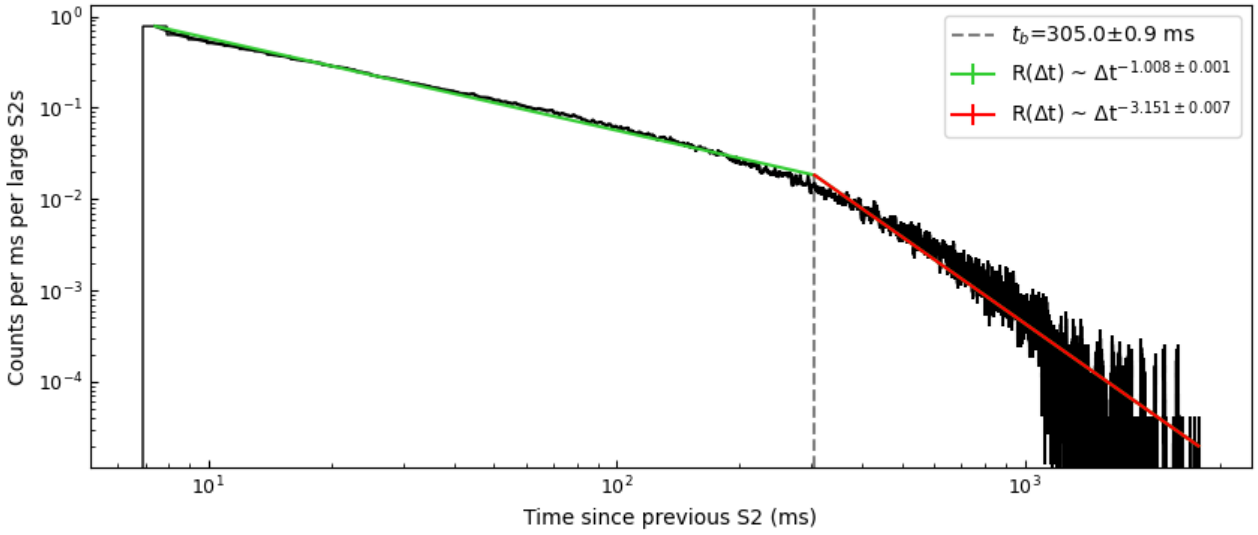

Figure 11: A broken power law fit to the data, again, to obtain the rate of SEs as a function of time from the most recent S2. The break point was allowed to be placed anywhere between t_{min}=6.9 ms to the last point in the dataset

A visual examination now shows a much better fit, with the model passing through the majority of errorbars. Regardless of initial values or limits provided for any of the parameters, the break point was consistently placed around $t_b = 305.0 \pm 0.9ms$. The exponent $n_1 = 1.008 \pm 0.001$ is more in line with the results in Table 1 that also fit to a similar restricted range. The second power law exponent is more than triple this value.

This result highlights an important issue with this model. While an increased rate of SEs can be correlated with large S2 signals, the S2 signals themselves are independent of each other. If we take the time windows between large S2 events, we find Fig 10(b)- they decay exponentially as expected from their Poisson nature, with a rate of $\lambda = 4.37 \pm 0.02 s^{-1}$. That is, we expect that on average, large S2s are separated in time by windows ~230 ms. This is in the same order of magnitude as t_b in Fig 11. So, it is possible that the bigger drop in counts beyond this point is artificially caused by the fact that it becomes increasingly rare to find two S2s separated by such large times.

This time window limitation has been noticed in the past. In [7], the study of SE-previous S2 time correlations is restricted to consecutive S2 signals with long time window separations, but this can be difficult to find given the large number of double scatters (DS) in the detector. Essentially, when a single particle interacting twice or even multiple times with the LXe will result in S2 signals occuring within one drift time (=2.3 ms) of each other. DS signals made up ~35% of all large S2 peaks in the dataset. A single electron from the first scatter at some $\Delta t > 2.3ms$ could become arbitrarily associated with the second scatter instead, which may have occurred at a different location in the detector. Indeed, when looking at the number of single electrons associated with each S2 signal, there is a marked increase for DS compared to single scatter signals despite them being identical in area, indicating that this increase is likely because many of the electrons are coming from a previous peak (see Fig 13 in Appendix A). This is then reflected in their non-correlation in positions in Fig 10. Particularly for smaller separations in Fig 10(a), it is likely that the progenitor S2 is *not* the most recent one. Rather, it may have originated from an earlier S2 event.

In Fig 9, 10(b) and 11 the DS peaks have been removed so that only the latest large S2 in a multiple scatter interaction remains. However, there is no reason why this problem should be limited to time windows <2.3 ms. Even for two independent S2 events separated by say, 10 ms, we can expect single electrons from the first peak to "bleed into" the second. What we are plotting then, does not accurately correlate the single electrons with their progenitor pulse. Indeed, it has been observed that including the time differences of position uncorrelated S2-SE pairs essentially has the effect of flattening the power law descent [8]. This is one possible explanation for why the exponent calculated from this method can sometimes fall to <1.

All of this evidence seems to point us towards the fact that a naive prescription of each SE signal to the S2 that came directly before it is wrong. That is, a single electron may be caused not just by the most recent interaction, but also potentially by all of the interactions that came before it. A power law model as described above is limited to a specific range dictated by the average time windows between S2 events.

Another effect that we have not considered here is the size of the S2 itself- given two double scatters, we would expect more of this "bleed" for a larger first peak, since this would show an elevated number of SEs as we saw in Section 4. From Fig 6 it is clear to see that our S2 signals span a few orders of magnitudes in both area and width, both of which can affect the number of emitted delayed electrons. This limitation has also been understood in past characterization studies. In [8], the effects of overlap from the size of the S2 was quantified using a "shadow parameter", with larger S2s casting a longer "shadow" on the SEs that occur directly after. The shadow region indicates potential correlations with the S2 peak- regions in heavy shadow could be cut out since they are likely generated by large S2 peaks.

The model in equation 3. is given in terms of Δt, i.e. the time difference between the observed SE and the most recent S2. If we instead convert this to be in terms of the time since the start of the run (t), we would essentially find several power laws that "start" and "stop" in the time intervals in between each S2 signal. Since we have established that the progenitor S2 peak can be any of the peaks before the SE signal, an easy fix is to get rid of this truncation, and allow each power law to "bleed" into successive S2 peaks. Since we can expect an elevated overlap for a larger S2 pulses, we can also add a multiplicative factor to the power law to control the starting rate of these emissions. The power law form of each S2 peak means that for any given point in time, the contribution to the SE rate comes primarily from the preceding peaks, but the effects of far away large S2 peaks may still be apparent (i.e. it remains non-zero).

With this model the contribution of each S2 pulse to the background rate R_p is given by the following function:

$$R_p(t,t_p,A_p,W_p|n,k,t_{min},\alpha,\beta) = \begin{cases} 0 & t < t_p + t_{min} \\ A_p^{\alpha} W_p^{\beta} (t-t_p)^{-n}, & otherwise \end{cases} \tag{6}$$

where t_p is the time at which the S2 peak occurs, and A_p and W_p are its area (in PE) and width (in ms). The factors α and β controls the scale at which the two parameters affect the SE emissions- i.e. the differential rate. The multiplicative factor represents the most generalized form of this rate- if the observed SEs have no dependence on width, for example, β would simply go to zero.

The overall background electron rate from the start of the run (t) is now the sum of these contributions:

$$R(t|k,C) = k + C * \sum_{p} R_p(t, t_p, A_p, W_p | n, k, t_{min}, \alpha, \beta) \tag{7}$$

where p is all of the S2 peaks in the run. During plotting, each individual power law was first normalized similar to equation 3 and then multiplied by the differential rate. The sum in equation 7 is multipled by some constant C so the extended MLE fit scales the results to the sample size. This value of C multiplied by the differential rate therefore gives an idea of the number of SE signals contributed by each S2 peak. Finally, a constant k has also been added to the function. This returns the rate of signals *not* attributable to the power law decay after each S2. I.e. these could be potential low mass DM signals.

Because of the computational requirements of this cumulative model, the fit was restricted to 2000 S2 peaks, which corresponds to ∼300 s of data. The results are visualized for a small section of ten peaks in Fig 17- more snapshots from this fit can be found in Fig 15, 16 and 17 in the Appendix.

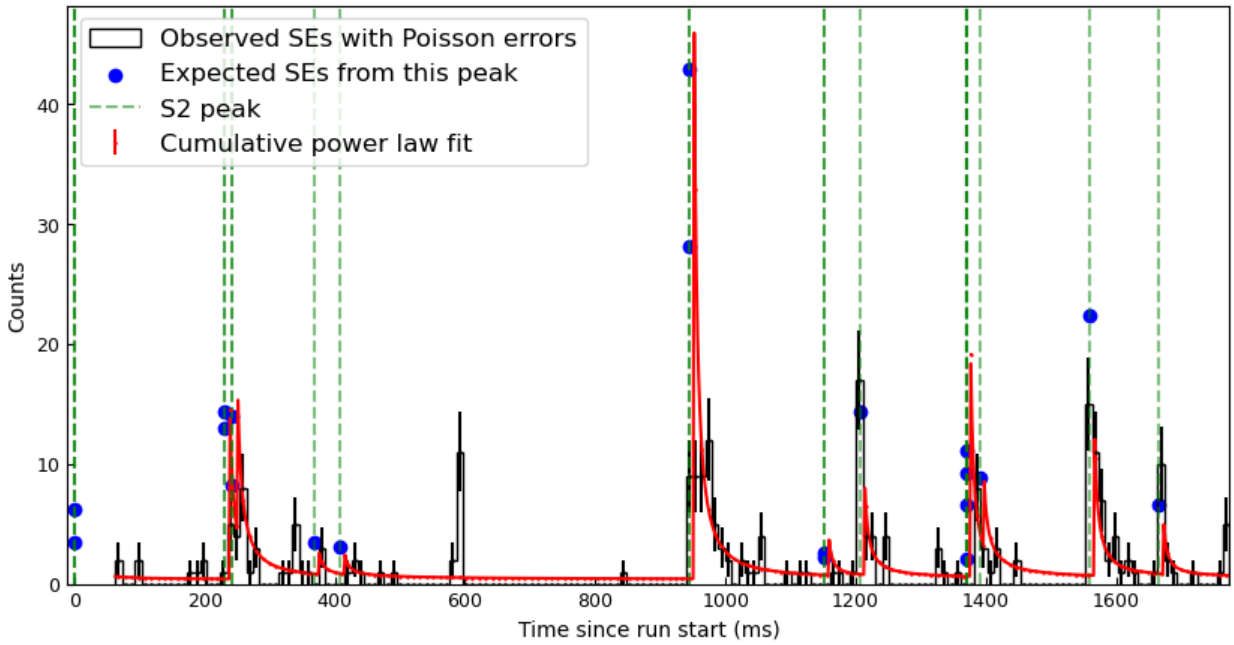

Figure 12: The cumulative power law fit visualized for a snapshot of some 10 peaks. The factor n has also been plotted as the blue scatter points. The resolution is 10 ns. The error bars are the square root of the counts in each bin. The fit passes through the majority of these errorbars.

It is clear to see from the plot that the scaling based on area and width does somewhat match the observed distribution- there is an increase in the number of events after S2s with larger differential rates. When SEs up to 100s after each peak were plotted separately, there is a small uptick in delayed electron rates for both area and width immediately after the S2 with a comparatively smaller variation in the latter (see Fig 13 and 14 in Appendix). The model predicts $\alpha = 0.880 \pm 0.004$ and $\beta = 0.505 \pm 0.006$ consistent with this.

This approach seems to be especially good at predicting SEs around clusters of several S2 peaks like around 1400 ms in Fig 17. It is also capable of predicting small features like around 400 ms. The covariance errors were also observed to be orders of magnitude smaller than the values, which provides an initial check against overfitting [32]. When the contribution from each peak is added up, the fit found a total SE count very close to the observed sample size.

Interestingly, the k constant gives a rate of 36.5$\pm$0.6 SEs per second. Compared to the rate from the uniform distribution, this is a threefold decrease- i.e. this model reduces the dark matter limit by a factor of three. When a similar k constant is added to the power law fits, both the basic power law and broken power law set go into the negatives, which lacks a physical interpretation. It is likely a result of the statistical limitations discussed above and so should be interpreted with caution. The positive value for k in this summed power law is a promising result.

The power-law exponent found from the fit is n=1.448 ± 0.005. While lower than the value of the basic power law, it is still higher than the values in Table 1. This introduces an important limitation in this current application of the cumulative power law fit because of computation requirements. In equation 6, we can see that the contribution of each peak only "turns on" some t_{min} after each peak since we have cut out this part of the data. Yet, the tails of previous S2 peaks can still continue in this region, i.e. the PDF predicts a non-zero amount of electrons in the region when there should be none. This would result in an artificial steepening of the predicted exponent and would especially affect predictions close to DS. The function should therefore be adapted so that for some t_{min} after each peak, it is *always* set to 0. This is an easy fix in theory, but the required normalization of the resulting PDF to use with iminuit quickly became computationally heavy.

7 Conclusions & Future Outlook

In this thesis, we attempted to develop a model for the delayed single and few electron emission background in XENONnT. These emissions are currently a limiting factor for low mass dark matter searches with liquid xenon time projection chambers, since such sub-GeV particles would also result in a similar signature. The models were built from the results of previous characterisation studies of the delayed electron tail backgrounds, particularly in XENONnT's predecessor. Three models of increasing complexity were tested, and a new approach that predicts the rate of single and few electron emissions as a function of all previous S2 times and sizes is proposed as a viable option to be further developed on. This approach is physically justified under the two main competing hypotheses for the production of these delayed electrons, and tentatively points towards a three-fold reduction in the dark matter limit. It can be used to predict SE signals even at long timescales after an S2.

This is however, still a preliminary result and there are many possible future steps from here. Until now, we have stuck to a mostly visual and analytical analysis of the results. A quantitative comparison of the quality of the fit is difficult- in fact, this is an open question for unbinned MLE methods [34]. The likelihood function values themselves cannot be compared, since the power law and cumulative power law model are fundamentally different. They are non-nested, i.e. they do not reduce down to each other for some special cases [35]. A common approach is to bin the predictions and perform a chi-squared test, for which the general rule of thumb is to have at least five events in each bin. While a crude binning is sufficient for the constant and power law models, it is clear to see that in the case of Fig 17, doing so would destroy any ability to observe the power law decay of the data. That being said, this cumulative power law model is well justified physically as opposed to just a naive previous-S2 only model. Exactly *how* much better still requires some further analysis.

The results of the fit can be significantly improved with more computational power. Firstly, this then allows the region after each S2 to be always set to 0, without which the steepness of the power law decay tends to be overestimated as outlined in the previous section. It also means that the model can be fit to longer run times, which can provide more insight into whether the results can be extended to different runs. The initial development of the models was done with a 100s snapshot of data on a different day. This run firstly observed a smaller photoionization period limited to 1 drift time, and it also resulted in a value >1 for β. It is possible that the form of the differential rate factor changes under different detector conditions, seeing as the rate of single electron emissions changes with electron lifetime and detector purity. With longer time scales, the annual modulation of these single and few electron emissions can also be studied to further distinguish between correlated backgrounds and potential dark matter signals.

Furthermore, signals from the entire volume of the detector was considered in this work. At times, the dataset had some isolated spikes similar to the one around 600 ms in Fig 17 that could not be predicted by the cumulative power law model. A possible reason for this could be spontaneous fluorescence from the walls of the detector or emissions due to impurities from these edges. A possible improvement would be to select SEs from only the inner fiducial volume of the detector, since this is method that is already implemented to remove backgrounds [21]. Finally, the method of data selection though suitable for a preliminary study, is fairly crude. A team of XENONnT data analysts have recently been testing more sophisticated classifications of these low-energy signals, based on their proximity to S2 signals. It would be interesting to see how the model fit changes when the selected signals to study is more restrictive. The results from these improvements would take us one step further in the hunt for dark matter.

8 References

[1] M. Roos, “Dark Matter: The evidence from astronomy, astrophysics and cosmology.” 2010. arXiv: 1001.0316 [astro-ph.CO].

[2] J. Remco. Aangevare, ”First WIMP results of XENONnT and its signal reconstruction,” Ph.D dissertation, Universiteit van Amsterdam, 2023. [Online].

[3] M. Misiaszek and N. Rossi, “Direct Detection of Dark Matter: A Critical Review,” Symmetry, vol. 16, no. 2. MDPI AG, p. 201, Feb. 08, 2024. doi: 10.3390/sym16020201.

[4] L. Baudis, “Dual-phase xenon time projection chambers for rare-event searches”, 2023, doi: arXiv:2311.05320 [physics.ins-det].

[5] R. Essig. (2024). Some progress & challenges for the direct-detection of sub-GeV dark matter. In Nuclear Physics B (Vol. 1003, p. 116484). Elsevier BV. DOI: 10.1016/j.nuclphysb.2024.116484

[6] R. Essig, T. Volansky, and T.-T. Yu, “New Constraints and Prospects for sub-GeV Dark Matter Scattering off Electrons in Xenon”, 2017, arXiv:1703.00910 [hep-ph].

[7] Akerib, D.S. et al. (2020) Investigation of background electron emission in the Lux Detector, arXiv.org. Available at: https://arxiv.org/abs/2004.07791 (Accessed: 24 June 2024).

[8] Aprile, E. et al. (2022) Emission of single and few electrons in XENON1T and limits on Light Dark matter, arXiv:2112.12116.

[9] Kopec, A. et al. (2021) Correlated single- and few-electron backgrounds milliseconds after interactions in dual-phase liquid xenon time projection chambers, arXiv:2103.05077 [physics.ins-det]

[10] E. Aprile et al., “Search for New Physics in Electronic Recoil Data from XENONnT”, 2022, arXiv:2207.11330 [hep-ex].

[11] P. Gorenstein and W. Tucker, “Astronomical Signatures of Dark Matter,” Advances in High Energy Physics, vol. 2014. Hindawi Limited, pp. 1–10, 2014. doi: 10.1155/2014/878203.

[12] A. Blanchard, “Clusters of galaxies,” New Astronomy Reviews, vol. 45, no. 4–5. Elsevier BV, pp. 401–407, Mar. 2001. doi: 10.1016/s1387-6473(00)00162-7.

[13] E. Corbelli and P. Salucci, “The Extended Rotation Curve and the Dark Matter Halo of M33”, 1999, arXiv:astro-oh/9909252.

[14] A. Kopec, “Few electron signals in liquid xenon dark matter detectors,” Ph.D dissertation, Purdue University, 2021, [Online]. Available: https://docs.lib.purdue.edu/dissertations/AAI30505198/

[15] 1E 0657-56. Harvard-Smithsonian Center for Astrophysics 60 Garden St. Cambridge, MA 02138 USA.

[16] T. Lin, “TASI lectures on dark matter models and direct detection.”, 2019. arXiv: 1904.07915 [hep-ph].

[17] R. Essig, J. Mardon, and T. Volansky, “Direct Detection of Sub-GeV Dark Matter,” 2011, arXiv:1108.5383 [hep-ph].

[18] R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022) and 2023 update, Chapter 27

[19] J. Aalbers et al., "A Next-Generation Liquid Xenon Observatory for Dark Matter and Neutrino Physics", 2022, arXiv:2203.02309 [physics.ins-det].

[20] E. Hogenbirk, "A spark in the dark Scintillation time dependence and neutron-induced signals in dual-phase xenon TPCs," Ph.D dissertation, Universiteit van Amsterdam, 2019. [Online]. Available: https://pure.uva.nl/ws/files/32523142/Thesis.pdf

[21] E. Aprile et al., "The XENONnT Dark Matter Experiment.", 2024. arXiv:2402.10446 [physics.ins-det].

[22] M. Schumann, "Dual-Phase Liquid Xenon Detectors for Dark Matter Searches", 2014, arXiv:1405.7600 [astro-ph.IM].

[23] E. Aprile et al., "Search for Light Dark Matter Interactions Enhanced by the Migdal effect or Bremsstrahlung in XENON1T", 2019, arXiv:1907.12771 [hep-ex].

[24] E. Aprile et al., "Light Dark Matter Search with Ionization Signals in XENON1T", 2019, arXiv:1907.11485 [hep-ex].

[25] D. Yu. Akimov et al., "Observation of delayed electron emission in a two-phase liquid xenon detector," Journal of Instrumentation, vol. 11, no. 03. IOP Publishing, pp. C03007–C03007, Mar. 02, 2016. doi: 10.1088/1748-0221/11/03/c03007.

[27] T. J. Anderson, "The LZ Dark Matter WIMP Search and Treatment of Fundamental Signals," Ph.D dissertation, Stanford University, 2023, [Online]. Available: https://arxiv.org/abs/2207.03764

[28] H. B. Prosper, "Practical statistics for particle physicists," arXiv:1608.03201 (accessed Jun. 24, 2024).

[29] Lista, L. (2017) "Practical statistics for particle physicists", arXiv:1609.04150 [physics.data-an] (Accessed: 24 June 2024).

[30] Cowan, G. (no date) Topics in Statistical Data Analysis for High Energy Physics. Available at: http://www.pp.rhul.ac.uk/cowan/statcourse.html.

[31] Goldstein, M.L., Morris, S.A. and Yen, G.G. (2004) Problems with fitting to the power-law distribution, arXiv.org. Available at: https://arxiv.org/abs/cond-mat/0402322 (Accessed: 24 June 2024).

[32] L. Lyons, J. D. Hansen, R. J. Barlow, and J. Orear, "Extended maximum likelihood," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, (accessed Jun. 24, 2024).

[34] Raja, R. (2006) A general theory of goodness of fit in likelihood fits, arXiv.org. Available at: https://arxiv.org/abs/physics/0509008 (Accessed: 24 June 2024).

[35] Heinrich, J. (2011) Pitfalls of goodness-of-fit from likelihood.
Available at: https://www.slac.stanford.edu/econf/C030908/papers/MOCT001.pdf.

Appendices

A Additional plots

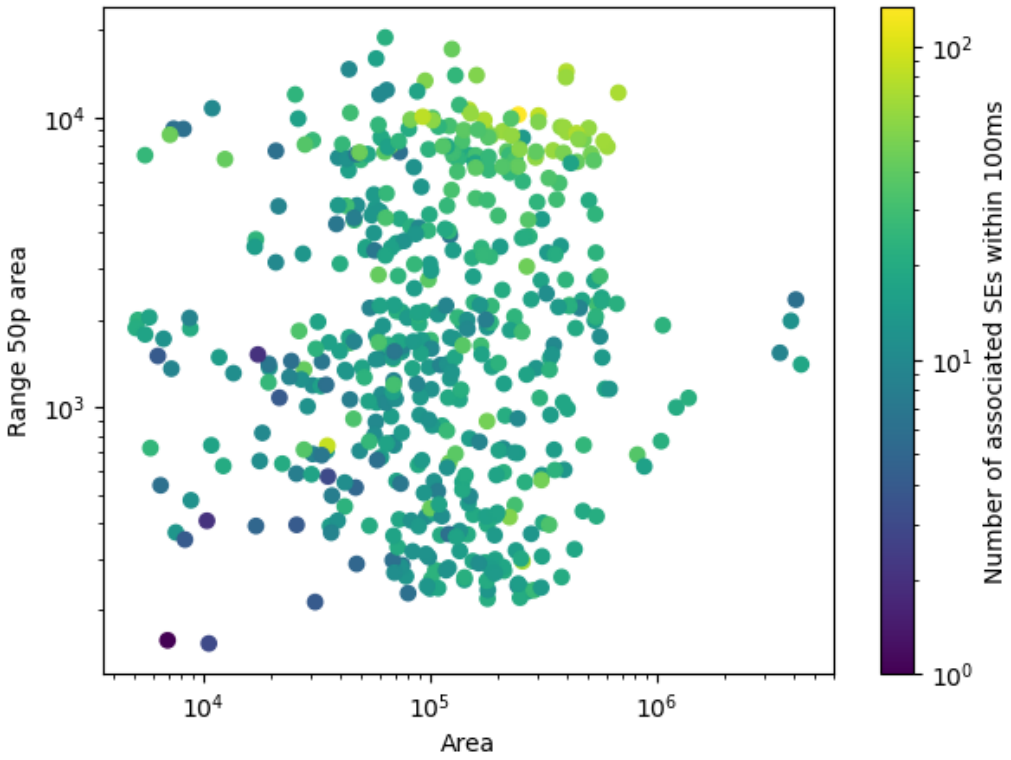

Figure 13: A scatter plot showing the elevated number of *delayed* single electrons as a function of the previous S2 signal properties. There is a clear increase with area and width. The scatter points were randomly selected from the set. Only 100 ms after each S2 is considered to avoid biasing longer time windows, and double/multiple scatters have been removed by associating the electrons only with the last peak.

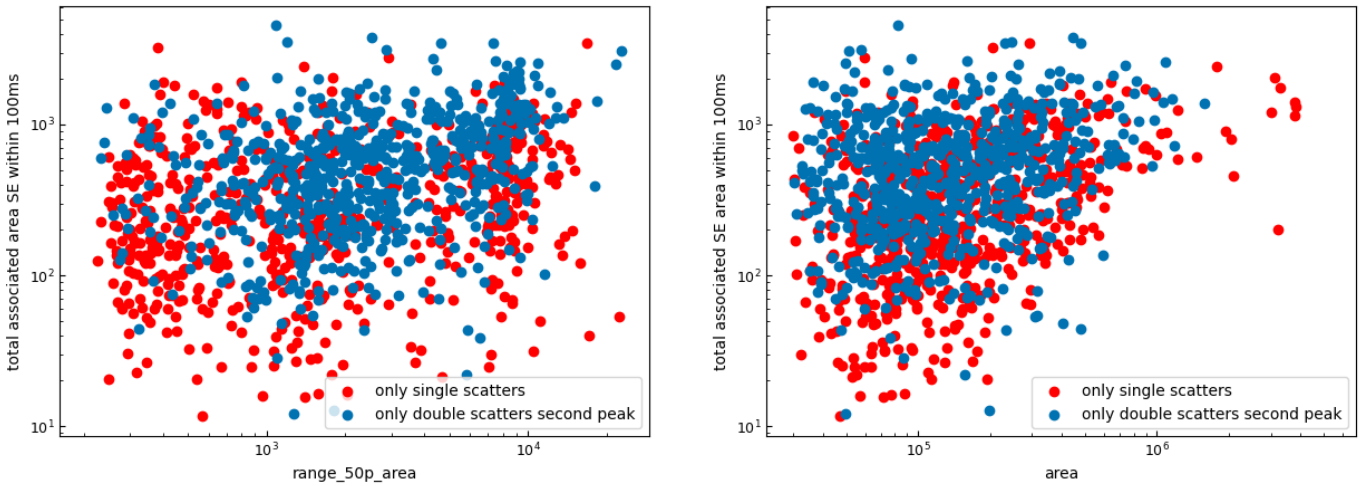

Figure 14: Scatter plot of single electrons after a S2 signal as a function of the area and width of the signal. The single and double scatters are differentiated here to show the sharper increase in the rate of single electrons after a double scatter for the same area. Data points were randomly sampled. Only SEs some 100ms after a main S2 is considered to avoid bias for longer time windows.

Some more visualizations of the cumulative power law model randomly sampled

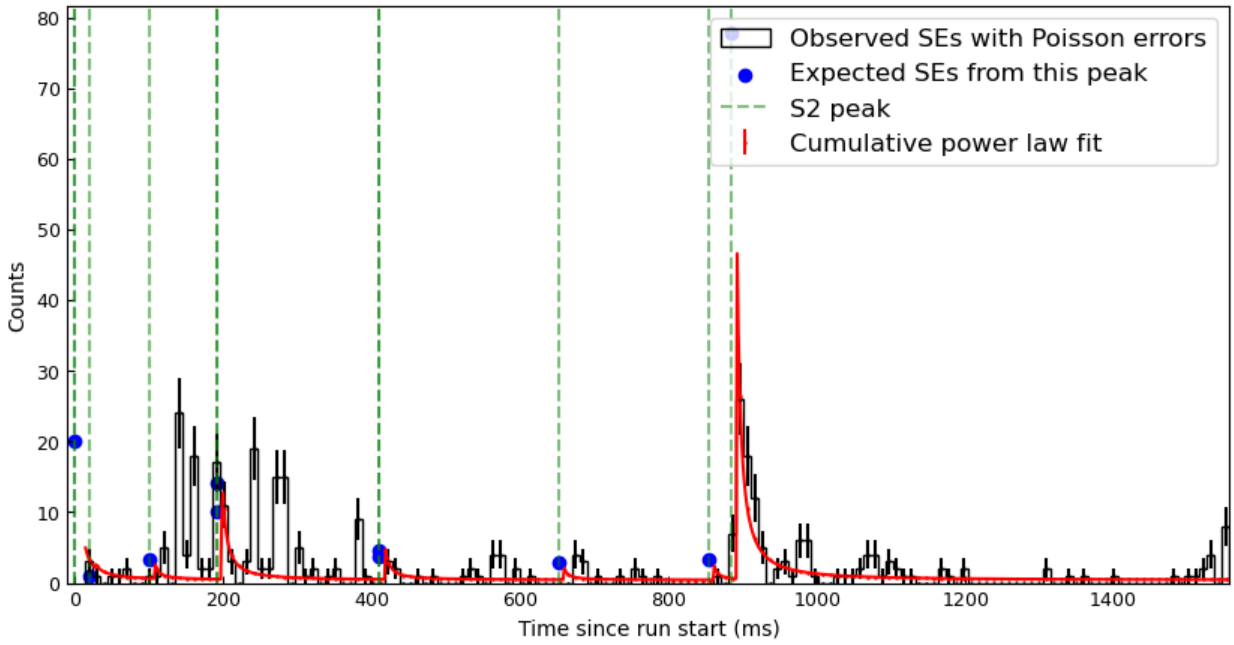

Figure 15: The cumulative power law fit visualized for a snapshot of some 10 peaks. The factor n has also been plotted as the blue scatter points. The resolution is 10 ns. The error bars are the square root of the counts in each bin. The fit passes through the majority of these errorbars.

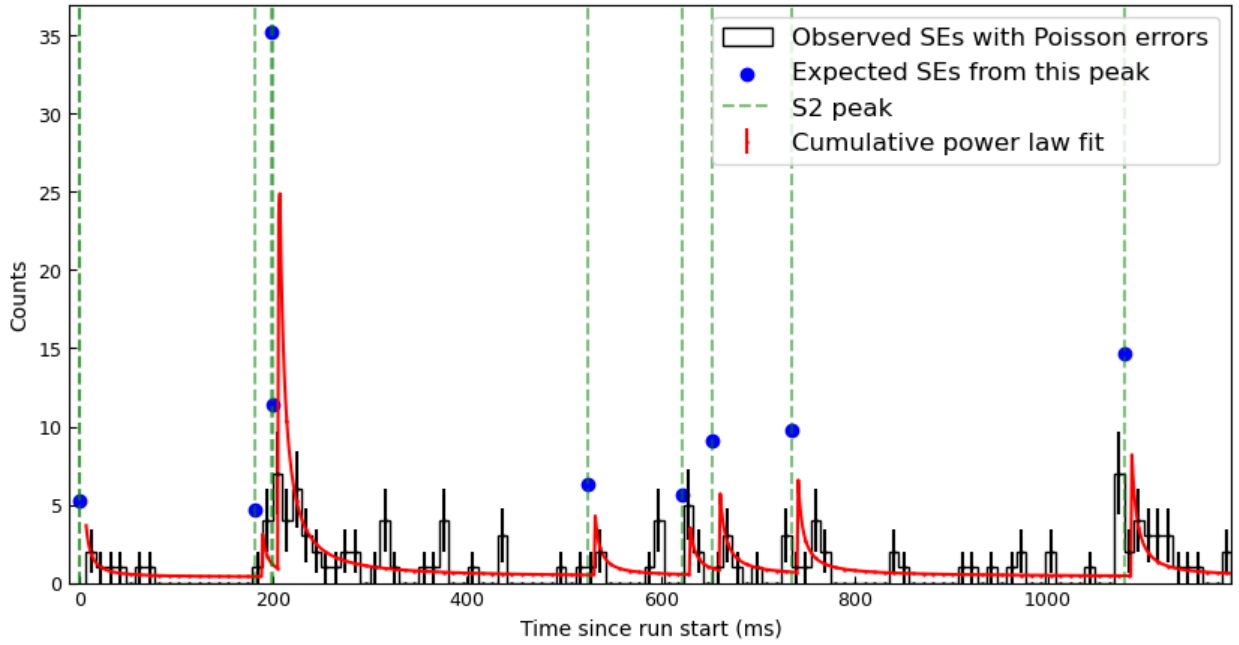

Figure 16: The cumulative power law fit visualized for a snapshot of some 10 peaks. The factor n has also been plotted as the blue scatter points. The resolution is 10 ns. The error bars are the square root of the counts in each bin. The fit passes through the majority of these errorbars.

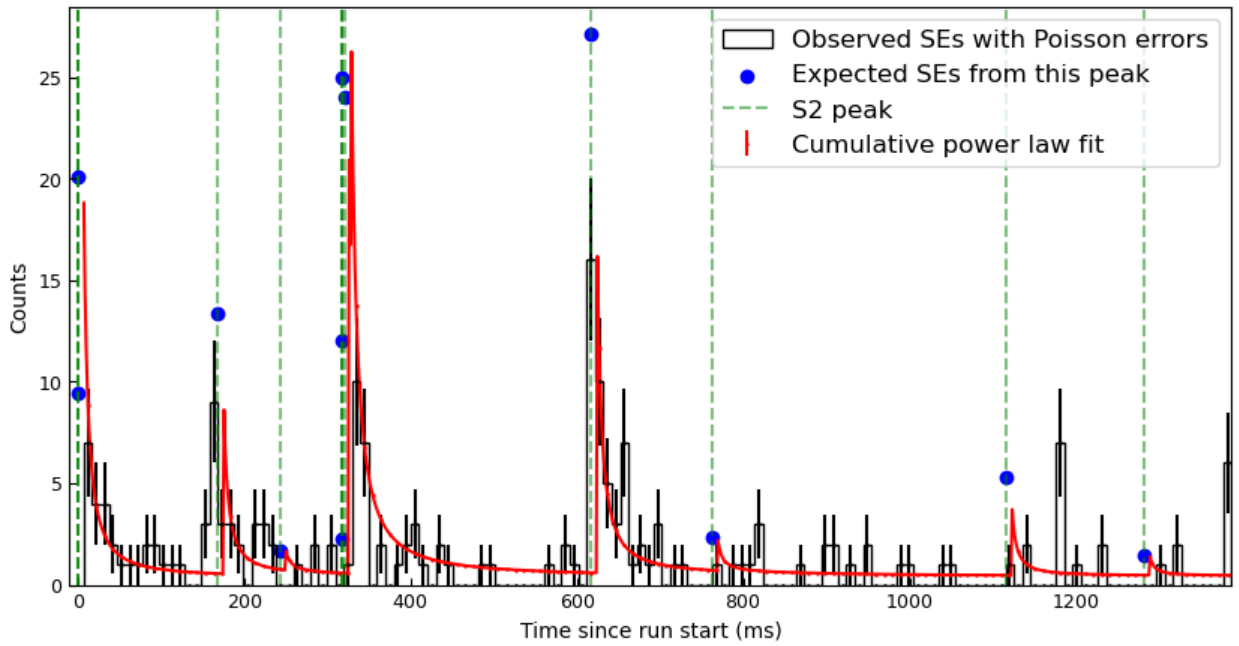

Figure 17: The cumulative power law fit visualized for a snapshot of some 10 peaks. The factor n has also been plotted as the blue scatter points. The resolution is 10 ns. The error bars are the square root of the counts in each bin. The fit passes through the majority of these errorbars.

B Code

The Github repository containing the code used in this investigation as well results of fits can be found **here**.