Johanna Jakob

Signal calibration for XENON1T with the new data reconstruction framework straxen

Münster, March 2021

WWU
MÜNSTER

Master thesis

Signal calibration for XENON1T with the new data reconstruction framework straxen

Johanna Jakob

First referee: Prof. Dr. Christian Weinheimer
Second referee: Prof. Dr. Alexander Kappes

Institut für Kernphysik
Westfälische Wilhelms-Universität Münster

March 31, 2021

Contents

1 Introduction

Consistent astrophysical observations during the last century led to the finding that our Universe includes an unknown, invisible type of matter. Dark matter constitutes about 26 % of its mass-energy content. Dark energy, responsible for the accelerated expansion of the Universe, makes up further 69 %, leaving only 5 % for ordinary, baryonic matter. Numerous new particles were postulated as candidates for dark matter. A promising candidate is the weakly interacting massive particle (WIMP). Despite its predominant abundance, its missing electromagnetic interaction and the small cross-section on the scale of the weak interaction make it extremely hard to detect. Several collaborations with low-background detectors are in search for the WIMP. However, until today, no direct observation of dark matter was made.

The XENON Dark Matter Project is one of the leading collaborations searching for signatures of WIMPs. With the XENON1T experiment, the collaboration set a world-leading exclusion limit on the WIMP-nucleon interaction cross-section. XENON1T was located at Laboratori Nazionali del Gran Sasso (LNGS) in Italy. As its successor XENONnT, currently under commissioning at LNGS, it employed a dual phase time projection chamber (TPC) with a liquid xenon target. Energy depositions in the TPC create both scintillation light and ionisation electrons. The latter generate a proportional secondary light signal. Both are detected with sensitive photomultiplier tubes (PMTs). The data are digitised and processed by the XENONnT-specific framework *streaming analysis for XENON* (straxen).

The low-background dark matter search requires a well-understood detector. Calibration campaigns and signal corrections are important in order to match this requirement. Homogeneously mixed calibration sources in the xenon target are used to perform an energy reconstruction and derive systematic corrections on both size and position of signals caused by energy depositions in the detector. In preparation of the science data taking of XENONnT, this thesis will utilise straxen mainly on XENON1T data in order to calculate and implement corrections on the reconstructed sizes of the calibration sources' signals. These analyses can then serve as a basis for XENONnT signal corrections.

After this introduction, the second chapter explains the concept of dark matter as well as the evidence that led to its postulation. In the third chapter, the experiments XENON1T and XENONnT are presented. The two calibration sources ^{83m}Kr and ^{37}Ar used for this thesis are introduced together with their signal corrections. XENON1T ^{83m}Kr calibration data are used in the fourth chapter for a correction of the light signal size. Also, the electron lifetime in the TPC is determined and used for a correction of the charge signal. In the fifth chapter, both charge and light signal analysis are repeated for ^{37}Ar. The low-energetic signal of this source has a deformation effect on the light signal, which is investigated with a simulation of the shape of the energy distribution. In the sixth chapter, the noble element simulation technique (NEST) software is introduced. Its model for signal generation is examined and applied to both calibration sources. A comparison of the simulation with the observed peaks in XENON1T data indicates a discrepancy occurring at the level of simulated electron recombination. Possible inhomogeneities in the drift field of the TPC that could explain this discrepancy are investigated in the seventh chapter. A method combining a data driven approach with NEST is developed, allowing a statement on the spatial behaviour of the field strength and a limited estimation of the absolute values. The last chapter contains a preliminary analysis of data taken by the next generation experiment XENONnT. At the moment, it is under commissioning, but is expected to start science data taking in the course of the year 2021. With data from the ^{83m}Kr commissioning calibration campaign, the correction of the light signals is successfully transferred to XENONnT, where it will be used as part of the data processing.

2 Dark matter

Visible, baryonic matter on cosmological scales has been observed to behave in a way that is not explained by the standard model of physics. The existence of a yet unknown type of matter not interacting with light offers an explanation of these effects. From the long list of evidences, a few are collected in this chapter. Further, possibilities for the detection via interactions with ordinary matter are discussed.

2.1 Evidence

The first indications for dark matter in chronological order were observed in the context of movements of astrophysical objects under the influence of gravity. Already in 1906, the term *dark matter* was used by Poincaré [1] for the description of objects in our galaxy not emitting or reflecting light. In 1922 by Kapteyn [2] and in 1932 by Oort [3], it was described that the mass of the visible matter does not suffice to explain the motion of objects in the milky way. It was Zwicky [4] in 1933 who performed an estimation of the mass and the velocity dispersion of galaxies in the Coma Cluster, based on the luminosity and the redshift of the galaxies. The comparison to the observed velocity dispersion showed a deviation that he explained with the existence of dark matter in a greater amount than visible matter. In the 1970s, direct measurements of rotation curves [5, 6] of several galaxies further supported the existence of a large amount of dark matter. The rotation velocity v at high radii r was observed to be at a constantly high value, instead of declining as predicted by Kepler's third law, following $v \propto \frac{1}{\sqrt{r}}$. The measured rotation velocity and its expected value without a dark matter contribution are shown in fig. 2.1 through the example of the spiral galaxy NGC 3198. The postulation of a dark matter halo with a dominant mass fraction pervading the galaxy and its surrounding can explain the missing decrease of the rotation velocity in the outer region of a galaxy.

Since these measurements were obtained, stronger evidence for the presence of an unknown type of non-luminous matter has been explored. One example is derived

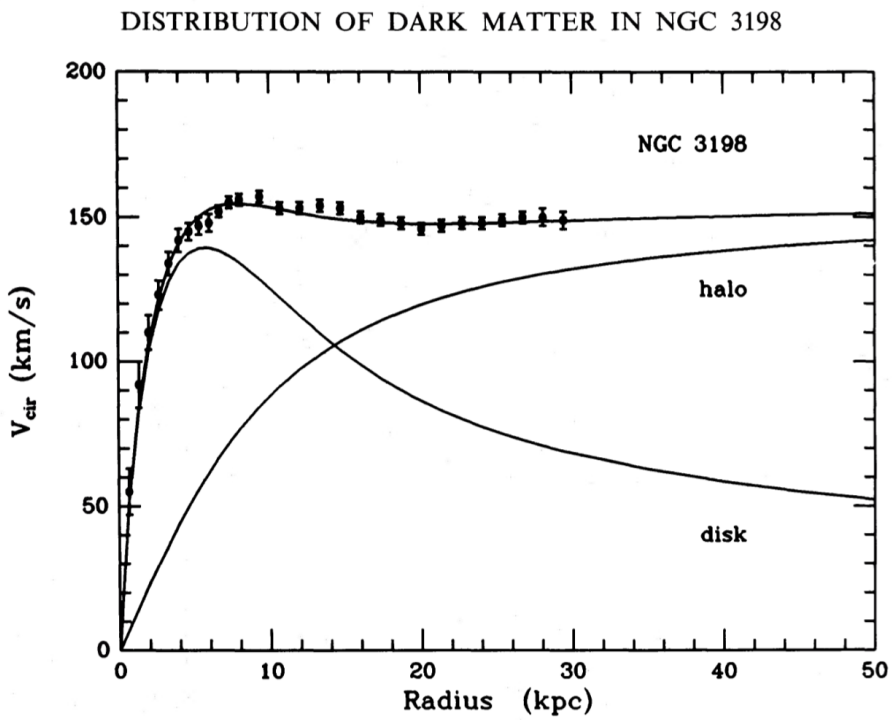

Figure 2.1: Rotation velocity curve using the example of the spiral galaxy NGC 3198 [7]. The measured rotation velocity for different radii (black points) is compared to the expected rotation curve without dark matter ("disk") and to the contribution of a dark matter halo. The rotation velocity expected from the visible disk of the galaxy is in disagreement to the measured curve.

from the cosmic microwave background (CMB), first reported in [8]. The CMB dates back to the recombination epoch, when the Universe was about 400 000 years old. The slight anisotropy of photons emitted at this time is connected to matter formation in the early Universe and is successfully described by the ΛCDM (Λ cold dark matter) model, which makes a statement about the fraction of baryonic and dark matter. Detailed measurements were conducted by WMAP [9] and the Planck collaboration [10]. Both used satellites for measurements of the anisotropies of the CMB. More information about the ΛCDM model and the CMB can be found in [10].

Another effect connected to dark matter was postulated by Einstein [11] and Zwicky [12]. Light is deflected by strong gravitational fields, therefore the presence of a massive object in the path between a light source and an observer leads to a deformation of the light source's image by gravitational lensing. This effect is stronger for more massive objects and can be used to reconstruct the gravita-

Figure 2.2: Pink: X-ray measurement of gas from two colliding clusters (Chandra telescope). Blue: Matter distribution calculated by gravitational lensing effects. In the background: Image from the visible light of the colliding clusters (Hubble telescope and Magellan telescope) [15]. The deviation of the pink and the blue distribution can be explained by two dark matter halos passing each other unaffected, while the luminous gas was decelerated by the collision.

tional potential responsible for the distortion. Measurements show that the mass reconstructed with the gravitational lensing is larger than the luminous matter found in the path (detailed information in e.g. [13]). A well-known example for gravitational lensing is the Bullet Cluster (1E 0657–558) [14]. Here, the collision of two galaxy clusters is observed. The centres of the matter distributions of the colliding clusters are investigated using the gravitational lensing effects and the emitted electromagnetic radiation. Both approaches lead to different results, suggesting that the collision strongly influenced parts of the luminous matter, while the dark matter halos of the galaxy clusters were able to pass each other mostly unaffected. A montage of the Bullet Cluster with the different measurements is shown in fig. 2.2.

The strong evidence for dark matter has made the idea of an unknown form of electromagnetically non-interacting matter become widely accepted, even though a direct detection of dark matter on earth is still missing. The observations are consistent with the concept of non-relativistic, weakly interacting massive particles (WIMPs) [16]. Other possible dark matter candidates (e.g. axion-like particles [17] or sterile neutrinos [18]) are not discussed in this thesis. The WIMP [19] is

described as stable, neutral and with a mass in the range from GeV/c^2 to TeV/c^2. Further, it has a low self-interaction and interacts with ordinary matter via gravitation and weak interactions. WIMPs are postulated to be generated by a freeze-out during the expansion of the early Universe, as the temperature of the plasma fell below the WIMP mass.

2.2 Detection channels for dark matter

Three different detection channels are possible for WIMP detection experiments [20]. One of them is the production in particle colliders and the detection in a disappearance experiment. A second one is the indirect detection. Here, signals from dark matter annihilations are searched, which are causing pairs of ordinary particles. The remaining channel, which will be in the focus of this thesis, is the direct detection: a WIMP scattering on a target creates a small recoil in the order of keV [21]. In order to detect these interactions, the signature of such recoils must be known in detail for the individual experiments. Following [21], the differential recoil spectrum can be written as

$$\frac{\mathrm{d}R}{\mathrm{d}E}(E,t) = \frac{\rho_0}{m_\chi m_A} \int v \cdot f(\vec{v},t) \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}E}(E,v)\mathrm{d}^3v, \tag{2.1}$$

where E is the recoil energy, ρ_0 is the local dark matter density, m_χ and m_A are the WIMP mass and the nucleus mass, respectively, v is the dark matter velocity in the detector's rest frame, $f(\vec{v},t)$ is the WIMP velocity distribution and σ is the WIMP cross-section. This cross-section is further discussed in [16] and [20]. It depends on the WIMP velocity v and the recoil energy E and is given by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}E} = \frac{m_A}{v^2\mu^2}\left(\sigma_{\mathrm{SI}}F_{\mathrm{SI}}^2(E) + \sigma_{\mathrm{SD}}F_{\mathrm{SD}}^2(E)\right), \tag{2.2}$$

where μ is the WIMP-nucleus reduced mass and σ_{SI} and σ_{SD} are two different contributions from spin-independent (SI) and spin-dependent (SD) interactions. $F_{\mathrm{SI}}(E)$ and $F_{\mathrm{SD}}(E)$ are the corresponding form factors, which are relevant for large target atoms with a high mass. For large momentum transfers, they account for the effect that the WIMP's particle wavelength can no longer be considered large in comparison to the nucleus' radius, suppressing the cross-section at large recoil energies. Under the assumption that the coupling strength of WIMPs to protons and neutrons is equal, the SI cross-section depends on A^2, where A is the

number of nucleons. For small momentum transfers, the scattering is a coherent process, resulting in the addition of the scattering contributions from each nucleon. In contrast, the cross-section of the SD interactions does not depend on the number of nucleons. Instead, it depends on the total nuclear spin and its spin structure, thus only atoms with an odd number of protons or neutrons contribute. Due to the A^2 dependence, heavier detector targets are in general more sensitive to SI interactions, even if the high recoil energies are suppressed by the form factor. Therefore, a low energy threshold is especially important for heavy detection targets [20].

In direct detection experiments, the recoil energy can be deposited in three different channels [16]. Since so far none of the experiments could measure all three of them simultaneously, each experiment is confined to one or two of those channels. The first channel is the relaxation of the recoil into atomic motion,[1] referred to as heat. The other channels are the ionisation or excitation of target atoms. In some materials, the deexcitation leads to detectable scintillation light. The lastly mentioned two channels are evaluated in the experiments of the XENON Dark Matter Project, which conducts a direct detection WIMP search employing a liquid xenon dual-phase TPC, as described in the following chapter.

[1] In solid detection materials, phonons are generated.

3 XENON experiments

There are several collaborations that search for the hypothetical WIMPs. A world leading limit on the WIMP-nucleon spin-independent elastic scattering cross-section of $4.1 \times 10^{-47}\,\mathrm{cm}^2$ at $30\,\mathrm{GeV}/c^2$ [22] was placed by the XENON1T experiment, which was part of the XENON Collaboration. The collaboration has conducted a row of experiments employing the noble element xenon. All of them were located underground at the Laboratori Nazionali del Gran Sasso (LNGS) in Italy, providing a shielding from cosmic radiation of 3600 m [23] water equivalent. In the following, the working principle and the two latest experiments XENON1T and XENONnT are described. XENON1T has completed its lifetime and has been upgraded to the successor XENONnT, which will start science data taking in 2021. Also, two signal corrections necessary for both experiments and calibration sources used for this purpose are introduced.

3.1 Time projection chamber

The core of all XENON experiments is a dual-phase time projection chamber (TPC). The TPC contains liquid xenon (LXe) used as detection material responding to various types of radiation and a gaseous xenon (GXe) phase at the top of the TPC. The cylindrical volume of the TPC is observed by two arrays of photomultiplier tubes (PMTs), located at the top and bottom of the cylinder.

The working principle of a TPC is shown in fig. 3.1. An incoming particle interacts either with the electronic shell of a xenon atom or with the nucleus itself, called an electronic or a nuclear recoil, respectively. This energy deposition starts an ionisation and excitation process.

LXe as a scintillator medium is able to produce both electrons and scintillation light as a response to incoming radiation. The production of scintillation light with a wavelength of 178 nm [24] is the result of both direct excitation of xenon atoms and recombination of electron-ion pairs. The emission of the scintillation light is in both cases coupled to the formation and the following decay of excited dimer states of xenon (Xe_2^*). The direct excitation of a xenon atom (Xe^*) induces

the process [24]

$$\begin{aligned} \mathrm{Xe}^* + \mathrm{Xe} + \mathrm{Xe} &\rightarrow \mathrm{Xe}_2^* + \mathrm{Xe} \\ \mathrm{Xe}_2^* &\rightarrow 2\mathrm{Xe} + h\nu. \end{aligned} \tag{3.1}$$

Recombination of electrons and ions (Xe^+) causes

$$\begin{aligned} \mathrm{Xe}^+ + \mathrm{Xe} &\rightarrow \mathrm{Xe}_2^+ \\ \mathrm{Xe}_2^+ + e^- &\rightarrow \mathrm{Xe}^{**} + \mathrm{Xe} \\ \mathrm{Xe}^{**} &\rightarrow \mathrm{Xe}^* + \mathrm{heat}, \end{aligned} \tag{3.2}$$

followed by the process outlined in eq. (3.1) for the de-excitation of the remaining excited atom. As given in eq. (3.2), the recombination process converts a small amount of the deposited energy into non-measurable signals referred to as heat. This is a channel that cannot be observed by the TPC.

The scintillation light promptly reaches the PMTs, where it is registered as a light signal. Due to the different refraction indices of GXe [25] and LXe [26], a large part of the light is reflected at the interface and is detected at the lower PMT array. This prompt light signal is referred to as S1. In order to suppress recombination and to separate the ionisation electrons, an electric drift field inside the TPC is used. The drift field is applied between the cathode, located closely above the bottom PMT array, and the gate, which is submerged in the LXe close to its surface. Spanning the GXe phase, between the gate and the anode another, a stronger electric extraction field is applied. Starting at the interaction site, the ionisation electrons are drifted towards the gate. Their drift velocity depends on the electric field strength [24]. The drift time through the TPC is in the order of a few µs. As soon as the electrons reach the gate, they are accelerated by the stronger extraction field in the gas phase. On their way to the anode, they ionise and excite further GXe atoms and trigger an avalanche effect. The light that is emitted by this cascade is proportional to the amount of ionisation electrons and is then registered as the charge signal, which is referred to as S2. The amount of separated electrons depends on the strength of the drift field. The more electrons are separated, the less recombinations are possible. Therefore, light and charge signal are anti-correlated [27].

An important feature of the dual-phase TPCs is the ability for a full three-dimensional position reconstruction. The x- and y- coordinates are determined by

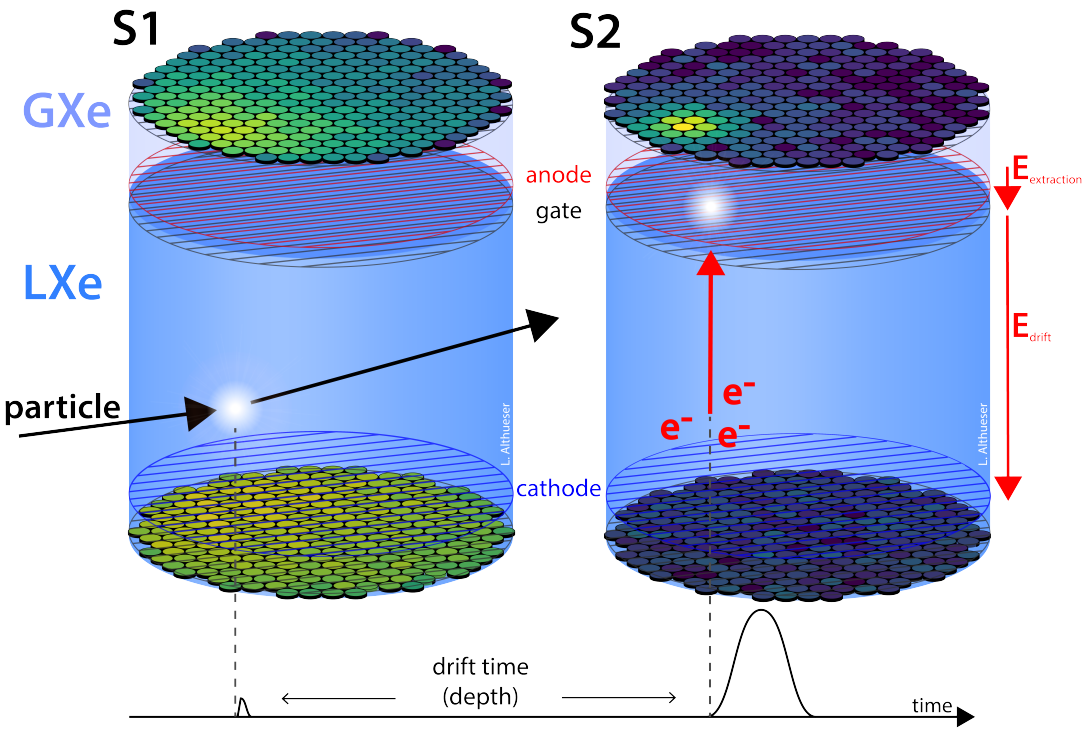

Figure 3.1: Working principle of a time projection chamber with the generation of a light signal (left) and a charge signal (right) [28]. The incoming particle interacts with the LXe shown in blue, creating scintillation light and ionisation electrons. The primary scintillation signal is immediately registered by the PMTs at the bottom and top of the TPC (S1). The ionisation charge is drifted towards the anode by an electric drift field E_{drift} and extracted into the gas phase (light blue) by the stronger extraction field $E_{\text{extraction}}$, leading to an avalanche effect. This secondary scintillation is registered as the S2. A three-dimensional position reconstruction is performed with the S2 top pattern (x-, y-coordinates) and the drift time between S1 and S2 (z-coordinate).

the intensity distribution of the S2 on the top array, since a homogeneous drift field projects the position of the interaction site onto the top PMT array. For the z-coordinate, the drift time between S1 and S2 is evaluated. Since the electrons move with a constant, field-dependent velocity [24] through the drift field, the relation between drift time and depth can be used to calculate z. The combination of S1 and S2 then allows to reconstruct the deposited energy [29].

Another advantage of the dual-phase TPC is the ability to discriminate between nuclear and electronic recoils [22]. Since most of the backgrounds are electronic recoils, while the possible WIMP signal would be a nuclear recoil, a good discrimination is important. This discrimination is based on a different ionisation density

for electronic and nuclear recoil events, which leads to a different ratio of ionisation and scintillation [24]. Therefore, the ratio of S2 and S1 is different for electronic and nuclear recoils.

The choice of xenon as a detection material has several advantages [24]. As mentioned previously, it generates both light and charge carriers as a scintillation medium. Being a noble gas, it does not react with impurities inside the TPC. Furthermore, the high atomic mass of xenon leads to a high stopping power for external γ- and β-radiation and allows the detection material to act as its own shielding. By defining a so-called fiducial volume inside the LXe of the TPC, the rate of external events, for example induced by decays in the detector materials, can be reduced drastically. By rejecting these external events, the low-background requirement of the dark matter search can be reached. The presented methods are employed by the two experiments XENON1T and XENONnT, which will be introduced in the following sections.

3.2 XENON1T

XENON1T [23] finished the commissioning phase in late 2016 and stopped taking data in December 2018. An illustration of the XENON1T TPC is shown in fig. 3.2. The TPC had a diameter of 96 cm and a height of 97 cm [27]. The coordinate system was chosen in a way that the gate lies at a value of $z = 0$ cm. The anode was located 5 mm above, while the cathode was at -96.9 cm [23]. Top and bottom PMT arrays were shielded from the electric field by the screening electrodes. All five electrodes consisted of meshes with wire diameters of $\sim$200 µm, allowing a high optical transparency for the scintillation light. In total, the TPC was equipped with 248 PMTs [23], whereof 127 were placed in the top array and the remaining 121 in the bottom array. The PMTs were manufactured by Hamamatsu and had a tube diameter of 76 mm each [31]. The stable level of the liquid-gas interface is ensured by a stainless steel diving bell, containing the top PMT array. The wall of the TPC was made of PTFE (polytetrafluoroethylene) panels, which possess a high reflectivity [32] for photons with the scintillation wavelength of 178 nm. Those panels were surrounded by 74 field shaping electrodes in order to ensure a homogeneous drift field. The TPC was housed in a cryostat were the LXe was kept at an operating temperature of -96 °C. The cryostat was contained in a water tank for further background reduction. The water tank was monitored by an active Cherenkov muon veto detector [33]. The infrastructure providing the

cooling power for the cryostat and for a purification system was housed in an adjoining service building. A detailed description of all detector subsystems can be found in [23]. In order to supply these systems with xenon, a total amount of 3.2 t was needed, whereof 2.0 t were contained inside the TPC as an active target.

3.3 XENONnT

XENONnT is the upgrade of the XENON1T experiment. By the start of 2021, it is under commissioning at LNGS. Employing the same working principle as its predecessor, it is able to reuse a large part of the XENON1T infrastructure.

3.3.1 Detector

The TPC of XENONnT [30] has a diameter of 1.33 m and a height of 1.49 m. It is equipped with 253 PMTs in the top array and 241 PMTs in the bottom array. With a new cryostat, the TPC is housed in the existing water tank, where it is connected to the cryogenic infrastructure of XENON1T. The background reduction is enhanced with an additional active neutron veto, for which the water will be loaded with gadolinium [30]. The Gd captures neutrons and emits a γ-ray cascade, which is then transformed into Cherenkov light and used as a veto signal. Also, a

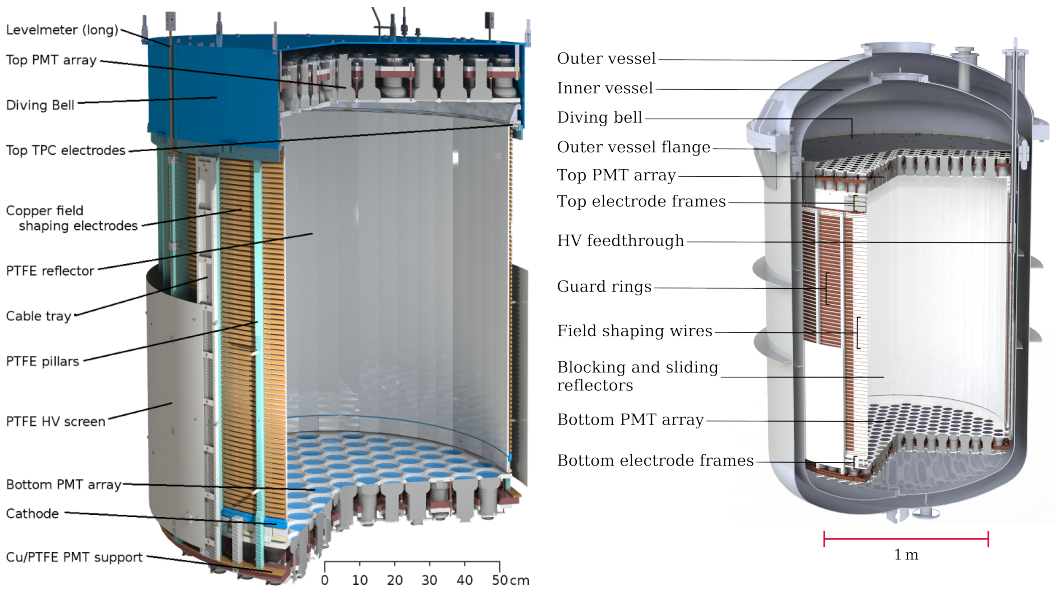

Figure 3.2: Illustration of the XENON1T TPC [23] (left) and of the XENONnT cryostat containing the TPC [30] (right). Note that the right figure is scaled down by a factor of approx. 2 in all dimensions compared to the left figure.

new radon removal system [34] and a liquid purification [30] are installed in order to reduce impurities in the detection material. The active LXe target has a mass of 5.9 t, while a total of 8.4 t is used to run all adjoining systems. XENONnT is projected to reach a sensitivity to spin-independent WIMP-nucleon interactions of $1.4 \times 10^{-48}\,\mathrm{cm}^2$ for the cross-section of a $50\,\mathrm{GeV}/c^2$ WIMP at 90 % confidence level [30], which is more than one order of magnitude below the current limit of XENON1T. An illustration of the XENONnT cryostat housing the TPC is shown in fig. 3.2.

3.3.2 Analysis framework

Apart from a larger target mass, XENONnT also utilises a new data processor. The novel analysis framework is called *streaming analysis for XENON* (straxen) [35]. Together with the more generic framework *streaming analysis for xenon TPCs* (strax) [36], it provides the algorithms that are needed for processing and analysing the data recorded by XENONnT. The naming already suggests the concept of a streaming data analysis. The processing starts before all data are acquired, as soon as the data taking begins. This causes a flow of data that is further processed on its way. This flowing property is referenced in the labelling and allows a live data processing.

The strax framework contains algorithms that are generic enough to be employed by any noble liquid TPC. Its basic assumption is that all data have a specific data type with a timestamp. In order to work with the various data types, each data type has a matching numpy [37] representation. If two data types describe the same physical property, they are of the same data kind. Based on their timestamp, strax divides the incoming data into chunks, each a self-contained interval of data. The chunking allows to conduct the data processing as a true streaming analysis. Also, strax gives the framework for the definition of contexts and plugins. In strax(en), a context contains information about how the data are processed. This includes where to load the data from and where to store the results. Additionally, a context contains details about the various data types which a plugin provides and its default configuration options. A plugin is a self-contained step of the data processing. It has to declare its output variables and data kind as well as the plugins it depends on itself, which is illustrated in fig. 3.3. A plugin consists of the code that is required for the determination of its output. This code can be modified modularly as an independent section. In summary, strax provides the

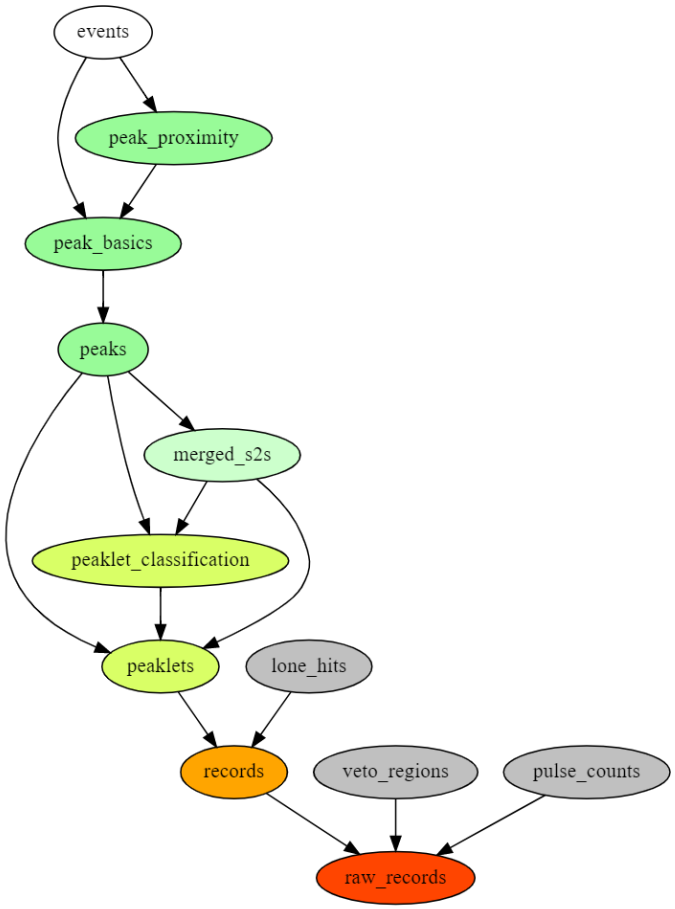

Figure 3.3: Succession of data kinds and plugins in the data flow of straxen [38]. Each data kind has its own color, while each item stands for a plugin. The arrows indicate the dependencies between the plugins. The data processing starts at the bottom with the data kind `raw_records` and ends at the top with `events`. After the plugin `events` there are several other plugins that follow the highest one that is listed here. These are used for high-level analyses without generating a new data kind.

basic environment for data classification without actually defining the individual routines and specific information that are used for XENONnT.

Straxen configures the concepts given by strax for XENONnT. Therefore, straxen contains contexts for various operation modes, for example commissioning of the TPC, regular online data taking or reprocessing of XENON1T data. These use cases also require different sets of plugins. Since each plugin provides a certain data type while itself depending on the previous analysis stage, a dependency tree is constructed. As mentioned earlier, it is possible to model and reorder these dependencies, so the exact configuration changes over time. Still, the overall suc-

cession of data kinds stays the same. Besides providing plugins for the TPC data analysis, straxen also contains plugins for the other sub-detectors of XENONnT: the muon veto and the neutron veto. In the following, the data flow for the TPC is explained step by step up to the event level, which is the data kind that will be used for the analysis later in this work.

The dependency tree for the analysis of TPC data is shown in fig. 3.3. Here, the data kinds are marked by different colours, while each plugin has its own item. The arrows mark the dependencies between the plugins, so that the highest level plugin is at the top, and the data processing starts at the bottom. The processing begins with the output of the data acquisition system and a plugin that provides the data kind `raw_records` by reading the incoming data. At this stage, the data are in the shape of pulses given separately by each PMT channel. This is the fundamental data kind that all other plugins are built on. In a next step, the processing begins with the data kind `records`. Here, the baseline for each pulse is calculated by evaluating the first samples of a record. The baseline value is then subtracted from all samples. Afterwards, the samples are multiplied[1] by -1. Additionally, some data rate reduction techniques are applied, like an optional software high energy veto and a first hit finding with the goal to set the samples to zero which are not contained in a hit. The hit finding algorithm iterates through the incoming data samples and checks if the current sample is above a given threshold. Consecutive samples above this threshold are summarised in a hit. At this stage, the determined hit properties are not considered further. In fig. 3.4, an overview of the classifications that are used in this section is depicted. The pulses that were discussed so far are shown in blue.

The next data kinds are devoted to the identification of individual incidents, called `peaks`. A peak can be S1, S2 or unidentified. In order to reach this classification, in a first step the hit finding algorithm is repeated. The resulting hit is defined by its starting and ending sample and by its duration in time. The duration is calculated by the time per sample. Also, a hit contains the integration of the samples over time. This property is called area and uses the unit photoelectrons (PE) to measure the size of a hit. The number of photoelectrons corresponds to the number of electrons that are ejected from the photocathode and is therefore a measure of the energy deposition on the interaction site for a particular hit.

Then the hits are passed to the peak finding algorithm. While a hit still is

[1] called flipping the data

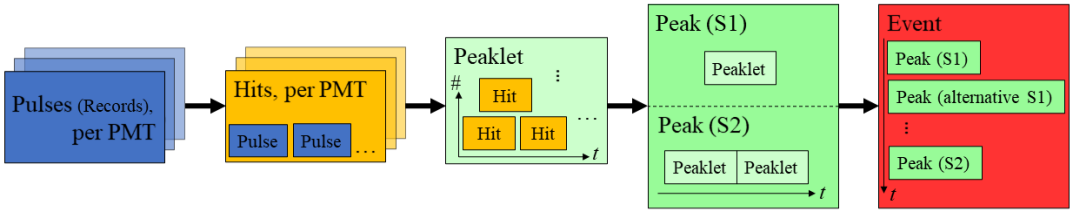

Figure 3.4: Illustration of the data flow in straxen. Each data kind is marked with a different colour. Pulses and hits are recorded for each PMT individually, which is marked by the transparent contours. Peaklet is the first combined step, where the different channels are summarised, as indicated by the axis marked with #. If required, the peak reconstruction merges separated S2 peaklets after the classification. In a last step, the peaks are assembled into the highest data kind of events.

defined on the PMT level, a peak summarises the different channels. This is visualised in fig. 3.4, where the hits are shown in yellow and the peak-related data kinds are shown in green. A peak has similar properties like a hit, but is required to contain a complete S1 or S2 (or unknown signal). During the peak construction, the intermediate data kind `peaklet` (for small peak) is constructed for a first estimation of the peak properties.

A peaklet starts as a group of hits that occur in a time range which is below a configurable threshold value, for example 350 ns for the reprocessed XENON1T data. As soon as this time span has passed without a hit in any PMT, the peaklet candidate is completed. This simple division does not always resolve peaklets that are very close together, therefore they are split further. The splitting algorithm calculates a goodness of split for each sample of the peaklet. This goodness of split is calculated by a modified version of the Jenks/Fischer natural breaks splitting algorithm [39]. If a threshold value is exceeded, a split is done at the maximum value. This procedure is repeated until either the peaklets are small enough, the threshold is not exceeded anymore or a recursion limit is reached. This splitting strategy is strict enough to ensure that a peaklet contains only one S1 or S2. But the other way round it is possible that an S2 is split into several peaklets. Therefore, at the end of the peakfinding algorithm, there is a merging step for S2s that are distributed over more than one peaklet.

After the splitting of the peaklets, the next plugin classifies them. The variables that are used for the classification are called rise time and tight coincidence. The

rise time is defined as the time that is required to reach 10 % of the peaklet's area. A short rise time is an indicator that the peak is the result of a light signal (S1), since the photons travel through the detector with the speed of light and reach the PMTs with a very short delay time. A signal that is caused by electrons (S2) exhibits a higher rise time. Both the diffusion during the electron drift and the S2 creation in the GXe cause the broadening. The second classification variable is the tight coincidence. The idea is to distinguish between the number of contributing PMTs in a time window around the maximum of a peaklet's waveform. An S1 or S2 should exceed a certain number of contributing PMTs in order to be classified as valid.

The exact values for these requirements can be fine tuned with the XENONnT data. For the reprocessed XENON1T data, the shorter rise time necessary for a peaklet to be classified as an S1 is set to a maximum value of 60 ns if the peaklet has an area smaller than 100 PE. Larger peaklets have a maximal allowed rise time of 150 ns. The second requirement for an S1 classification is the tight coincidence of 3 or more PMTs in a time window of 100 ns around the maximum. The remaining peaklets are classified as S2s if they exceed a tight coincidence value of 4 PMTs in the 100 ns window. Everything that is left is then classified as unknown.

At this point, only the S2 merging is left before the peakfinding is completed. Due to the aggressive splitting algorithm, some of the signals caused by electrons with their longer duration in time are cut apart into different peaklets. The S2 merging algorithm compensates that by merging consecutive peaklets. This is shown in fig. 3.4 in the dark green box. The merging is based on several criteria: in order to trigger a merging, the potential peak has to start with a peaklet that was previously classified as an S2. Also, the hits that are formed into a peak must not be separated by gaps larger than 3.5 µs. The merging of a peak is not allowed to result in an area of 5000 PE or higher. Also, a time limit for merged peaks of 15 µs is set.[2] With an additional plugin, the position is reconstructed, as described in section 3.1.

The final stage of data processing is the accumulation of peaks into the data kind of `events`. An event is required to contain a whole signature of a decay or interacting particle, namely the main S1 and S2 peaks and possible alternative peaks. It consists of a time range around a so-called triggering peak and all the peaks that are inside this time range. For the reprocessing of the XENON1T data,

[2] Of course, a peaklet without a merging step can surpass 5000 PE or a duration of 15 µs.

an event is extended 1 ms to the left and to the right of the triggering peak. The classification as a triggering peak is based on the peak's area and the number of other peaks that are close to the candidate. This information is provided by the `peak_proximity` plugin. In order to trigger an event, a peak has to exceed a certain area, which is set to a default value of 100 PE. For comparison, a single electron in straxen is observed to create between 20 PE and 30 PE [40]. Furthermore, the triggering peak must not reach a certain number of competing peaks, where the default limit is set to 7. After the assignment of the peaks to events, each event is further evaluated. This happens in several plugins of the data kind events which are not pictured in fig. 3.3. These plugins determine the main S1 and S2 in the event as well as an alternative S1 and S2. Main and alternative S1 and S2 are the largest and second largest S1 and S2 peaks of an event. For these peaks, several features are calculated. Next to peak areas, positions, time between main and alternative peak and many other variables that are provided by customisable plugins, several corrections are applied on the event level. Two of those corrections are explained in the following section.

3.4 Signal corrections

As explained in 3.1, the detector measures the light signal (S1) and the charge signal (S2). Both signals are influenced by effects that add a systematic, position dependent deviation to the signal size. In this section, some of these effects are identified and corrected for both S1 and S2.

3.4.1 Position dependent light correction

If an energy deposition in the TPC creates scintillation light, not all of the generated light is registered by the PMTs as part of an S1. Instead, part of it is lost inside the detector due to absorption by the TPC walls or impurities. This leads to a position dependent variation of the light yield LY [41]. The LY is defined as the light signal S1 per incident energy E, where $S1$ will refer to the area of the light signal. The LY is affected by the photon yield Y_{Ph}, which describes the number of photons that is generated per incident energy. The photon yield itself depends on the electric field strength F and the incident energy, since both influence the recombination and therefore the production of scintillation light. If the field was not homogenous throughout the TPC, this would lead to an indirect

position dependency. Further, the LY depends on the light collection efficiency ϵ_{L}, which describes the number of photons that hit a PMT per photon generated at the interaction site. ϵ_{L} takes into account effects that are based on the detector geometry and introduces the direct position dependency into the light yield. Lastly, the LY depends on the quantum efficiency ϵ_{QE}, which is the probability for one photon hitting the photocathode of a PMT to create a photoelectron, and the collection efficiency within a PMT ϵ_{CE}, which gives the probability that this first photoelectron succeeds to reach the first dynode. Since all PMTs are taking the data simultaneously, the average ϵ_{CE} and ϵ_{QE} of all PMTs contribute to the light yield. All these factors together result in [41]

$$LY = \frac{S1}{E} = \epsilon_{\mathrm{L}}(r, z, \varphi) \cdot Y_{\mathrm{Ph}}\left(E, F(r, z, \varphi)\right) \cdot \epsilon_{\mathrm{QE}} \cdot \epsilon_{\mathrm{CE}}. \tag{3.3}$$

Under the assumption of a constant electric field, the only position dependent factor is ϵ_{L}. The S1s spatial response can be homogenised by the observation of the relative S1 size from events with a uniform energy deposition on different positions in the detector. This yields the so-called LY map [41] and provides a correction factor for every location inside the TPC. The LY map is a data-driven approach and therefore needs data from a calibration source. A suitable calibration source requires a constant energy deposition per event and is required to be mixed with the LXe in order to provide homogeneously distributed events in the whole TPC.

For the creation of the LY map, the TPC is divided into small volumes in the z-, r- and φ-directions, which are called bins. Bins with the same z-coordinate form a slice. With the r-coordinate, the slice is divided into a number of annuli N. The number of bins on each annulus increases towards the outermost annulus to provide an equal volume per bin. For a homogeneously distributed calibration source, this leads to an approximately constant number of events in each bin. The number of bins n_{N_R} per annulus N_R is given by

$$n_{N_R} = 2N_R - 1, \tag{3.4}$$

keeping the volume of each bin constant for annuli with a constant thickness. The number of slices and annuli does not have an inherently defined value, but is adjusted to the available calibration statistics. A higher number of bins leads to a better resolution of the calibration, but also requires enough statistics to perform

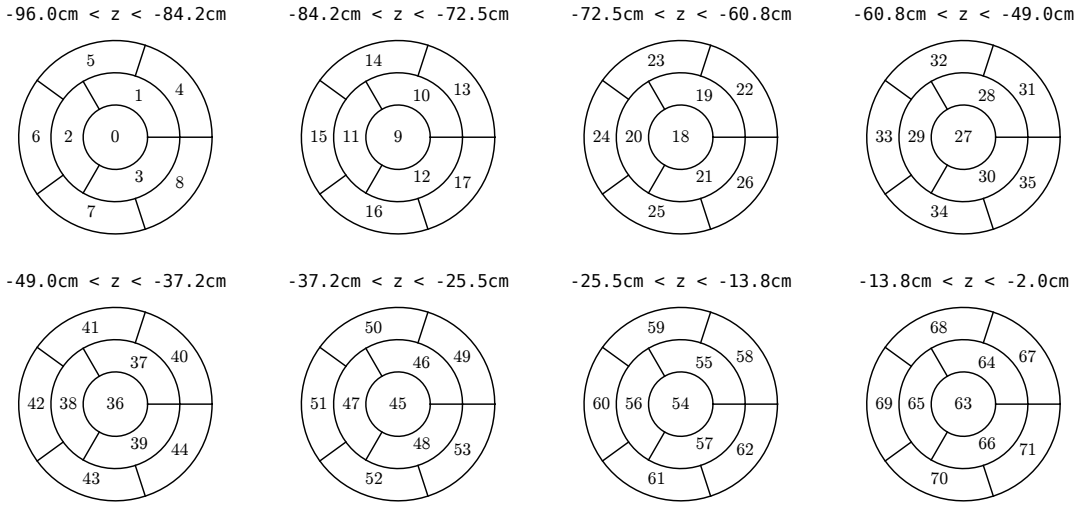

Figure 3.5: Division of the TPC into spatial bins in r, z and φ. Every circle references a slice in z-direction. For further reference, the bins are numbered outwards in r, counter-clockwise in φ and upwards in z.

a fit in each bin. A binning example is shown in fig. 3.5, where the number of slices is set to eight and the number of annuli is set to three. This leads to nine bins per slice and to an overall number of 72 bins for the whole TPC. The bins are numbered going outwards in r, counter-clockwise in φ and upwards in z.

The data from the calibration source are sorted in the bins by their coordinates. In each bin, a fit of the distribution of the S1 area is performed, determining the respective mean of the S1 area. Divided by the energy deposited by the calibration source, this gives the absolute value for the LY. In order to correct the S1 data, a relative LY value is required [41]. The obtained S1 areas are normalised with the mean of all areas, which gives the position dependent correction factor L_c:

$$L_c(r,z,\varphi) = \frac{LY(r,z,\varphi)}{\langle LY \rangle} = \frac{S1(r,z,\varphi)}{E} \cdot \left\langle \frac{S1}{E} \right\rangle^{-1} = \frac{S1(r,z,\varphi)}{\langle S1 \rangle}. \tag{3.5}$$

Together with the coordinates of the bins, the result is combined to a LY map and saved as a `.json` file. It can then be used to correct the S1 data of a given dataset. The pre-defined package `InterpolatingMap` [42] is used for interpolation of the discrete data points that are given by the binning of the TPC. Its interpolation algorithm is based on inverse-distance weighted averaging. In this way, the continuous version $\tilde{L}_c(r,z,\varphi)$ of the correction factor is calculated for every possible

S1 and its position (r, z, φ). The corrected value $cS1$ is then given by

$$cS1 = \frac{S1}{\tilde{L}_c(r, z, \varphi)}. \tag{3.6}$$

3.4.2 z-dependent charge correction

The S2 signal's area needs to be corrected [41], since it is not reconstructed as constant over the height of the TPC. On their way through the liquid xenon, some of the S2 electrons are captured by electronegative impurities, leading to a smaller S2. As the electrons reach their saturation speed in a short time compared to the overall drift time, they spend most of the drift time at a constant velocity. This means that the probability to be captured is also constant during their drift. A population showing a constant capture probability decreases exponentially over time. Therefore, the number of electrons and thereby also the signal size $S2$ is connected to the drift time t via

$$S2 = S2_0 \cdot \exp\left(-\frac{t}{\tau}\right), \tag{3.7}$$

where τ is the electron lifetime in the detector and $S2_0$ is the original signal size at the interaction site. An event from the lower part of the TPC has a longer drift time and therefore a smaller S2 than an event from the top of the TPC. If the electron lifetime τ is known, it is possible to correct an individual S2 by extrapolating it to its original size, depending on the associated drift time. Solving eq. (3.7) for the original signal size $S2_0$ provides the necessary extrapolation:

$$cS2 = S2_0 = \frac{S2}{\exp\left(-\frac{t}{\tau}\right)}. \tag{3.8}$$

The electron lifetime can be obtained by an exponential fit to a two-dimensional histogram of the drift time versus the S2 area. Again, the calibration source needs to be distributed homogeneously throughout the TPC and needs to possess a mono-energetic decay.

The electron lifetime correction is not the only S2 correction. For the calculation of the cS2 area, an additional x-, y-correction [43] is required, which is not considered in this work. The corrected effects are caused by factors like the bending of anode and gate electrode and different quantum efficiencies of the PMTs.

3.5 Calibration sources

There are several calibration sources that were used by XENON1T and that will be used for XENONnT. As described above, a good calibration source has to fulfil several requirements. Since a homogeneous distribution in the TPC requires to mix the source with the LXe, it is essential that it is easy to remove as soon as the calibration is done and it should create decays with a constant and known energy deposition.

3.5.1 Calibration with krypton

A calibration source that fulfils these requirements and that is commonly used is ^{83m}Kr, a metastable isomer of ^{83}Kr, with a half-life of 1.83 h [44]. The metastable state is at an energy level of 41.6 keV above the ^{83}Kr ground state and decays mainly via internal conversion. The isomer ^{83m}Kr is obtained by an electron capture of its mother isotope ^{83}Rb, which has a half-life of 86.2 d. Some of its decay modes result in the desired metastable state of krypton via further, short-lived excited states. A simplified decay scheme of ^{83}Rb is shown in fig. 3.6. The figure shows that the transition of ^{83m}Kr into the ground state is a two-stage decay with an additional short-lived excited state with a half-life of 154 ns at an energy level of 9.4 keV above the ground state. Therefore, the overall decay consists of a 32.2 keV decay that is quickly followed by a 9.4 keV decay.

For the purpose of calibration in dual-phase xenon TPCs, this metastable state has nearly optimal properties. Its relatively short lifetime prevents it from remaining active after the calibration is finished. Still, it is simple to supply the volume of LXe with ^{83m}Kr. While the ^{83}Rb is placed in zeolite beads [46], the produced ^{83m}Kr emerges as a gas and easily mixes with the xenon. If one of the excited krypton atoms decays inside the liquid phase of the TPC, it produces electrons and photons that are registered as S1 and S2. The time scale of the half-life of the lower excited state is similar to the resolution that is used in section 3.3.2 in the peakfinding algorithm. Because of that, the photon signal of the decays with a short time difference cannot be resolved as two different peaks. In that case, the two decays are merged into one main peak with an area that is equivalent to an energy deposition of the total 41.6 keV. Decays with a time difference that is considerably longer than the half-life are registered as an event with two separate S1 peaks, leading to three possible sizes of ^{83m}Kr S1 peaks. The electron induced

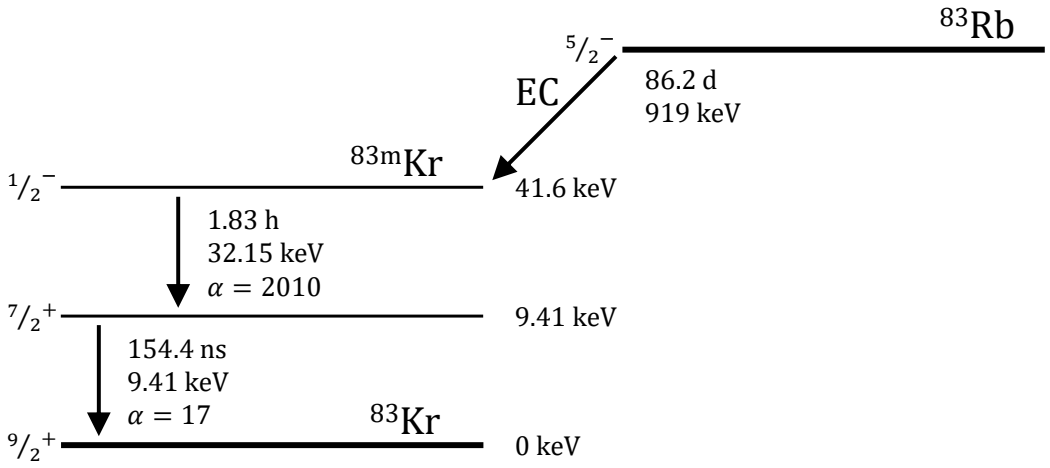

Figure 3.6: Simplified decay scheme of ^{83}Rb, containing the transitions that are used for calibration. The metastable state ^{83m}Kr with an energy of 41.6 keV above the ground state is populated by an electron capture of ^{83}Rb. The state decays with a lifetime of 1.83 h via one short-lived excited state at 9.4 keV. This second decay has a half-life of 154 ns ([45], adapted from [44]).

signal has a different signature. Due to the extraction and to diffusion of the electron cloud during the drift, the S2s have a higher width in time. While the S1 has a duration of less than 0.5 µs, the S2 has a duration of about 10 µs [23]. Since this is two magnitudes above the half-life of the short-lived 9.4 keV transition, the two states of ^{83m}Kr are not resolved for the S2. Instead, the electron signal of ^{83m}Kr is nearly always merged into one 41.6 keV peak.

3.5.2 Calibration with argon

Another similar calibration source used in XENONnT and XENON1T is ^{37}Ar. The isotope ^{37}Ar has a half-life of 35.0 d and decays directly into the stable ground state of ^{37}Cl via an electron capture [47]. Its simplified decay scheme is shown in fig. 3.7. The main part of the decay energy from the electron capture is transferred to the generated neutrino and it leaves the detector without creating a signal. Instead, the energy deposition which is used for the calibration comes from the de-excitation of the atomic shell after one electron from an inner shell was captured. In most cases, the captured electron is taken from the K shell. In 8.90 % of the decays, it is captured from the L shell, while the rest is provided by the M shell or higher shells. The resulting hole in one of the inner atomic shells is filled by an electron from an outer shell. This is accompanied either by the emission of X-rays or of

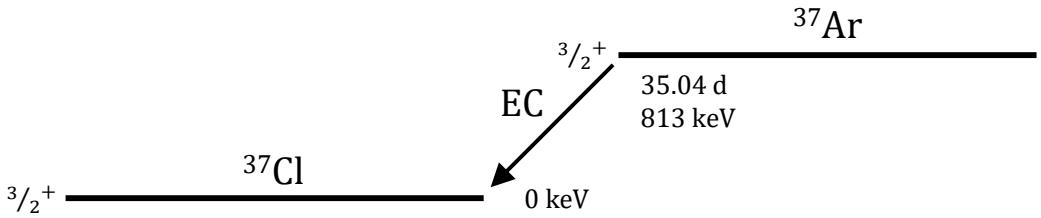

Figure 3.7: Simplified decay scheme of ^{37}Ar, which decays via an electron capture with a Q value of 813 keV into the ground state of the stable ^{37}Cl. It has a half-life of 35.04 d. Since the energy that is released by the primary decay is carried by the emerging neutrino, the energy deposition that is visible for the detector is created by the de-excitation of the atomic shell ([45], adapted from [47]).

auger electrons, which then create electron and photon signals in the detector. An overview over the energy depositions and the branching ratios for the K, L and M shell is given in table 3.1. Since the capture of a K shell electron is the most frequent decay, the observed calibration peak has an energy deposition of 2.8 keV.

As a gaseous source, ^{37}Ar is directly inserted into the xenon gas system. The isotope ^{37}Ar is synthetic and can be produced by neutron capture [47]. Both being noble gases, it has similar properties as ^{83m}Kr and is homogeneously distributed in the TPC. Due to its longer half-life it needs to be removed by distillation [48] and thus is not as flexibly usable as ^{83m}Kr, but offers a very low-energetic calibration source. Other than the krypton source, ^{37}Ar has a one-stage decay with a mono-energetic peak (considering only the K shell capture).

Table 3.1: Branching ratios for the ^{37}Ar electron capture from K, L and M shells with their corresponding energy depositions [47].

Decay Mode	Branching Ratio (%)	Energy deposition (keV)
K capture	90.17	2.8224
L capture	8.90	0.2702
M capture	0.93	0.0175

4 Signal corrections with ^{83m}Kr

One of the most important calibration sources in XENON1T was ^{83m}Kr, and it will play an equally important role for XENONnT. Therefore, the first ^{83m}Kr calibration of the new detector was performed during the commissioning phase. In order to prepare for the science runs (SR) of XENONnT, data from XENON1T are used for the implementation of the analysis code in straxen. In this chapter, the two introduced corrections are calculated and implemented with ^{83m}Kr data.

4.1 Data selection

With the introduction of straxen, 28 one-hour periods of data taking, called runs, with an open krypton source from XENON1T were reprocessed with straxen. These runs were taken at the beginning of SR1 of XENON1T, which corresponds to early February of 2017.

The data processing in straxen is divided into several levels, as discussed in section 3.3.2. The position dependent signal correction is an analysis that uses highly processed data, namely the data kind of events. An event ideally contains one ^{83m}Kr decay with its associated peaks. For the krypton analysis, a plugin named `event_info_double` is used. This plugin is optimised to process events with multiple physical S1s and S2s. It provides information about the largest peak as well as the second largest peak, for both light and charge signal. The main peak is abbreviated with the letter 'a', the second largest peak with 'b'. In total, the plugin provides information about S1a, S1b, S2a and S2b for each event. The dataset used in this chapter has an overall size of 6.26×10^6 events.

The krypton runs do not only contain the desired krypton events, but also events from other sources as well as misclassified krypton events as background. The definition of a suitable dataset is a necessary preparation for the analysis. Further event parameters provided by straxen are used to create several requirements on the data with the goal to sort out the background data. These requirements are grouped into so-called cuts. The S1a versus the S2a of each event of the whole available dataset without any selection is plotted on a logarithmic scale in a two-

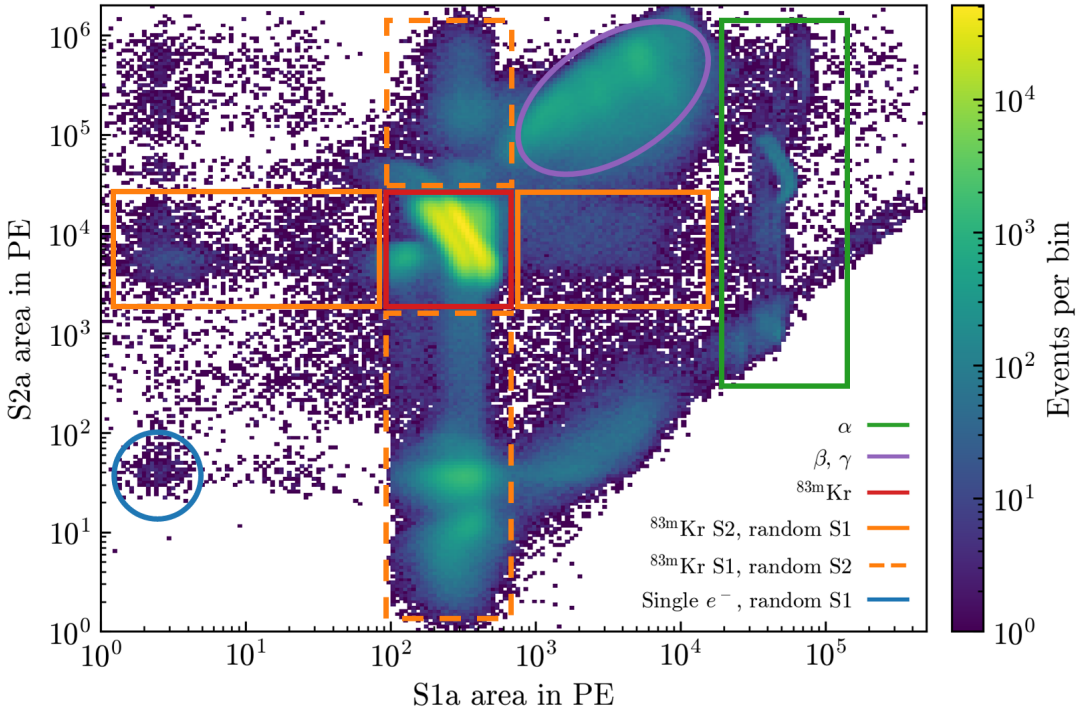

Figure 4.1: Primary S2 versus primary S1 area for the complete 28 runs with krypton data from SR1 of XENON1T that are available for straxen. No cuts and no corrections were applied. Several populations are highlighted.

dimensional histogram in fig. 4.1. The S1a area is mainly between 100 PE and 10 000 PE. The sharp decrease of events with less than 100 PE is caused by the peak classification described in section 3.3.2, where the value of 100 PE is chosen as the border between two rise time requirements. The S2 areas are usually larger than the S1s and can reach up to 10^6 PE. In fig. 4.1, the region with the most events per bin is marked with a red box. It contains the correctly reconstructed ^{83m}Kr events. Several further regions are marked and labelled with their corresponding origin.

As specified in section 3.5.1, ^{83m}Kr has two decays which are separated by a half-life of 154 ns. These consecutive decays produce two light signals in a quick succession. For some of the events, these two peaks are combined in one large peak. In this case, the complete area of the reconstructed scintillation light signal is stored in S1a and S1b has a vanishing area or is occupied by noise. The other possibility is a successful resolution of the two light signals. Here, S1a includes the reconstructed light signal of the first, 32.2 keV decay, while the light signal that is generated by the remaining 9.4 keV is stored in S1b.

This is different for the S2s. As described earlier, they have a width on a higher order of magnitude as the half-life itself. Thus, S2a and S2b are reconstructed as one merged signal in the vast majority of events, which is then stored in the S2a variable. Therefore, the S2a of an event will simply be referred to as its S2.

Due to the two possibilities for S1a ^{83m}Kr signals and the combined S2, the plot of S1a versus S2a contains two populations inside the red box which are close to each other in S1a and at the same S2 values. The entries of the left population, consisting only of the signal from the 32.2 keV decay are plotted at a lower area, since fig. 4.1 does not display the S1b area. The anti-correlation of S1 and S2 gives the two populations a tilted elliptical shape on the logarithmic scale.

Further events with a similar S1 area (marked with a dashed orange line in fig. 4.1) have a misreconstructed S2. For example, the short-lived second decay of ^{83m}Kr can be misidentified and matched as the corresponding S2. Such misreconstructed events will be removed by the cuts that are explained in the next subsections. Many of them are based on considerations in [49] and use the same variables. The cuts that are listed as part of the preselection aim to remove the majority of events not induced by ^{83m}Kr in general and are valid for both an S1a from a 41.6 keV peak (single S1) and from a 32.2 keV peak (double S1). Afterwards, the focus is on the separation of single S1 and double S1 events.

4.1.1 Preselection

At the level of `event_info_double`, one correction was applied to the data so far, which is the field distortion correction [43]. This correction aims to compensate the effect of a distorted electric drift field in the TPC on the position reconstruction. Events at higher radii tend to be reconstructed inwards with respect to the actual interaction site. Electrons which have to pass most of the TPC during their drift to the anode have a longer exposure to the distortion of the applied drift field and therefore reach the gate with a larger deviation. Since the x- and y-position reconstruction is based on the S2, this especially applies to events from the lower part of the TPC. The field distortion correction takes this shift into account and corrects the x- and y-position.

TPC boundaries cut

The field distortion corrected positions are then used for the first cut of the ^{83m}Kr data selection. Events that have a reconstructed position outside of the TPC

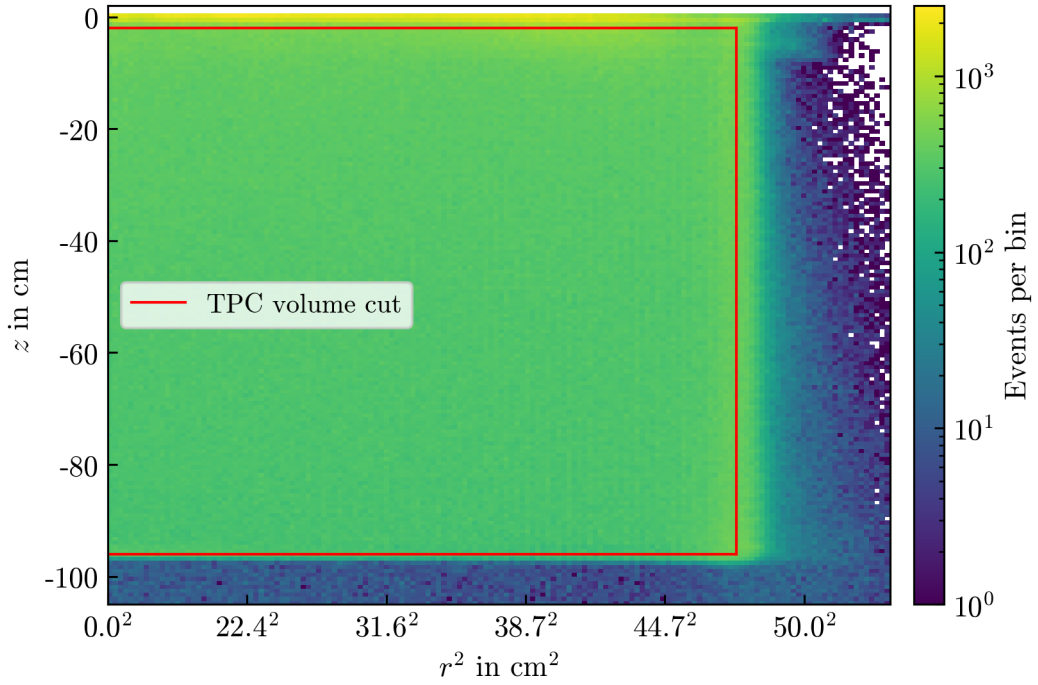

Figure 4.2: Histogram of the event distribution in the volume of the TPC. The z-position of each event is plotted against r^2 in order to represent the quadratic growth of the volume with r. The red lines mark the applied data selection of events inside the TPC. Events with a reconstructed radius of more than 47.5 cm or height outside of $-96\,\text{cm} < z < -2\,\text{cm}$ are considered as unphysical and are therefore neglected.

are considered as unphysical, leading to $r < 47.5\,\text{cm}$ and $-96\,\text{cm} < z < -2\,\text{cm}$. Events with a z-coordinate higher than $-2\,\text{cm}$ are neglected in order to exclude gas events from the dataset, where ^{83m}Kr atoms decay in the gas phase.[1] The red box in fig. 4.2 visualises this cut. The cut removes 20.5 % of the events, which corresponds 1.28×10^6 events. Nearly half of the discarded events are removed due to the upper z-limit.

S1b-S1a time difference cut

The plugin `event_info_double` is configured to assign S1a to the largest peak being classified as an S1 which is timed before the largest S2. The peak with the second largest area becomes S1b. The time difference between S1a and S1b is defined as $\text{dt} = t_{\text{abs, S1b}} - t_{\text{abs, S1a}}$. Since the assignment of S1a and S1b does not

[1] The placement at $-2\,\text{cm}$ is a conservative choice. But since the edges of the TPC are discarded with the consideration of a fiducial volume, further analyses are not affected by this.

depend on the ordering, dt can become negative. A correctly reconstructed and processed ^{83m}Kr event has a positive dt due to the specific decay structure where the 9.4 keV transition follows the 32.2 keV transition. The cut dt $\geq$ 0 ns removes events where this is not the case. Also, an upper limit of dt $\leq$ 3000 ns is set. As mentioned in section 3.5.1, the desired ^{83m}Kr decay has a half-life of 154 ns. Only 4×10^{-7} % of the 9.4 keV transitions will occur with a time difference larger than 3000 ns to the previous 32.2 keV decay. Therefore, nearly all of the remaining events can be excluded to be ^{83m}Kr.

Events with such high time differences between S1a and S1b mainly exist due to so-called afterpulses [50, 51]. Afterpulses originate from residual gas molecules that remain in the vacuum tube of a PMT. If they are ionised by passing photoelectrons, they begin to drift towards the cathode where they can generate an additional signal. This signal is registered as an afterpulse. Its time difference depends on the number of nucleons and the ionised charge units, but is often in the order of a few µs. There are also afterpulses at time differences lower than 3000 ns. Those are reduced by the further cuts. The consequences of the cut in dt on the original dataset are visualised in fig. 4.3. The events in panel b) with an S1b area of more than 100 PE are mainly caused by the afterpulses.

Cut on the number of channels contributing to S1a

Another helpful variable which is provided by `event_info_double` is the number of PMTs that contribute to a peak, called number of contributing channels (N_{PMTs}). Normally, a ^{83m}Kr S1 is registered mainly by the bottom PMT array of the TPC. Therefore, a correctly reconstructed ^{83m}Kr S1 has a number of contributing channels corresponding to a large part of the bottom array. This number is increased by several PMTs from the top array, but does not include the activation of nearly all PMTs of the detector. Instead, such high N_{PMTs} are reached by high-

Table 4.1: Effects of the individual cuts that are part of the preselection. The additional removal of a single cut is given as the number of the additionally rejected events if the cuts are performed consecutively, its effect on the complete ^{83m}Kr dataset of 6 260 217 events is given as the absolute removal.

Cut	absolute removal	additional removal
TPC cut	1 280 393 events (20.5 %)	1 280 393 events (20.5 %)
Time difference cut	809 944 events (12.9 %)	406 495 events (6.5 %)
$N_{\mathrm{PMTs,\ S1a}}$ cut	602 098 events (9.6 %)	195 089 events (3.1 %)

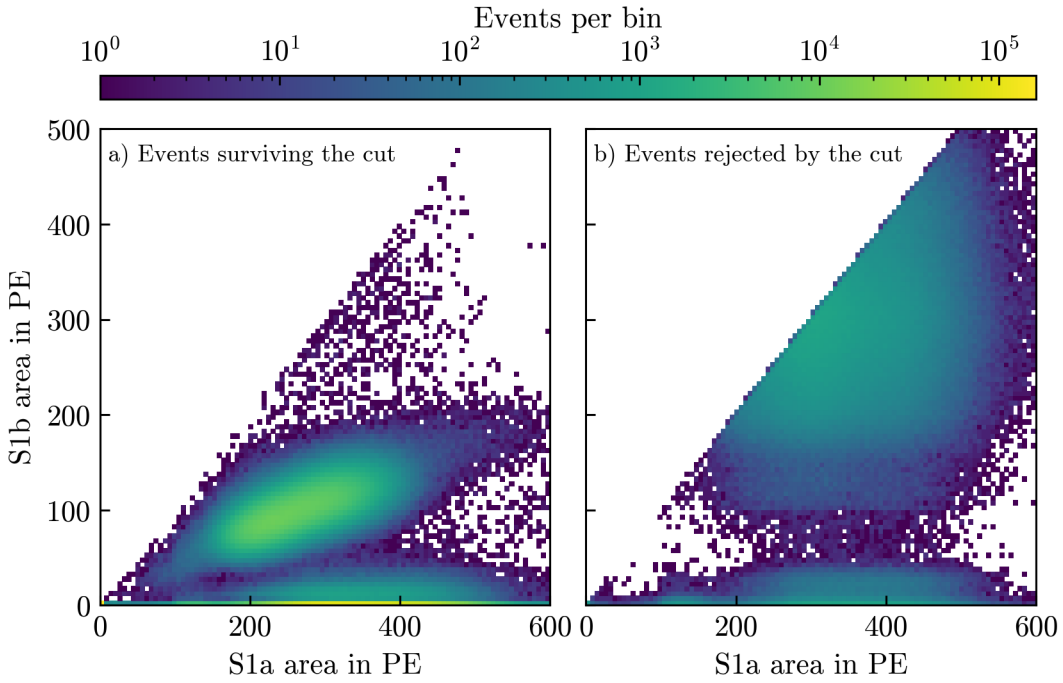

Figure 4.3: Both histograms show S1b versus S1a area of the krypton data. In panel a), the events which are left after the cut in dt are plotted, panel b) shows the rejected events. The cut in the time difference removes all events with a negative dt, where S1b happens before S1a. Also, events with a time difference larger than 3000 ns are discarded.

energy events that are supposed to be excluded from the data selection. The cut is defined as $N_{\text{PMTs, S1a}} < 150$, which is shown in the appendix (fig. A.1). A further limit is set with $N_{\text{PMTs, S1a}} > 50$ for the lower number of channels contributing to S1a. This excludes events where the S1a triggers too few PMTs to be generated by a ^{83m}Kr decay. The placement of this cut is taken from [49]. The specified cut is applied again to the complete dataset, removing a total of 8.1×10^5 events, or 12.9 %. Figure 4.4 shows the effect of this cut on logarithmic axes in order to make the different magnitudes of areas visible. Because of that, all events with S1b = 0 PE are not represented in this plot, which corresponds to most of the 41.6 keV events.

Result of the preselection

The three presented cuts are employed to get to a preselection of events that is further subdivided into the two populations of 32.2 keV and 41.6 keV. These cuts are applied one after the other, each one removing additional events. Table 4.1

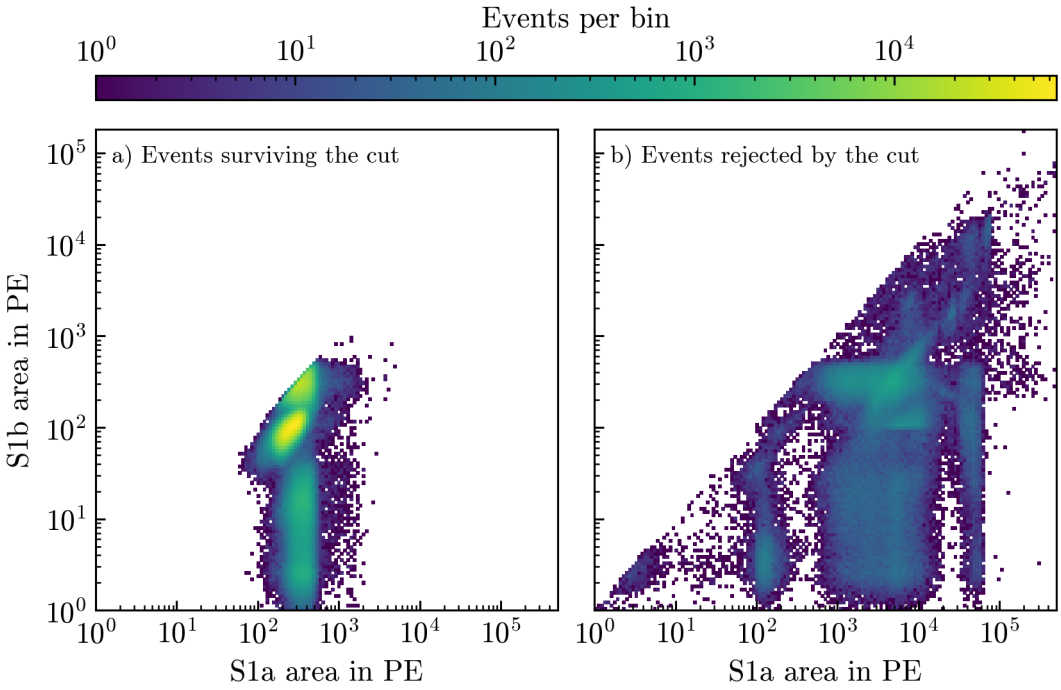

Figure 4.4: Histograms of the complete ^{83m}Kr dataset, plotted as S1b area versus S1a area. In panel a), the events surviving the cut of $50 < N_{PMTs,\ S1a} < 150$ are plotted, while panel b) shows the removed events.

gives an overview of how many events each cut rejects in addition to the previous one and also referring to the complete ^{83m}Kr dataset. The result of this preselection is shown in fig. 4.5.

4.1.2 Differentiation between 32.2 keV and 41.6 keV

In principle, the analysis of the ^{83m}Kr data is not limited to one of the two S1 populations. But since the analysis requires a single population in order to fit the requirement of a mono-energetic calibration source, it is important to separate the 32.2 keV signals from the 41.6 keV signals.

Single S1s with 41.6 keV

The main difference between the single S1 and the double S1 is the S1b peak. For the single S1, it either does not exist at all, or it is very small and corresponding to noise. Therefore, the main criterion to distinguish the double S1 from the single S1 is the number of contributing channels to S1b ($N_{PMTs,\ S1b}$). In most cases, this number is zero, since no S1b is recorded. For the events with a noisy S1b, an

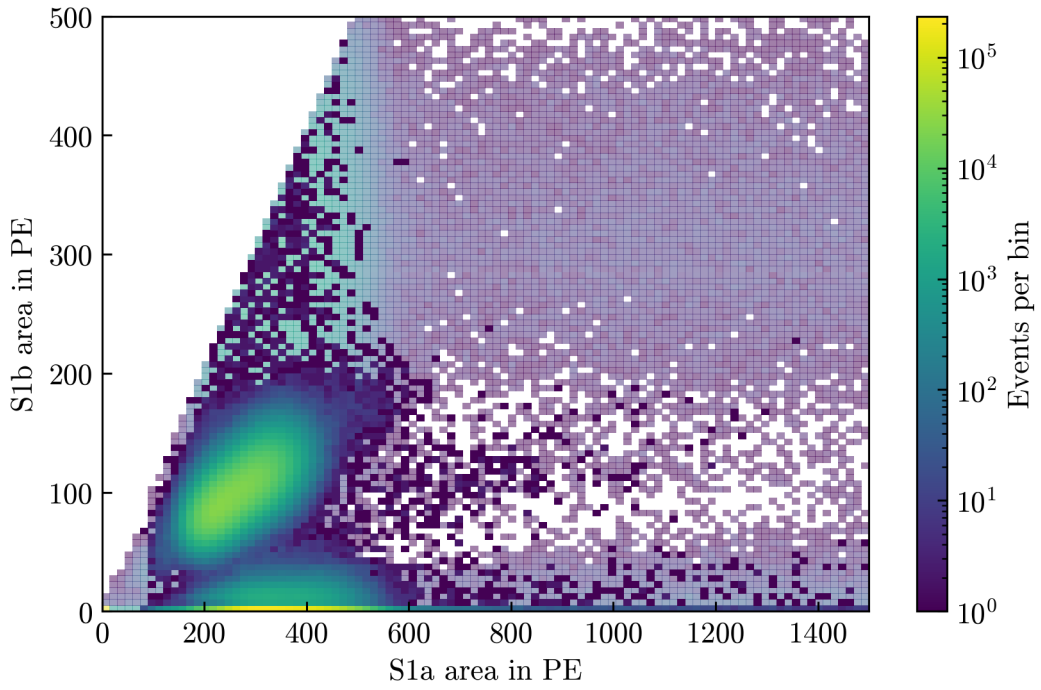

Figure 4.5: Two dimensional histograms in the S1a and S1b space for the complete ^{83m}Kr dataset (transparent) and the dataset with applied preselection (solid). Due to the signal definitions, no events with S1b > S1a are present.

upper limit of 15 PMTs is set in the cut.

Under the so far selected events with a vanishing S1b, there are several events with an extraordinarily large S1a area, visible at the bottom of fig. 4.5. Those are mainly seen by the top array. A measure for that is the variable area fraction top (AFT), defined as

$$\mathrm{AFT} = \frac{\mathrm{area}_{\text{top array}}}{\mathrm{area}_{\text{total}}}. \tag{4.1}$$

By requiring for all surviving events $0.075 < \mathrm{AFT} < 0.5$, the population above S1a = 600 PE is reduced by 69.15 %. At the same time, the events with S1a beneath this value are reduced by 2.32 %. After the application of these cuts in addition to the preselection, the selection of the 41.6 keV population is complete. The absolute numbers of remaining and removed events are given in table 4.2.

Double S1s with 32.2 keV

Also for the double S1 selection, the S1b is the most important feature for the selection. As in the cut from the preselection using $N_{\text{PMTs, S1a}}$, the number of contributing channels to S1b is restricted at the same upper limit of 150. Also,

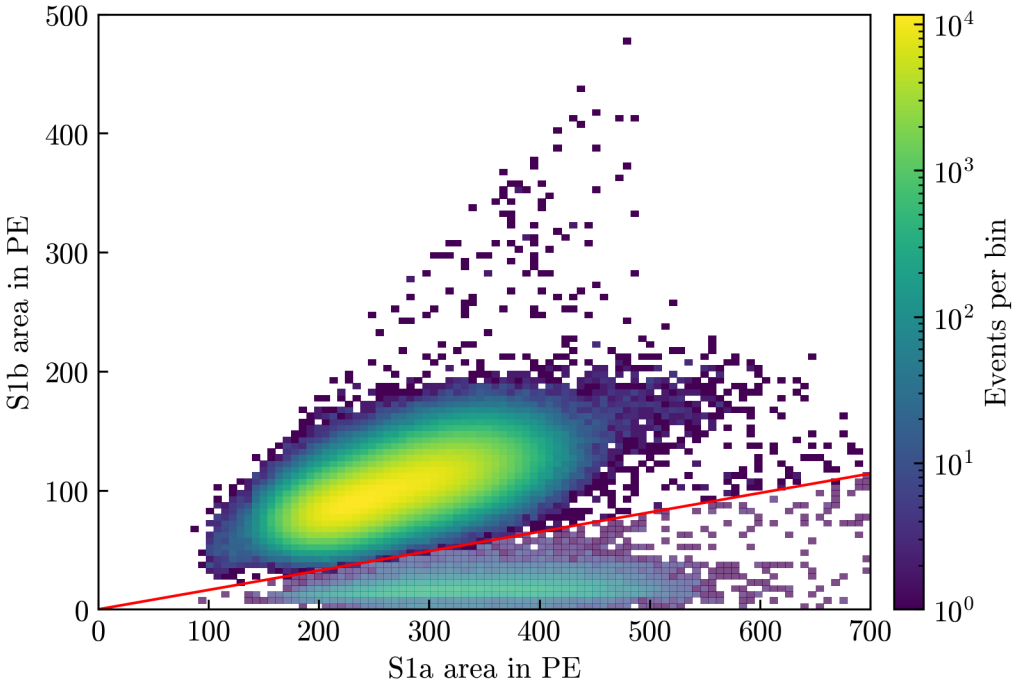

Figure 4.6: Two-dimensional histograms in S1a and S1b space for the events surviving (solid) the cut given by the red line and for the rejected events (transparent). The red line marks the direct cut in the S1a and S1b area plane. The cut is added to the data of the preselection after the cuts in $N_{\text{PMTs, S1b}}$ and $N_{\text{distinct PMTs}}$.

$N_{\text{PMTs, S1b}}$ has to be larger than 25 in order to exclude the single S1 events. The placement of the cut is plotted in the appendix (fig. A.2).

Another variable that is used is the number of distinct channels $N_{\text{distinct PMTs}}$. It contains the number of PMTs that is triggered by S1b, but not by the previous S1a and offers a tool to reject events whose main and second peak do not originate from the same interaction (pile-up events, see [52]). In order to reject such events, an upper limit of 30 distinct channels for S1b is set, which is also plotted in the appendix (fig. A.3).

Table 4.2: Absolute lengths of the datasets after the selections of single S1, strict and loose double S1 as well as the percentages of surviving events based on the complete dataset without any cuts.

	single S1	double S1 (loose)	double S1 (strict)
absolute values	2522311	1757966	165032
relative values	40.29 %	28.08 %	2.64 %

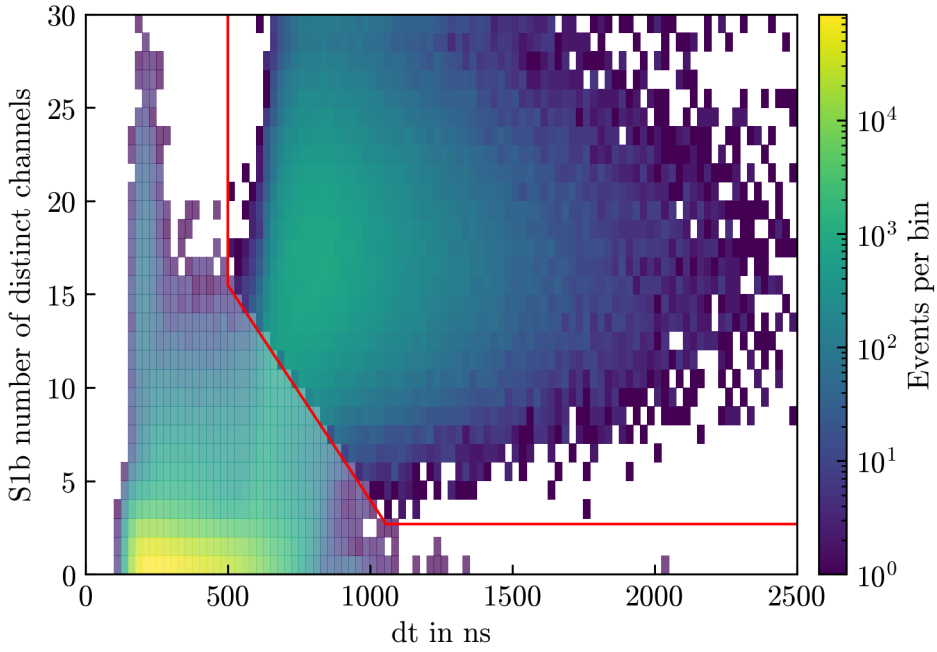

Figure 4.7: Two-dimensional histograms of the parameter space dt and $N_{\text{distinct PMTs}}$ for events surviving (solid) and being rejected (transparent) by the cut which is marked by the red line.

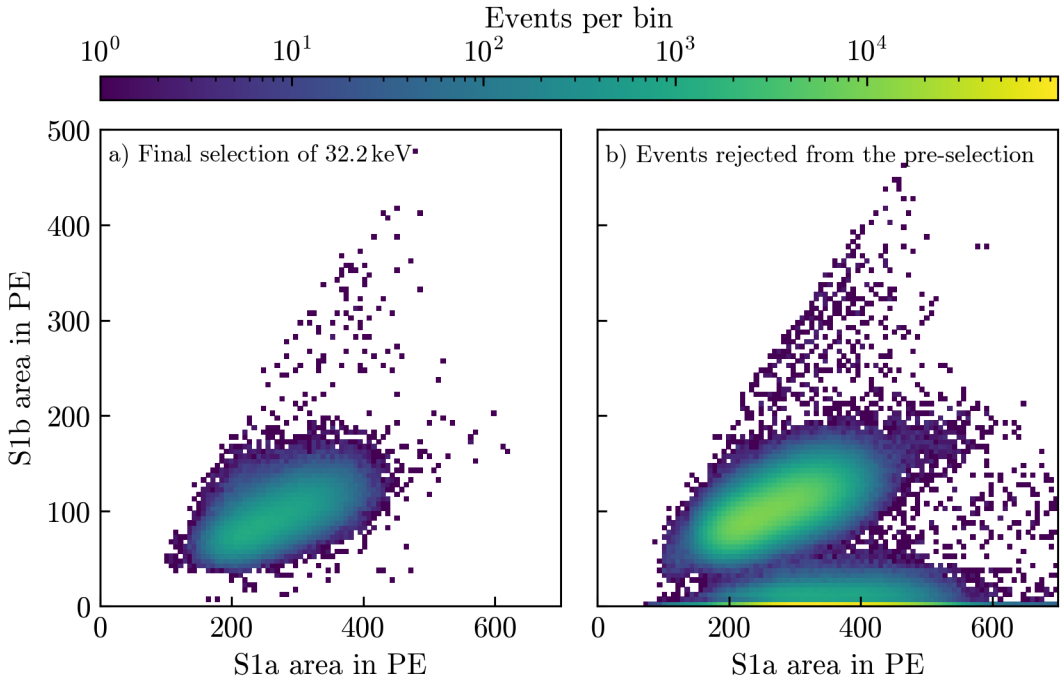

Figure 4.8: Two-dimensional histograms with S1b versus S1a area at different stages of the data selection. In panel a), the strict selection of 32.2 keV events with an accurate recording of the following 9.4 keV peak is shown. In panel b), all events are plotted that were rejected since the preselection.

As visible in fig. 4.6, there are several events with an S1b area that is too small to originate from the 9.4 keV transition. One way to reject these events without losing most of the statistic is to perform the cut directly in the plane of the areas of S1a and S1b, even though it is desirable to avoid a direct cut in S1a and S1b area in order to prevent a biased selection. Also, a direct cut in the area might not be able to exclude events that lie in the required energy range but do not originate from a correctly recorded and processed ^{83m}Kr event. Still, for the sake of a high statistic, a direct cut separating the single and double S1s can be done, which is shown in fig. 4.6. The red line marks the cut, which is defined by $\mathrm{S1b_{area}} > 0.163\,\mathrm{S1a_{area}}$. This selection will be referred to as the loose ^{83m}Kr selection.

A more restrictive way to reject these events is to look at the time differences and the number of distinct channels simultaneously. Events with a time difference of less than 500 ns to 1000 ns have a lower number of distinct channels than the events at higher dt, which is shown in fig. 4.7. The fact that events with a lower dt have different properties than events at high time differences indicates that the data processing does not work accurately for peaks that are close in time to each other. By separating these two populations visible in fig. 4.7 at the red line, a stricter double S1 selection is created. The vertical part of the cutting red line gives the requirement of $\mathrm{dt} \geq 500\,\mathrm{ns}$, the horizontal part of $N_{\text{distinct PMTs}} > 3$. Those two lines are connected by $N_{\text{distinct PMTs}} \geq 27 - 0.023\,\frac{1}{\mathrm{ns}} \cdot \mathrm{dt}$. Since the half-life of the intermediate state of ^{83m}Kr is only 154 ns, the largest amount of events happens before the required minimum time difference and therefore is lost for the analysis. The remaining events are called the strict selection and are shown in fig. 4.8. For both the strict and the loose selection, an overview is given in table 4.2.

If the selection of the double S1 events was successful, the half-life of the ^{83m}Kr decay should be visible in the selected dataset. When the events are plotted over their time difference, the events per bin should decrease exponentially. In fig. 4.9, due to a limited reconstruction and selection efficiency, the loose selection at lower dt does not exhibit an exponential decay: the decay rate is not constant. Opposed to that is the strict dataset, showing an exponential decrease of events at time differences larger than ca. 800 ns. A fit of the number of events of the strict dataset over the time difference should yield the half-life of the short-lived ^{83m}Kr state. It is performed on the strict double S1 selection in the range of 950 ns to 2050 ns.

For the fit, a χ^2-minimisation is used. This method reduces in each of the N

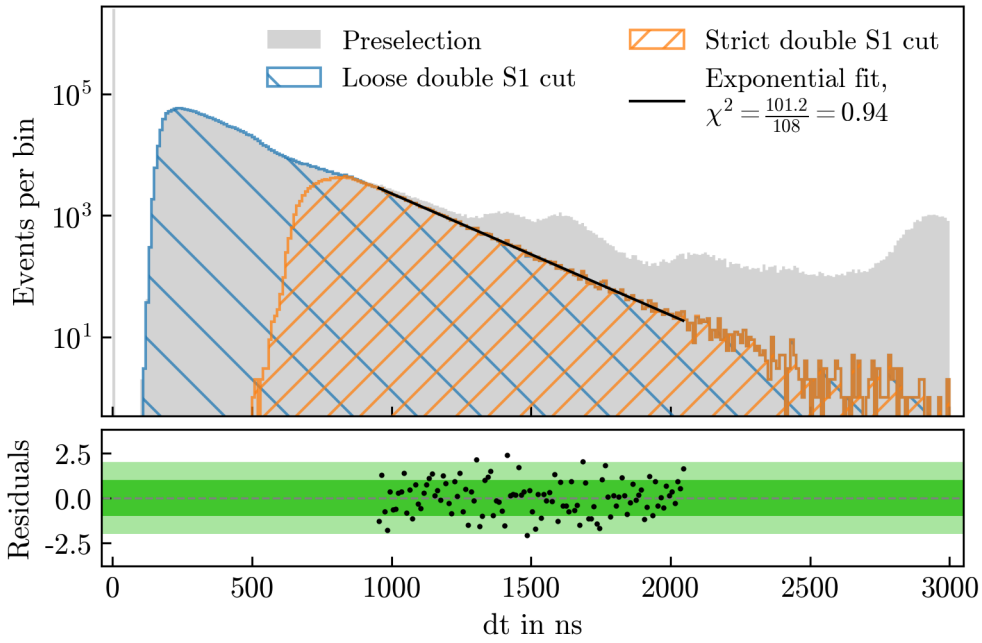

Figure 4.9: Histograms of the preselection, the strict S1 selection and the loose S1 selection over the time difference with a bin width of 10 ns. Also, the result of the half-life fit of the strict selection is shown, yielding $T_{\frac{1}{2}} = (150.4 \pm 0.7)$ ns.

bins of the fit the squared distance between the data point or the events per bin y_i and a fit function $f(\vec{p})$, which depends on the fit parameters $\vec{p}$. This distance is normalised by the uncertainty of the bin entry Δy_i. The minimised quantity is given by

$$\chi^2 = \sum_{i=1}^{N} \chi_i^2 = \sum_{i=1}^{N} \left(\frac{y_i - f_i(\vec{p})}{\Delta y_i} \right)^2 . \tag{4.2}$$

For the minimisation, the fit parameter set $\vec{p}$ is adapted in such a way that the χ^2-value becomes minimal. This is done with the package iminuit [53], which is used for minimisation throughout this work. The uncertainty estimation is given by $\Delta y_i = \sqrt{y_i}$. This estimation assumes Gaussian distributed uncertainties for the events in each bin. For bins with small numbers of entries, the presumption of symmetrical uncertainties is not valid anymore. Therefore, the fit range is restricted to 2050 ns. The goodness of a fit is described by the reduced χ^2-value. It is calculated by the χ^2-value of the last iteration of the minimisation normalised by the number of degrees of freedom. The degrees of freedom are given by the number of bins subtracted by the number of free fit parameters. Reduced χ^2-values close to 1 would then be the optimum for a successful fit.

In order to detect systematic deviations between the fit model and the data, the residuals are defined. They are plotted in the lower panel of fig. 4.9 and give the distance between the value of the fit function f_i and the data point y_i, normalised by the uncertainty of the bin entry Δy_i:

$$r_i = \frac{y_i - f_i}{\Delta y_i}. \tag{4.3}$$

In a successful fit, they are distributed symmetrical around zero and show no kind of a systematic pattern. The half-life fit employs the fit function

$$f(x, T_{\frac{1}{2}}, c) = c \cdot \exp\left(-\frac{x \cdot \ln(2)}{T_{\frac{1}{2}}}\right), \tag{4.4}$$

yielding a half-life of $(150.4 \pm 0.7)\,\mathrm{ns}$. Even though the two values are not inside their uncertainty intervals, the relative difference is less than 3 %, confirming the selection of a $^{83\mathrm{m}}$Kr dataset. Furthermore, the rejection of the events with lower time differences in favour of a strict selection is supported by the fact that these events do not show the characteristic exponential decay.

Since it is possible to select both single and double S1 events, a decision needs to be made which selection is used for the S1 correction. For this, the temporal widths of the S1s are investigated. Here, the width of a peak is defined as the time interval that includes 50 % of its area. As plotted in fig. 4.10, the width of the signals in the single S1 selection is wider than the width of the double S1s. This expected, because the single S1s consist of two separate decays. The width of the 41.6 keV events is not only larger, it is also depending on the z-position. This indicates a position dependent ability to resolve the two decays. Since the $^{83\mathrm{m}}$Kr events are employed as a spatially homogeneous calibration source, this makes the 41.6 keV events the less favourable choice. Also, the possibility to verify the double S1s as $^{83\mathrm{m}}$Kr decays by the half-life fit makes the 32.2 keV population the preferred choice. Since the half-life fit is performed on the strict data selection, it will be used for the relative S1 light yield map.

4.2 S1 correction

As explained in section 3.4.1, the volume of the TPC is subdivided into spatial bins with equal volumes. For this analysis, eight z slices and six radial bins are

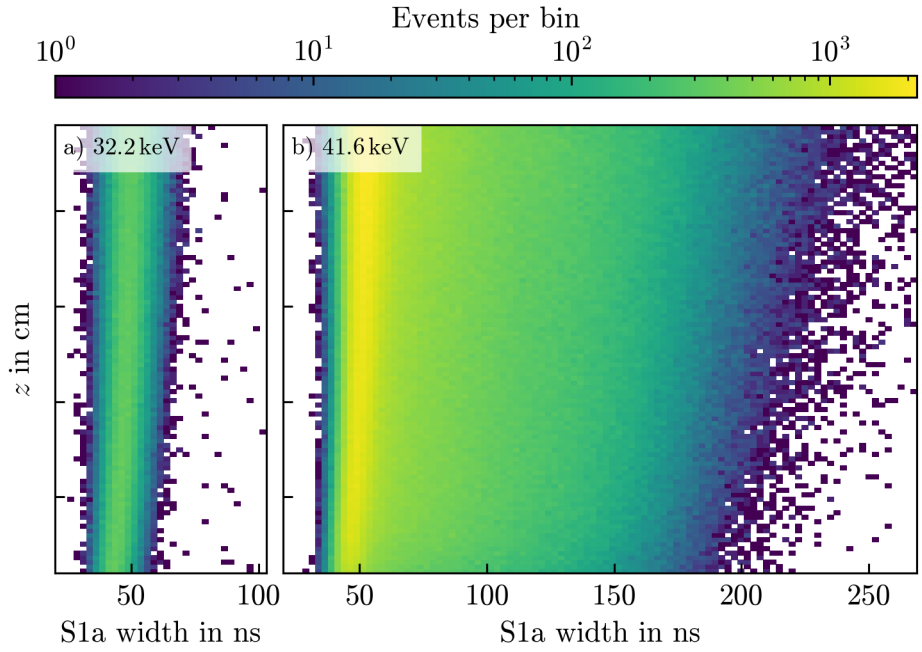

Figure 4.10: Width of the double and single S1 selection in panel a) and b), respectively, plotted in a histogram against the z-position in the TPC. The width of the double S1 selection shows less spatial dependence, making it the favourable choice.

chosen. With the azimuthal bins given by eq. (3.4) in order to maintain equal three-dimensional bin volumes, this makes 288 bins in total with 36 bins per slice. Due to the choice of the strict cut with lower available statistic, the fitting is done with a binned Poisson maximum likelihood fit, avoiding the uncertainty estimation of a χ^2 fitting method. Instead of minimising the deviation of a model to the data like the χ^2-method, the Poisson likelihood fit attempts to maximise how well a model describes the data. The probability for k counts in a bin is given by the Poisson distribution

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{4.5}$$

where $\lambda(\vec{p})$ depends on the fit model. By varying the fit parameters $\vec{p}$ and with that also $\lambda(\vec{p})$, the probability for the given number of counts is maximised. For a total number of N bins, the overall likelihood $\mathcal{L}$ is obtained by the multiplication

$$\mathcal{L} = \prod_{i=1}^{N} \frac{\lambda_i^k}{k_i!} e^{-\lambda_i}. \tag{4.6}$$

For a numerical treatment, a sum is easier to handle than a product. This is achieved by evaluating the logarithm of $\mathcal{L}$, yielding

$$\ln \mathcal{L} = \sum_{i=1}^{N} k_i \ln(\lambda_i) - \ln(k_i!) - \lambda_i. \tag{4.7}$$

Since the term $\ln(k_i!)$ is constant under the variation of λ, it is omitted. Also, since a minimiser is used for the fit, the negative likelihood is minimised instead of a maximisation of the positive likelihood. With $\lambda_i = f_i(\vec{p}) - y_i$, where f_i is the fit function and y_i are the counts per bin, the minimised quantity χ^2_{P} is given by

$$\chi^2_{\mathrm{P}} = \sum_{i=1}^{N} \chi^2_{\mathrm{P},i} = -2 \sum_{i=1}^{N} \left(y_i - f_i(\vec{p}) - y_i \cdot \ln\left(\frac{y_i}{f_i(\vec{p})}\right) \right). \tag{4.8}$$

Here, the factor 2 is introduced in order to obtain the same definition of parameter uncertainties as in the least squares method [54].

In each bin, a two-dimensional Gaussian fit is performed in the S1a and S2a space. The vertical axis is divided by a factor 10 in order to have the fit parameters for S1 and S2 on a similar order of magnitude. The fit function

$$\begin{aligned} G_{\mathrm{2d}}(x, y, A, \mu_x, \mu_y, a, b, \theta) = A \cdot \exp\Bigg(& - \frac{\Big((x-\mu_x)\cos\theta - (y-\mu_y)\sin\theta\Big)^2}{2a^2} \\ & - \frac{\Big((x-\mu_x)\sin\theta + (y-\mu_y)\cos\theta\Big)^2}{2b^2} \Bigg) \end{aligned} \tag{4.9}$$

describes an ellipse that is tilted by an angle θ, which creates the terms in the exponential function similar to a two-dimensional rotation matrix. The rotation of the data in the (S1,S2)-space is based on the anti-correlation between the light and the charge signal. The more electrons escape the interaction site, the larger the S2. In this case, the light signal is smaller, since each electron escaping the interaction site cannot generate scintillation light by recombining. The parameters a and b determine the width of the ellipse in both dimensions, while the parameters μ_x and μ_y correspond to the coordinates of the centre point of the ellipse with x = S1a and y = S2a. Lastly, the amplitude A determines the maximum of the ellipse.

One example fit of a bin from the outermost annulus of the fourth z slice (number 138) is shown in fig. 4.11. Even though the binning is different, the numbering is done after the same scheme as presented in fig. 3.5. From each fit, the parameter

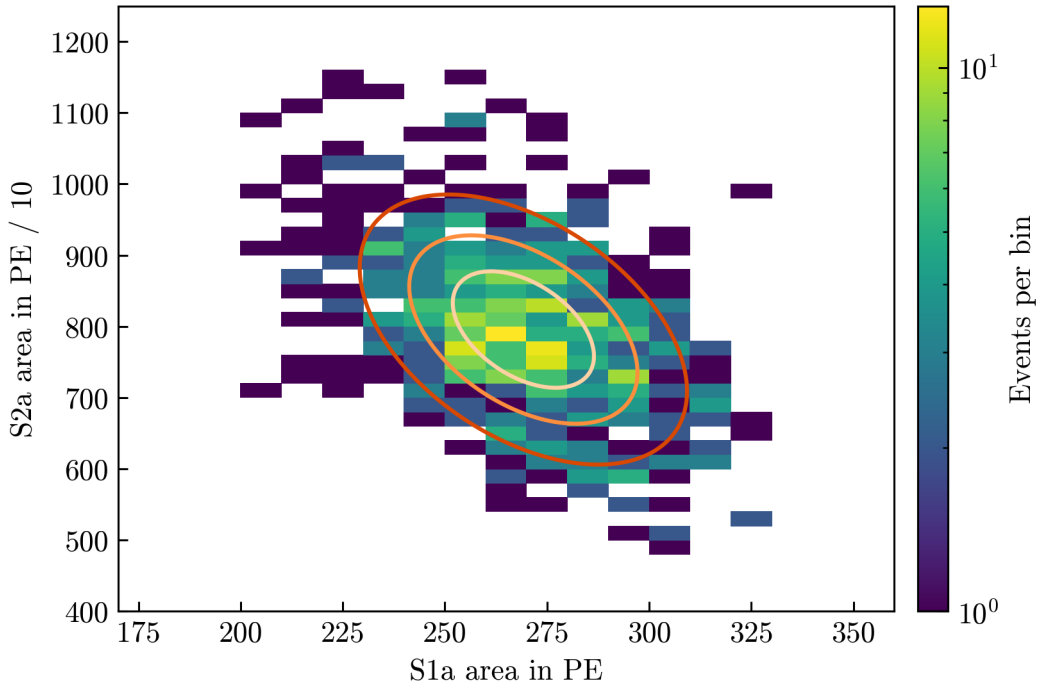

Figure 4.11: Histogram of the S1a and S2a/10 space for an example bin (nr. 138). In each bin, a two-dimensional Gaussian fit is performed. The ellipses give the amplitude of the fit function in eq. (4.9). The pictured bin is one of the outermost radial bins of the fourth z slice.

μ_x of eq. (4.9) is the key value. Therefore, μ_x is stored as the result of a fit.

After all bins were fitted, the mean of all fitted S1a areas is calculated. This value is then used to normalise the absolute fitted areas in order to receive a relative value for the light yield in each TPC bin. The result is shown in fig. 4.12, where all TPC bins are plotted. The relative LY is indicated by the colour scale. The closer a z slice is to the top of the TPC, the lower is the relative light yield. This can be explained by the fact that S1s are mainly registered by the bottom array. Photons that are generated in the upper part of the TPC are reflected by the TPC walls and the liquid gas interface more often and in general have to cover a longer distance, which causes a signal loss from the upper part of the TPC.

Only the slices closest to the bottom array show a radial or angular dependency. The angular dependency is likely due to several deactivated PMTs during SR1 [41] in the bottom array, which reduce the light yield from events that happen directly above them. For events at high radii inside the lowest slice, the same argumentation applies as before. A higher fraction of the photons needs to be reflected at the walls before reaching a PMT, which leads to a higher signal loss.

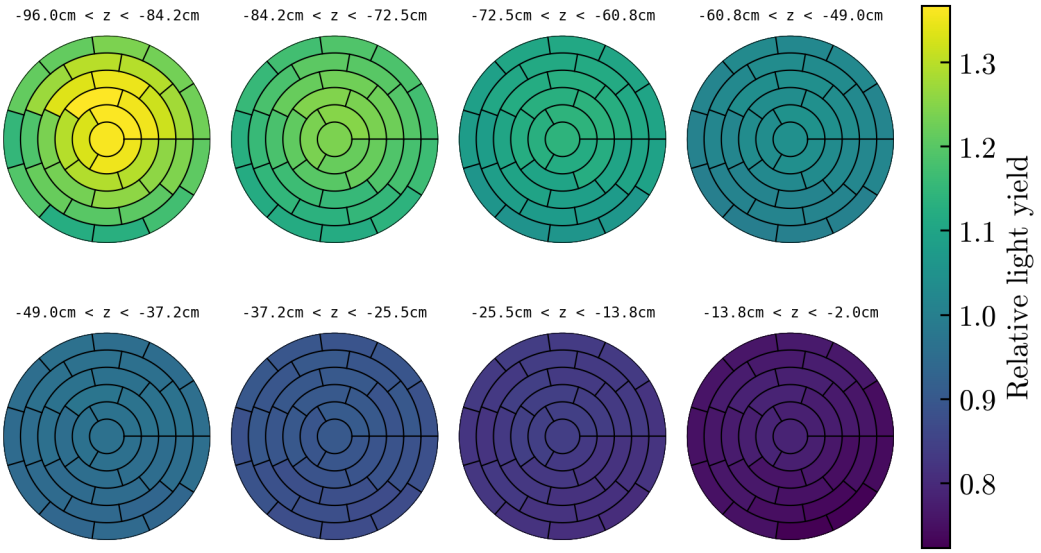

Figure 4.12: Relative light yield for the 288 bins of the TPC indicated by the colour map. Each circle represents a slice in z-direction. The number of φ-bins grows with the radial binning in order to maintain a constant volume per bin.

The relative light yield is assigned to the coordinates of each bin's centre point resulting in a discrete LY map. With the `InterpolatingMap` package [42], a continuous, three-dimensional map is constructed in between the centre points of the bins. This is based on the nearest neighbouring bins, with a default setting of $2n_{\text{dim}} = 6$ neighbours in a three-dimensional space. The average of these bins is calculated with a weighting $w_i = 1/\Delta l$ by the inverse of their distance Δl.

Based on eq. (3.6), the correction is applied to the data. As visible in fig. 4.13a), with a low number of z slices, the corrected data display clearly visible steps in z that correlate with the number of slices from the correction map. This is due to the fact that the influence of bins from the neighbouring slice is smaller, if the distance between them is larger. This can be avoided with either a higher total number of slices[2] or a higher number of neighbouring bins that are considered for the interpolation, which is shown in fig. 4.13b). Especially visible with few slices but a high number of neighbours is that the outer half of first and last slice in z benefit less from the interpolation than the slices in the middle, since they have no neighbours from the next slice influencing their relative light yield. Therefore,

[2]Using the high statistic of the loose ^{83m}Kr data selection, the binning can be chosen fine enough to stay at the default setting of 6 nearest neighbours. In the appendix, a correction map with a finer binning, created with the loose data selection, and its application to the data using 6 nearest neighbours are plotted in figs. A.4 and A.5, respectively.

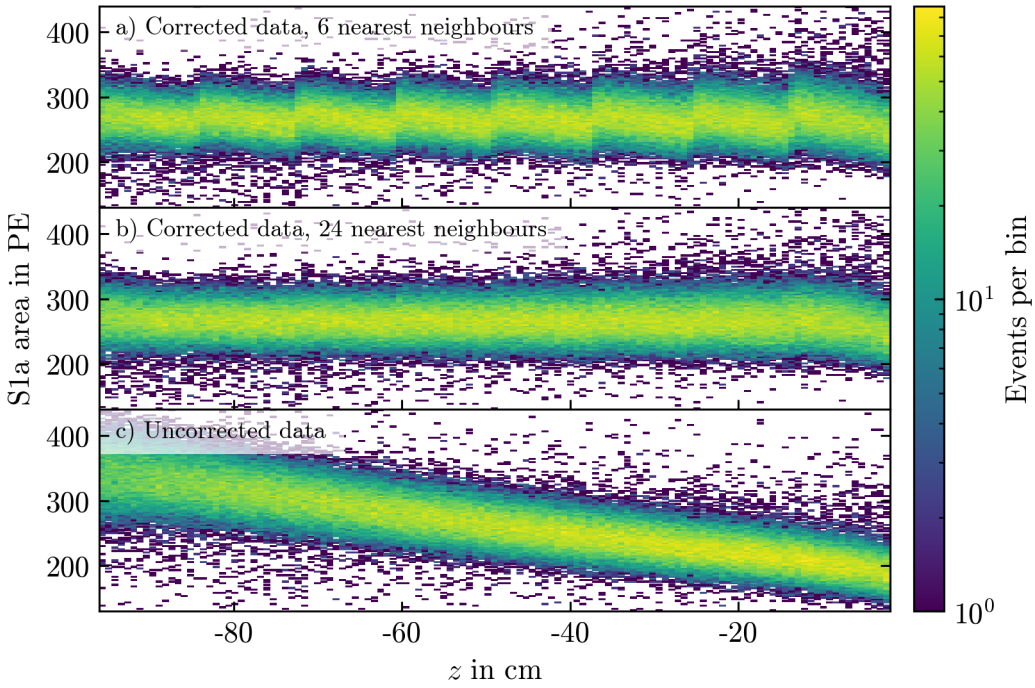

Figure 4.13: S1a areas in the full radius of the TPC as a function of the z-coordinate for three different S1 corrections. The interpolated correction map from fig. 4.12 is used with eq. (3.6) to perform the correction on the strict ^{83m}Kr dataset. Panel a) shows the correction with a considered number of six nearest neighbours, in comparison to 24 nearest neighbours in panel b). The last panel c) shows the uncorrected data.

the slope in S1a area remains visible at the ends of the TPC, even though the centre of the slice is aligned at the average. Most further analyses use a fiducial volume, cutting the top and the bottom part of the TPC, and thus are not affected. For comparison, the uncorrected data are plotted in fig. 4.13c). Summarised, the presented S1 correction achieved a homogeneous cS1 provided over the whole TPC.

4.3 S2 correction

In order to determine the electron lifetime, the same strict ^{83m}Kr dataset as before is used. One additional requirement is made with $\mathrm{S2a_{area}} > 1000\,\mathrm{PE}$. This removes a few events with a misidentified S2, which has not been considered in the cuts so far. The liftime is commonly estimated using the $\mathrm{S2_{bottom}}$ area. This variable is defined as the the part of an S2 registered by the bottom PMT array. It is used due to saturation that can occur at high energy [29]. Since ^{83m}Kr has a low-

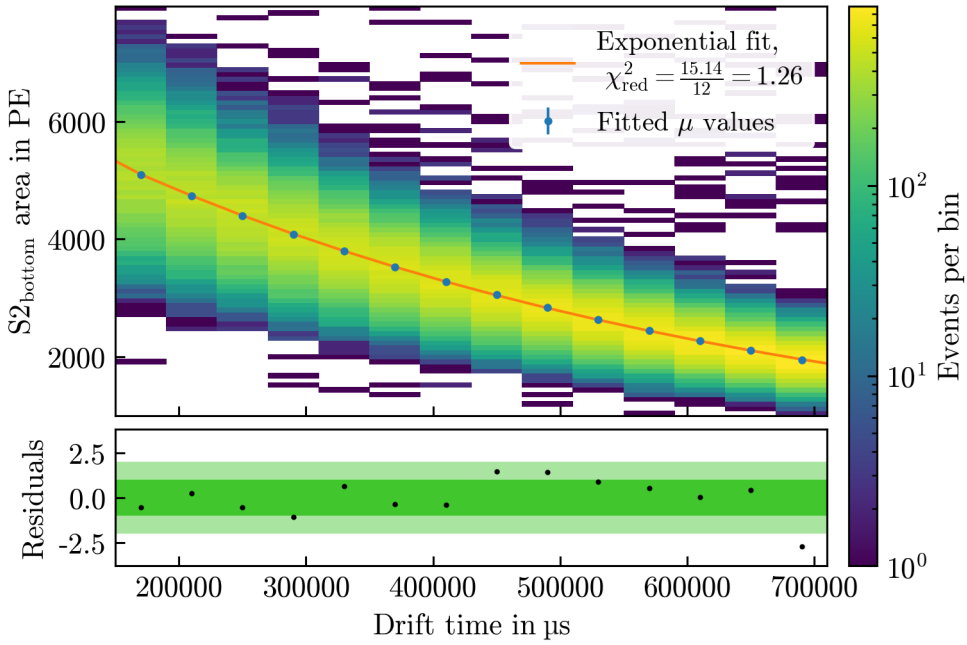

Figure 4.14: Two-dimensional histogram in the $S2_{\text{bottom}}$ and drift time space with the electron lifetime fit (orange line), yielding $(543.2 \pm 0.7)\,\mu s$. It is performed with the fit function eq. (3.7) on the blue data points, which themself are the results of further fits. Those Gaussian fits are made in order to assign a mean S2 area to every drift time bin. The histogram shows the binning in drift time that is used for the acquisition of the S2 values.

energetic decay, saturation is not an issue here. In order to maintain comparability, a reconstructed S2 bottom signal based on the AFT is used. It is given by

$$S2_{\text{bottom}} = (1 - \text{AFT}_{\text{S2}}) \cdot S2, \tag{4.10}$$

where AFT_{S2} is the S2a area fraction top.[3]

For the lifetime fit, first the data are binned in the drift time. The events that are contained in a single bin are again binned in their $S2_{\text{bottom}}$ area, where they form a Gaussian distribution. Again, due to the low statistic, the events binned in S2 are fitted by a maximum likelihood fit with the one-dimensional Gaussian function

$$G(x, A, \mu, \sigma) = \frac{A}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\pi\sigma^2}\right). \tag{4.11}$$

For each bin in the drift time, its central value is chosen as the drift time represent-

[3] This definition is only working as long as no saturation is present, since it does not fully exclude a contribution from the top PMT array.

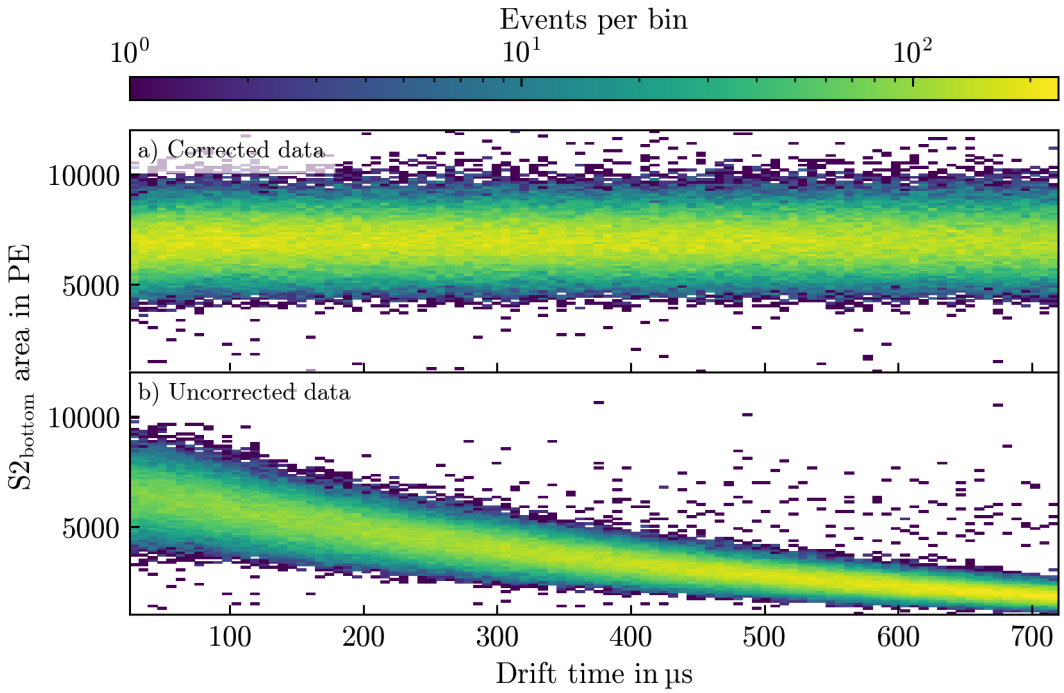

Figure 4.15: Two-dimensional histograms of drift time and $S2_{bottom}$ area. Panel a) shows the S2 area of the strict ^{83m}Kr dataset after the electron lifetime correction. The uncorrected data are shown below in panel b).

ing this bin. The parameter μ from the Gaussian fit is assigned as its representative S2 area. This combination leads to the blue data points in fig. 4.14, which are then fitted with the fit function given in eq. (3.7). The uncertainties on the parameter μ from the previous Gaussian fit Δy_i are used for the χ^2-minimisation of the exponential function, leading to a reduced $\chi^2_{\text{red}} = \frac{15.14}{12} = 1.26$. The residuals in the lower panel do not show a strong systematic effect. The fit results in an electron lifetime of $(543.2 \pm 0.7)\,\mu s$, which is in a similar range as the value that can be read off for the ^{83m}Kr model in [41].

With the known electron lifetime, the S2 can be corrected for the charge loss during the drift. Using again eq. (3.7), depending on the measured drift time, for each event a correction factor is calculated. A division by this factor reconstructs the original size $S2_0$ of the main S2. In fig. 4.15, a comparison between uncorrected data and corrected data is shown. The correction successfully aligns the events such that there is no dependency on the drift time anymore.

5 Signal corrections with ^{37}Ar

As discussed in section 3.5.2, another calibration source that can be used for calibration tasks is ^{37}Ar. Due to the small energy deposition, the light signals that are created by ^{37}Ar are close to the detectors S1 detection threshold. An important detail for this threshold is the tight coincidence requirement in the event reconstruction process, which was introduced in section 3.3.2. For the classification as a light signal, a peak needs at least 3 contributing PMTs. For the charge signal, a minimum number of 4 PMTs is required. This is part of the noise reduction, since it prevents dark counts or the possible accidental coincidence of dark counts in two PMTs from being recognised as a peak.

In this chapter, a suitable ^{37}Ar dataset is composed. As for the ^{83m}Kr calibration data, a correction for the relative LY is constructed and applied while taking the threshold effect into account. The chapter ends with a determination of the electron lifetime in combination with a correction of the S2 data.

5.1 Data selection

Next to the ^{83m}Kr runs from SR1, also 25 one hour ^{37}Ar runs were reprocessed for straxen. They were recorded at the end of the data taking period of XENON1T in SR2, corresponding to the end of October in 2018. Beside the argon calibration source, also the ^{83m}Kr source was open. Because of that, the data also contain the signature of the ^{83m}Kr events, which was described in section 4.1. Other than ^{83m}Kr, the ^{37}Ar signals consist of only one peak in light and charge signal. Therefore, they are simply called S1 and S2 and the runs are processed with the plugin `event_info`. Similar as in section 4.1, the first step is to perform a data selection. In order to avoid a direct cut in the area variables, the data selection is performed in the number of PMTs contributing to the S1 and S2 signals. This parameter space with the corresponding cut is visualised in fig. 5.1. It spans from 0 PMTs to 213 PMTs, since 35 PMTs of XENON1T's 248 PMTs were turned off or excluded from the analysis for SR2 [55]. Since an event is allowed to consist of solely S1 peaks or S2 peaks, the baselines with either $N_{\text{PMTs, S1}} = 0$ or

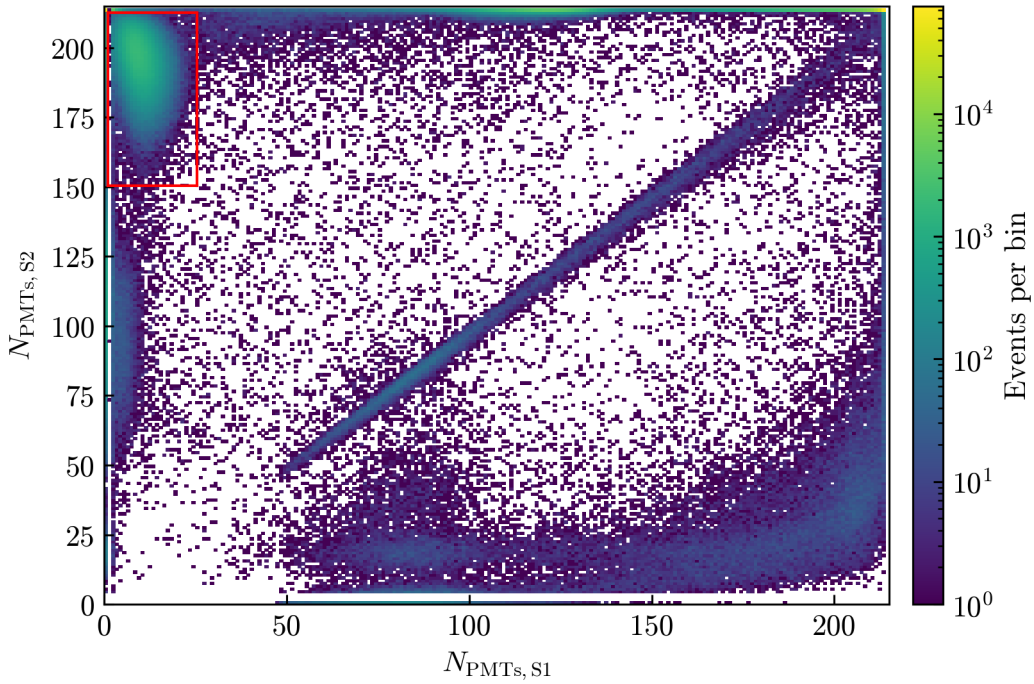

Figure 5.1: Two-dimensional histogram of the number of contributing channels of S1 and S2, offering a possibility to select the ^{37}Ar population. The red box gives the cut that is applied.

$N_{\mathrm{PMTs,\,S2}} = 0$ are populated. $N_{\mathrm{PMTs,\,S1}}$ being 1 or 2 and $N_{\mathrm{PMTs,\,S2}}$ being 1, 2, or 3 is not possible, since the classification as an S1 or S2 requires 3 or 4 contributing channels, respectively. These requirements lead to the gaps in the population on the lower and the left side of the plot.

The ^{37}Ar events are expected to have a small S1 and therefore a small $N_{\mathrm{PMTs,\,S1}}$ close to the minimum of 3 PMTs. Opposite to that, a correctly recorded S2 is large enough to be seen by many of the PTMs. Therefore, the ^{37}Ar events are expected to be the population in the top left corner of fig. 5.1. The cut is given by the requirements $0 < N_{\mathrm{PMTs,\,S1}} \leq 25$ and $150 < N_{\mathrm{PMTs,\,S2}} \leq 212$, which is indicated by the red box.

The other correctly reconstructed populations contained in these runs, for example the $^{83\mathrm{m}}$Kr population, are located at the top of the plot. Since ^{37}Ar is at the detection threshold of the detector, the other signals are larger and therefore N_{PMTs} is higher for S1 and also for S2. Events with equal numbers of contributing channels in S1 and S2 as well as events where $N_{\mathrm{PMTs,\,S2}} < N_{\mathrm{PMTs,\,S1}}$ are misreconstructed events, where for example an alternative S1 is identified as the matching

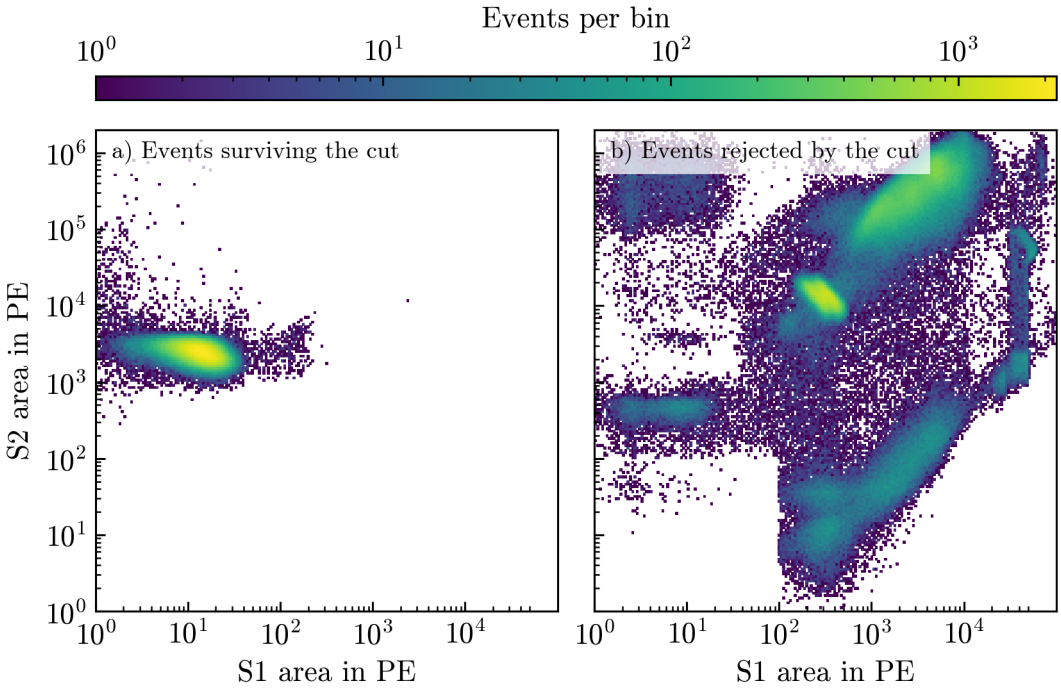

Figure 5.2: Two-dimensional histograms of the S1 and S2 areas with logarithmic scales on both axes. The left panel shows the argon events surviving the cut on the number of channels in fig. 5.1, while the right panel shows the events that were rejected by the cut.

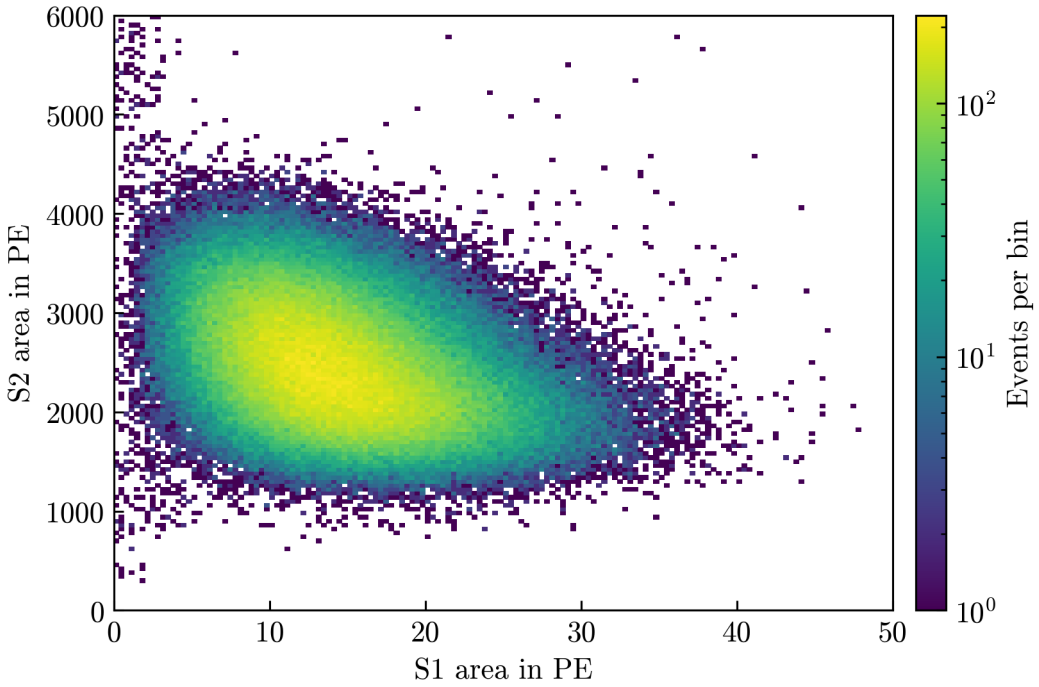

Figure 5.3: Two-dimensional histogram of S1 area and S2 area for the final selection of ^{37}Ar events.

S2.

The result of the cut in N_{PMTs} is shown in fig. 5.2, the events in the left panel remain in the selection, while the events in the right panel are excluded. In order to further clean the selection, several additional cuts are applied. An overview of the amount of removed events in each cut is given in table 5.1. As for the $^{83\mathrm{m}}$Kr selection, all S1 events with an S1 AFT higher than 0.5 are rejected. Also, the same TPC dimension cut as in the $^{83\mathrm{m}}$Kr selection is applied in order to exclude events with an unphysical reconstructed position. The final event selection is shown in fig. 5.3.

5.2 S1 correction

As for the krypton calibration source, the principle of the S1 correction follows section 3.4.1. The events are sorted into three-dimensional bins and for each bin a mean S1 value is calculated with a fit based on eq. (4.8). Since the argon selection consists of fewer runs than the $^{83\mathrm{m}}$Kr population, a coarser binning than in fig. 4.12 is applied. The number of annuli is reduced to 3 (compare fig. 3.5).

As visible in fig. 5.3, the ^{37}Ar population in the S1 and S2 space shows a deformation in the S1 area, which is related to its closeness to the S1 detection threshold. Therefore, a different fit function is required in order to determine the mean S1 in each bin. In the following paragraphs, two approaches are described in order to deal with the unusual shape of the ^{37}Ar data.

Threshold function

The first approach aims to describe the low-energy side of the argon population with an additional threshold function in S1 [56]. This function is supposed to

Table 5.1: The table summarises the effects of the individual cuts on the runs including the ^{37}Ar calibration. The additional removal of events by a single cut is given as the number of the additionally rejected events if the cuts are performed consecutively. Its effect on the complete ^{37}Ar dataset of 891 296 events is given as the absolute removal.

Cut	absolute removal	additional removal
N_{PMTs} cut	579 200 events (65.0 %)	579 200 events (65.0 %)
TPC cut	322 930 events (36.2 %)	24 428 events (2.7 %)
AFT cut	49 144 events (5.5 %)	21 461 events (2.4 %)

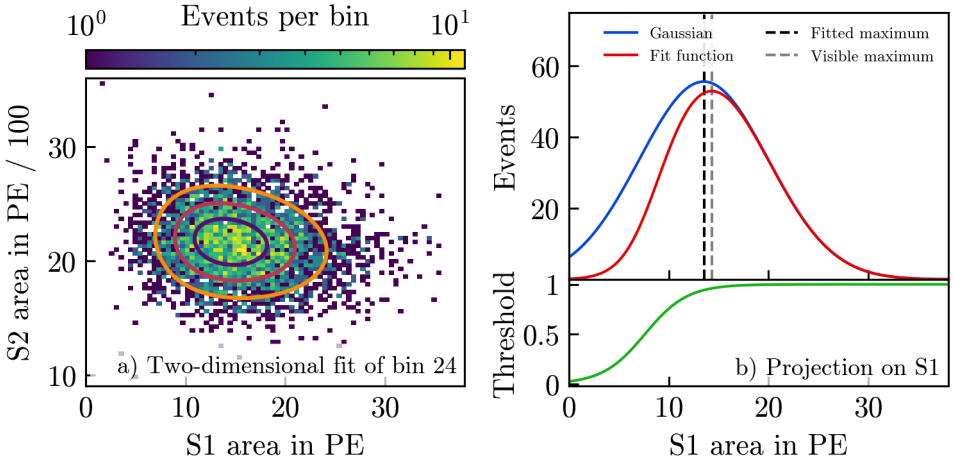

Figure 5.4: Exemplary fit of ^{37}Ar data. The left panel shows a histogram of S1 and S2 areas in bin 24 from the lower half of the TPC. The S2 area is divided by 100 in order to have similar magnitudes for S1 and S2. The concentric lines visualise the two-dimensional fit function. In the right panel, a projection of the same fit function onto the x-axis (red) is plotted. It is composed by the underlying Gaussian peak (blue) and the threshold factor (bottom panel). The two vertical dashed lines represent the positions of the fitted mean value of the Gaussian μ_x (black) and of the effectively observed maximum (grey).

modify a Gaussian function in a way that its rising flank is lowered to the level of the argon data. The threshold function is determined by two parameters, which correspond to the position x_{thr} and the width w_{thr} of the threshold. If this threshold originates merely from the tight coincidence requirement, the two parameters are assumed to be constant, because a signal originating from a position where usually far more than the three required PMTs are activated should not be affected. The fit function $f(x, y, A, \mu_x, \mu_y, a, b, \theta, x_{\mathrm{thr}}, w_{\mathrm{thr}})$ has 8 free fit parameters and is given by

$$f = \frac{A}{1 + \exp\left(\frac{x - x_{\mathrm{thr}}}{w_{\mathrm{thr}}}\right)} \cdot \exp\left(- \frac{\Big((x-\mu_x)\cos\theta - (y-\mu_y)\sin\theta\Big)^2}{2a^2} - \frac{\Big((x-\mu_x)\sin\theta + (y-\mu_y)\cos\theta\Big)^2}{2b^2} \right). \tag{5.1}$$

A visualisation of the threshold function and the resulting fit function is shown in the right panel of fig. 5.4. The left panel shows an example fit of bin 24 with the parameters $\mu_x = (13.5 \pm 0.8)\,\mathrm{PE}$ and $x_{\mathrm{thr}} = (7.6 \pm 1.1)\,\mathrm{PE}$.

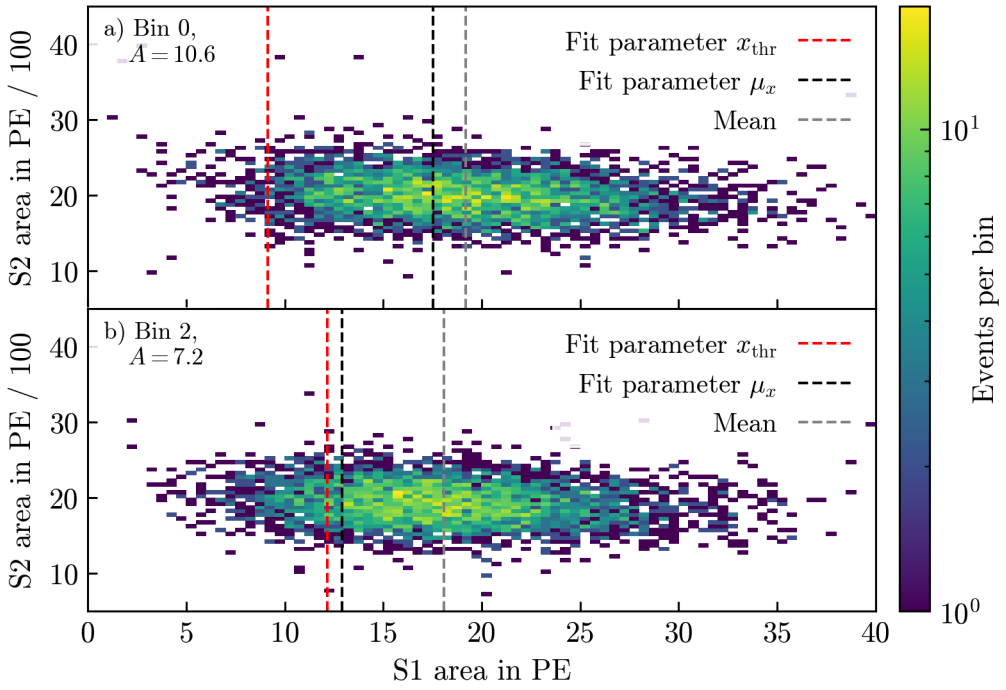

Figure 5.5: Histograms in the S1 and S2 area of two different bins from the lowest z slice. The best fit positions of μ_x and x_{thr} as part of the threshold fit function are marked by red and black dashed lines, respectively. For comparison, the mean value of the S1 area is given by the grey dashed line. The amplitude is indicated by the parameter A.

One drawback of this approach comes with the high number of fit parameters, which causes the successful convergence of the fit to be extremely dependent on the initial parameter values. Also, a slight variation in the data can lead to a noticeable change in the placement of the best fit, since the parameters are correlated. A smaller value of μ_x can be compensated by a higher placement of the threshold position x_{thr}, accompanied with a slight rise of the amplitude. This behaviour is illustrated with the example of two bins of the bottom slice in fig. 5.5. While μ_x is lower in panel b), x_{thr} is increased in comparison to panel a). Correspondingly, the fit of the amplitude in panel a) is higher than in panel b), although the overall number of entries is around 7 % higher in the lower example.

Therefore, a fit unconstrained in all eight parameters does not generate a reliable estimation of the S1 mean value. This is shown in the top panel of fig. 5.6, where the relative light yield is plotted per TPC bin. The bin numbering corresponds to the numbers that are given in fig. 3.5, the transition to the next z slice is marked by the dashed grey lines. Especially the bins of the top slice with their low light

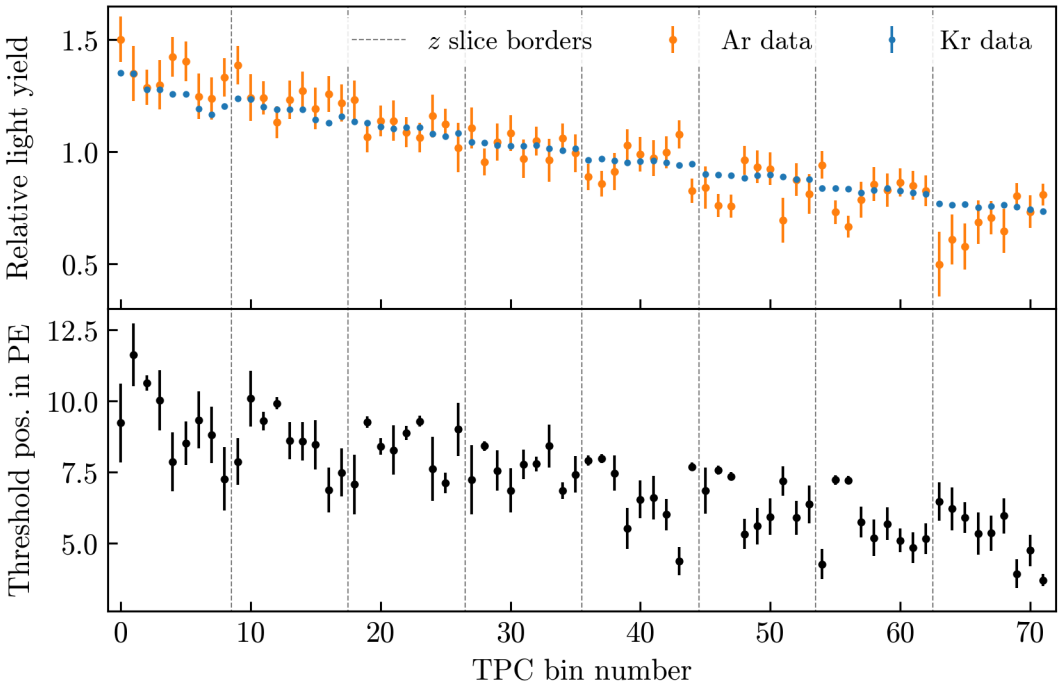

Figure 5.6: On the x-axis, the numbered TPC bins are plotted, where the numbering is done as in fig. 3.5. The grey lines mark the borders of the z slices. The upper plot shows the relative light yield from the argon calibration data (orange), derived by the fit function eq. (5.1). In comparison, a krypton fit with the same binning, performed as in section 3.4.1, is plotted (blue). Bottom: The position of the thresholds are given for every TPC bin in the lower plot. It decreases together with μ_x.

yield deviate from the ^{83m}Kr result. Also, they show relative uncertainties up to 20 %. The bottom panel pictures the development of the threshold position for each TPC bin, showing a clear decrease for bins at higher positions. This disagrees with the assumption that a threshold designed to compensate the tight coincidence requirement should have a constant value. Furthermore, the ^{83m}Kr analysis already showed that the LY is highest at the bottom of the TPC (corresponding to low bin numbers). If the threshold position would somehow be connected to the LY, the opposite trend should be observed. Thus, either the actual position of the threshold is not fitted properly due to the high number of free parameters, or the asymmetry of the ^{37}Ar S1 data is not solely caused by the tight coincidence requirement.

With a fixed threshold, the fit is more stable and convergence is less affected by the starting parameters. When setting the threshold fit parameters to a fixed

value, the resulting μ_x value is anti-correlated with the specific chosen value of x_{thr}.[1] Since the ^{37}Ar data do not offer a possibility for a direct investigation of the threshold position, the simple approach of a fixed threshold is not justified.

The threshold could be further investigated with a calibration source providing a continuous energy spectrum, for example the β-decay of ^{212}Pb as part of the ^{220}Rn calibration decay chain. Ideally, the shape of the reconstructed spectrum in each bin might give a direct handle on this threshold. As a result, the parameter x_{thr} and w_{thr} from eq. (5.1) could then be fixed. This analysis was not done since the corresponding XENON1T data were not converted to straxen, but could be worth investigating with data from XENONnT.

Another factor that discourages the usage of the fit function in eq. (5.1) is that this is a phenomenological model without a physical explanation why the threshold function should be formed exactly like this. Therefore, another way to describe the ^{37}Ar data is considered.

Smeared Poisson function

In this subsection, a physically motivated model is applied to the ^{37}Ar data. The model is inspired by the PMT gain calibration and is a one-dimensional model, focused solely on the distribution of the S1 area. It assumes that the asymmetrical shape of the ^{37}Ar data is not directly connected to the tight coincidence requirement but originates from the light source itself. A similar approach was used in [57].

The gain calibration of a PMT measures the amplification factor of the photoelectrons that are extracted at the photocathode. For an ideal PMT, the amplification of a single photoelectron ($k = 1$) would be perfectly constant and therefore have the shape of a δ-peak. An ideal PMT would have $\epsilon_{\text{QE}} = \epsilon_{\text{CE}} = 1$ and a vanishing probability to generate more than one photoelectron.

For an event generating more than one primary photon ($k = 2, 3, ...$), the response of the ideal PMT would be extended by further δ-peaks, describing the multi-photoelectron peaks. The actual number of detected photons generated by the source of such a light signal has a certain expectation value λ. For small light signals, the number of generated photons and therefore the number of photoelectrons in the ideal PMT is assumed to follow a Poisson distribution.

For a real PMT and its readout chain, the gain factor for a specific number of

[1] See fig. A.6 in the appendix

photoelectrons k is not δ-shaped, but is smeared to a Gaussian peak with a width of $\sqrt{k} \cdot \sigma_0$. The overall response function is then given by the sum over several Gaussian peaks with the corresponding width for each number of photoelectrons. The amplitudes of these peaks are weighted by the probabilities of the underlying Poisson distribution.

For the modelling of the ^{37}Ar data, the whole PMT array is considered to behave like a single PMT, since all PMTs are calibrated to yield the same gain. The smearing effect of the individual PMTs is summarised into an effective smearing $\sqrt{k} \cdot \sigma$ of the whole array. Therefore, according to this model, the parameter σ is determined by the detector. This leads to a fit function depending on the expectation value of the Poisson distribution λ, the effective smearing σ and an amplitude parameter A, which is determined by the statistic in each bin. It has only 3 free parameters, compared to 8 parameters from the previous approach with a two-dimensional fit function, and is given by

$$f(x, A, \lambda, \sigma) = A \sum_{k=1} \frac{P_\lambda(k)}{\sqrt{2\pi} \cdot \sqrt{k}\sigma} e^{-\frac{(x-k)^2}{2(\sqrt{k}\sigma)^2}}, \tag{5.2}$$

with the Poisson distribution $P_\lambda(k)$. The parameter λ is used as the mean value for each bin and therefore also for the calculation of the relative LY. The fits are performed with the binned maximum likelihood method. The individual Gaussian peaks and the fit function as their sum are visualised in fig. 5.7b). For the parameters λ and σ, the results of the fit from bin 24 are used, A is set to unity. The fit is given as an example in panel a). Since the binning is chosen as in the previous threshold model and the example bin shown here is the same as in fig. 5.4, the mean value of $\mu_x = (12.5 \pm 1.1)\,\mathrm{PE}$ and $\lambda = (15.62 \pm 0.09)\,\mathrm{PE}$ can be directly compared. The largest difference is that the smeared Poisson model does not take the additional threshold caused by the tight coincidence requirement into account. This may lead to a shift of the mean value λ, placing it at higher numbers of PE.

In order to define a residual r_i for each data point, the square root of the summands in eq. (4.8) is used [54], as in a regular χ^2 fit. This leads to $r_i = \sqrt{\chi^2_{\mathrm{P},i}}$, where the sign of $\chi^2_{\mathrm{P},i}$ needs to be considered and added accordingly. For bins with no entries ($y_i = 0$), the logarithmic term has to be discarded before calculating r_i. The residuals are given in the lower panel of fig. 5.7a).

The same approach can be applied to the krypton data, since for large λ the Poisson distribution approaches a normal distribution. This leads to three different

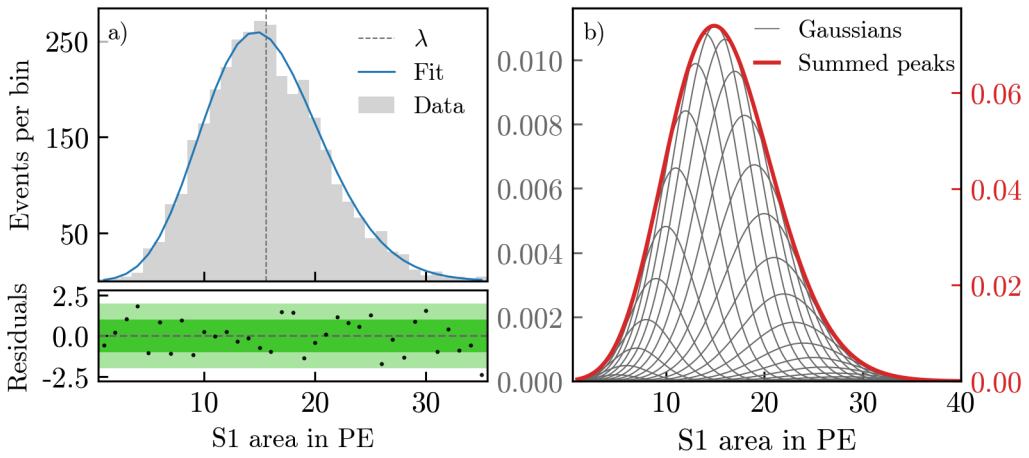

Figure 5.7: a) Histogram of the S1 area of the events in bin 24. The blue line is the fit with the smeared Poisson function eq. (5.2). The vertical dashed line marks the location of the mean value λ, being in this example at (15.62 ± 0.09) PE. The lower plot gives the residuals. b) Schematic plot of the smeared Poisson function versus the S1 area. It is constructed of overlapping Gaussian peaks (grey). Their amplitudes are weighted by the value of a Poisson distribution with the expected value λ at the Gauss peaks mean values. The sum of those peaks (red) is the fit function.

versions that are compared: the ^{37}Ar data fitted with the smeared Poisson function and the ^{83m}Kr data fitted with both, the same function as the argon data and a simple normal distribution. All three resulting relative LYs are plotted in the upper panel of fig. 5.8. As expected from the lower number of free parameters, the argon fit is much steadier and has smaller uncertainties than in fig. 5.6. While there is no or only a very small systematic difference between the two krypton fits, the relative LY from argon data tends to be smaller than the ^{83m}Kr results at the bottom of the TPC and larger at the top of the TPC. Since the data are normalised by their mean value and therefore centred in the middle of the TPC, this means that the overall range in which the argon data occur is smaller than for the krypton fits.

In the lower panel of fig. 5.8, the fit parameter σ is plotted for both the argon and the krypton fit. Without showing a clear up- or downwards trend throughout the TPC, σ can be roughly considered constant for one calibration source. Between the two sources, there is an offset of about 0.2 PE in the respective values of σ. This source dependency contradicts the earlier assumption that the parameter σ is exclusively determined by detector properties. Instead, there is a dependency

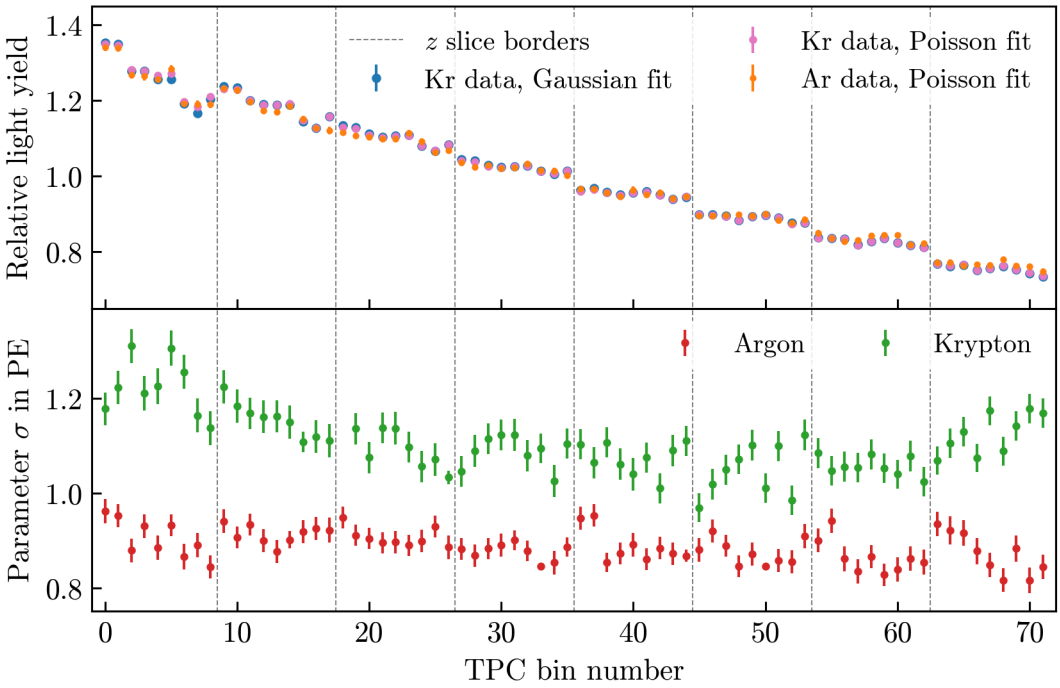

Figure 5.8: Relative light yield and fit parameter σ for the single TPC bins. The upper plot shows the relative LY of argon and krypton with the smeared poisson as a fit model. Also, the Gaussian fit of the krypton data is plotted. The lower plot shows the fit parameter σ for both the argon (red) and the krypton (green) fit with the smeared Poisson function. While it is approximately constant for each source, it differs between argon and krypton, opposed to the assumption that σ is a detector parameter.

of the source or the amount of deposited energy.

One approach of resolving this dependency is with an additional fit parameter σ_0. The idea is to split the former parameter σ into a source related broadening and into a constant part σ_0, determined by the detector properties. Adding them up in a squared sum, this leads to

$$\sigma_{\mathrm{Kr,\,ges}} = \sqrt{k\sigma_{\mathrm{Kr}}^2 + \sigma_0^2} \qquad \text{and} \qquad \sigma_{\mathrm{Ar,\,ges}} = \sqrt{k\sigma_{\mathrm{Ar}}^2 + \sigma_0^2} \tag{5.3}$$

with a shared fit parameter σ_0. The fit is performed bin-wise on both datasets simultaneously, connected via the shared parameter. But since in most of the bins σ_0 is fitted to zero without an impact on the source dependent broadening, the idea is discarded.

In this section, two different approaches for the handling of the asymmetric

light signals of ^{37}Ar were investigated. While the first approach is based on a direct model of the detection threshold, the second approach instead concentrates on other detector properties combined with an asymmetrical behaviour of the calibration source. The smeared Poisson method gives a more reliable result than the direct threshold model, even if the source dependency of the parameter σ is not explained by this model. However, the idea of a direct modelling of the tight coincidence requirement cannot be rejected completely, since an unphysical, not matching threshold function could also lead to a bad performance, without the concept of a threshold function being completely wrong.

5.3 Simulation of threshold effects

In order to investigate the possibility of the tight coincidence requirement leading to a strong influence on the observed signal, a simulation of a light signal in the XENON1T TPC is conducted. The simulation is based on a per-PMT map of the light collection efficiency which was obtained as part of Monte Carlo simulations for XENON1T, as explained in [58]. This map contains the probability for each PMT and for each point in the TPC that a generated photon reaches the PMT. The 248 PMTs and the possibility that a photon does not reach any of the PMTs at all (missed photons) are represented by a channel number. The per-PMT map is used to derive a position dependent probability for each of the channels to be hit. In a next step, the quantum efficiencies for each PMT are considered. For the PMTs that were turned on during the data taking, the QE is at about 0.3 [50]. Several PMTs were not turned on because of either a complete failure or a too high afterpulse rate [55]. These 36 PMTs are considered to have a QE of 0. Employing a binomial distribution, each channel is evaluated separately with the according QE. The entries that are sorted out in this step are added subsequently to the channel with the missed photons. For each simulated event, a list of hit channels is generated, with the length of the initially simulated photons at the interaction point.

In each event, the amount of different channel numbers is calculated, under exclusion of the channel containing the missed photons. If the number of different channel numbers is 3 or more, the event is marked as passing the tight coincidence requirement. By repeating the simulation step, a probability to overcome the requirement is derived for a certain point in the TPC and a specific number of initially generated photons.

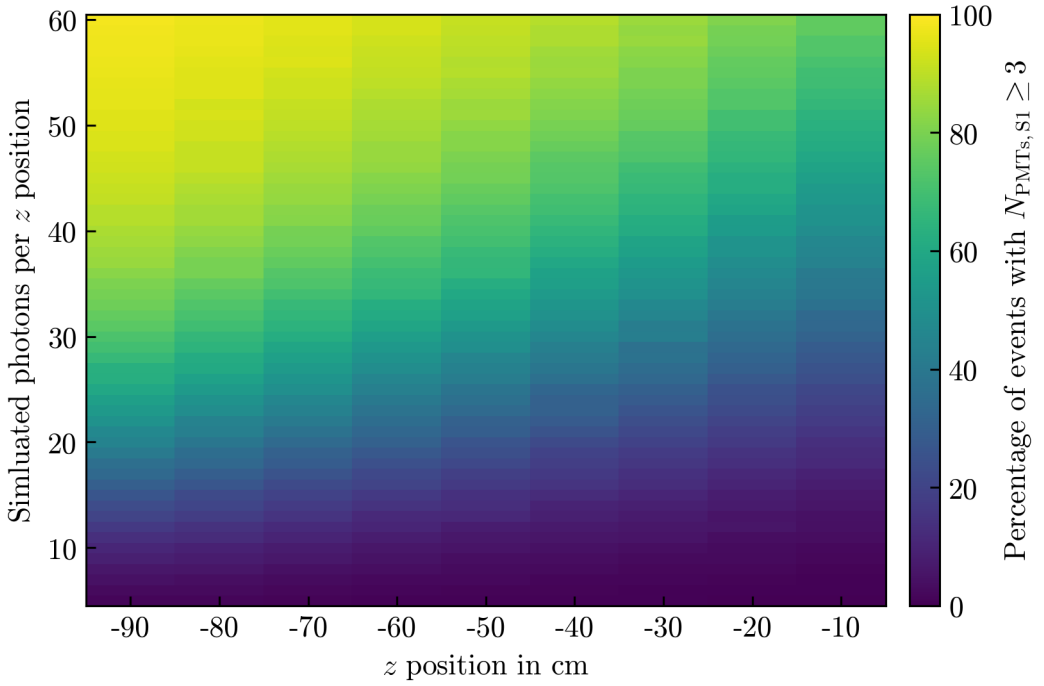

Figure 5.9: Two dimensional histogram with a simulation on how often a signal exceeds the tight coincidence requirement. The probability for an event based on a number of simulated photons to overcome the threshold of $N_{\mathrm{PMTs,\,S1}} \geq 3$ is plotted depending on the z-position of the interaction point and the primary number of simulated photons.

For this simulation, events with simulated numbers of photons from 5 to 60 are considered, since this range covers nearly the complete probability spectrum. Each of those photons is placed as explained above in one of the channels. The simulation is done at 9 different z-positions inside the TPC (at $z = -90, -80, ..., -10$), while the x- and y-coordinates are kept at 0, leading to $56 \cdot 9 = 504$ different simulation points. A total number of 7500 simulations is performed in each of these points.

Figure 5.9 shows the result of the threshold simulation. Note that the number of simulated photons is not the same as the previously used unit of photoelectrons. The registered PE are reduced in comparison to the generated or simulated photons by e.g. the QE. Thus, the ^{37}Ar S1s correspond to the upper part of fig. 5.9. In general, the chance is higher that at least three different PMTs register a photon for a larger number of initial photons in each simulated event. The number of generated photons in an event corresponds to the incident energy and should, therefore, be at a similar level for all events in a real, mono-energetic calibration source.

Referring to the same source, i.e. approximately the same number of photons per event, the probability to overcome the coincidence threshold is higher for a z-position close to the bottom array. Transferred to the behaviour of a possible threshold function, this would suggest a low detection threshold for events occurring close to the bottom of the TPC. At positions close to the gate, the threshold or the percentage of events that are not registered in the detector would be higher. Since most of the light signal is registered by the bottom array, the distance to be covered is smaller for events originating from positions close to the cathode, making a lower threshold for these events plausible.

This assumed threshold behaviour shows the opposite trend as the fit parameter x_{thr} from eq. (5.1), which is plotted in the lower panel of fig. 5.6. Here, the threshold position is higher for events from the bottom of the TPC. For events being positioned in a very short distance of the cathode, a possible explanation for an effect like this would be that, viewed from the interaction site, a single PMT takes in a larger solid angle than from a position further up in the TPC. This would lead to the majority of generated photons being registered by the same PMT, which could cause the tight coincidence requirement not to be fulfilled.

The simulation suggests that if this effect exists at all, it only plays a minor role for events very close to the cathode. In conclusion, the model without a direct detection threshold is supported by this simulation.

5.4 S2 correction

Being a mono-energetic calibration source with a homogeneous distribution inside the TPC, ^{37}Ar can also be used to derive an electron lifetime. The procedure is the same as in section 4.3.

The argon data are not as homogeneous as the krypton data. As investigated in [59], they show an anomalous behaviour in about the top quarter of the TPC. At these heights, the S2 area shows a slower increase than predicted from the data in the rest of the TPC. This effect is associated with the charge signal width. The width of an S2 is depending on its z-position in the TPC: a longer drift time, corresponding to a position at the bottom of the TPC, leads to an increasing signal width due to diffusion. The observed effect of smaller S2 areas at the top of the TPC is assumed to result from the combination of the position dependency of the width with afterpulses in the PMTs. Afterpulses have a constant time behaviour, they occur a few µs after the triggering signal, regardless of its width or energy.

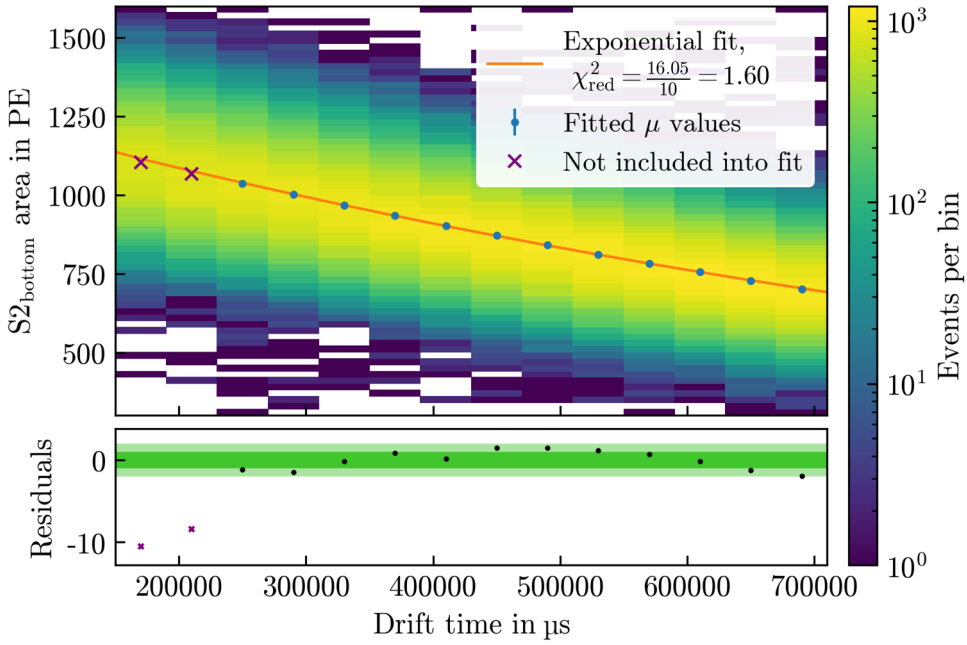

Figure 5.10: Two-dimensional histogram in the $S2_{\text{bottom}}$ and drift time space with the ^{37}Ar lifetime fit (orange line). The blue points used for the fit are obtained by a Gaussian fit of the $S2_{\text{bottom}}$ area in each drift time slice. The first two values (marked with crosses) are excluded from the fit due to the anomalous behaviour of the ^{37}Ar data at the top of the TPC. The resulting lifetime is (1.13 ± 0.01) ms.

Now, the hypothesis from [59] is that afterpulses of S2s with a large width are merged into the main S2. Meanwhile for a smaller S2 width, the afterpulses are separated from the S2 and registered as a following, alternative S2. Due to this better resolution in the top of the TPC, the S2 area does not increase as in the rest of the TPC.

A possible explanation why this effect is not seen for ^{83m}Kr data is that the width is also influenced by the energy of an event. With a higher energy, the number of generated electrons is higher and therefore also the time window where they arrive at the extraction field is larger, leading to a higher width. This higher overall S2 width prevents the afterpulses from being registered as a separate S2, biasing the S1s towards larger areas.

In reaction to this behaviour, the two highest slices in drift time are not considered for the argon electron lifetime fit,[2] corresponding to a bit more than the highest 20 cm. Apart from that, the fit is performed exactly as in section 4.3. It

[2]Utilising the largest part of the TPC, even though the more accurately reconstructed S2s are discarded.

is shown in fig. 5.10. The inclusion of the first two data points would lead to an increased value of $\chi^2_{\text{red}} = \frac{112.9}{12} = 9.41$, compared to $\chi^2_{\text{red}} = \frac{16.05}{10} = 1.60$. The residuals of the remaining fit range are not distributed symmetrical around zero. This could be caused by a remaining z-dependency in the data selection, other than the electron lifetime. Since the size of the afterpulse signal is observed to noticeable influence the complete S2 area, a varying contribution of the afterpulses with z could lead to the structure in the residuals. Another possibility is that the mean values of the S2 distribution were not determined accurately, for example due to the S2 distribution deviating from a Gaussian.

The electron lifetime of $(1.13 \pm 0.01)\,\text{ms}$ determined by the fit is at a much higher value than the previous lifetime from the krypton fit. Partially, this can be explained by the different recording times of the two datasets. While the argon data originate from SR2, the krypton data were recorded during SR1. Due to the continuous work of the xenon purification system and a pump upgrade [60] for SR2 increasing the purification flow the overall lifetime was raised during the operating time of XENON1T [41]. At the period of ^{37}Ar data taking, a lifetime of $\sim$850 µs was determined [61]. Thus, also under consideration of the higher lifetime for SR2, there remains a difference between both calibration sources. The source-dependent electron lifetime is also observed using background data from the radon decay chain (^{222}Rn and ^{218}Po) [41]. While it is not fully clear why this is the case, a hypothesis explaining this effect might be based on a different reaction of the calibration sources to inhomogeneities in the drift field.

6 Noble element simulation technique model

In chapter 5, the data from the argon calibration source were analysed in a similar way as the ^{83m}Kr data before. In both the considered charge and light signal corrections, differences were observed between the two sources. With the goal of gaining a higher understanding of these differences, in this chapter the noble element simulation technique (NEST) [62] is introduced and employed.

6.1 NEST and NESTpy

NEST is a simulation software for light and charge signal generation in noble elements [63], especially for argon and xenon. It uses a semi-analytic model in order to derive charge and light yields of electronic and nuclear recoils for various interaction types. Here, only the electronic recoils are considered. NEST establishes a model which depends on the interaction energy and also takes an applied electric field into account. The parameters of the model are adjusted to the available data on electronic recoils supplied by several experiments. The model from [63] is shortly reflected in the following.

As introduced in section 3.1, LXe has three channels where the energy of an interacting particle can be deposited: excitation, ionisation and heat. Since heat cannot be recorded in a LXe detector, it is not provided by NEST. For simplicity, the energy per ionisation W_{ion} and per excitation W_{ex} is described by the same unified work function $W = 13.7\,\mathrm{keV}$. With the ratio $\alpha_{\mathrm{ex}} = N_{\mathrm{ex}}/N_{\mathrm{ion}}$, this leads to

$$E_{\mathrm{dep}} = N_{\mathrm{ion}}W_{\mathrm{ion}} + N_{\mathrm{ex}}W_{\mathrm{ex}} = (1 + \alpha_{\mathrm{ex}})N_{\mathrm{ion}}W, \tag{6.1}$$

where N_{ex} is the number of excitons per interaction and N_{ion} is the number of ionisations. For α_{ex}, values from 0.06 [64] to 0.20 [65] are discussed. This means that most of the scintillation light is created by recombination, since the recombined electron-ion pairs contribute to the scintillation light. Therefore, the number of photons is strongly dependent on the recombination probability r of the ionised atoms. The numbers of produced quanta translate into photons N_{ph} (S1) and

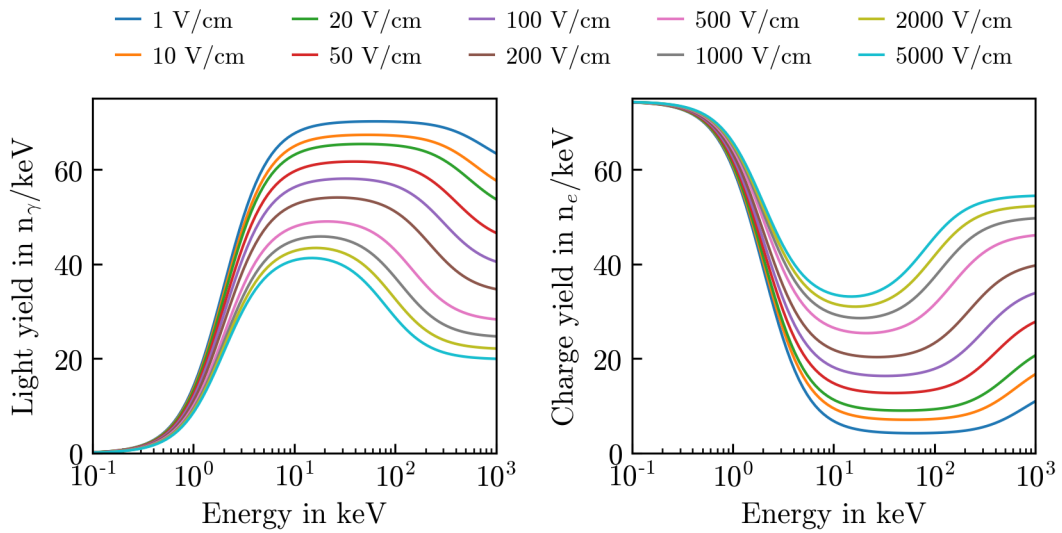

Figure 6.1: Light and charge yields resulting from the NEST model for a selection of drift fields. The left panel shows the light yield, while the right panel shows the charge yield. Light and charge yield are anti-correlated.

electrons N_e (S2) via

$$N_\mathrm{ph} = N_\mathrm{ex} + rN_\mathrm{ion} \qquad \text{and} \qquad N_\mathrm{e} = N_\mathrm{ion}(1-r). \tag{6.2}$$

NEST uses a combination of two models for the recombination probability. For a low deposited energy or short electron tracks, the Thomas Imel box model [66] describes the recombination as constricted to a box close to the interaction position. For a high deposited energy or long particle tracks, the Doke/Birk model [67] based on Birk's Law [68] is employed, assuming that the ionised electrons form a track of electron-ion pairs. The parameters of the models and the transition between them are determined by an adjustment to experimental data. The light and charge yields of the resulting NEST model are plotted in fig. 6.1. The light yield and the charge yield are defined as the number of photons and electrons per deposited energy, respectively. The light yield decreases for high energies (a feature of the Doke/Birk model) as well as for low energies (from the Thomas Imel box model), leaving a maximum light yield region at the transition region in the order of 10 keV. The yields with the underlying recombination probability as well as the transition between the two models are influenced by the applied electric field strength. With a growing electric field, the recombination probability decreases due to the electric field induced scintillation quenching [24]. As more electrons

succeed to escape a recombination, a higher electric field leads to a larger S2 and a smaller S1.

Based on W and α_{ex}, NEST begins with calculating the light and charge yield values. The calculation depends on the interaction type (various models for NR like neutrons or WIMPs and ER like photons or $^{83\mathrm{m}}$Kr), the incident energy and the applied drift field. The yields have deterministic values and do not change as long as the underlying properties are not changed. Based on the yields, the number of excited and ionised atoms is determined. At this stage, a statistical fluctuation is introduced by obtaining the number of actual particles by random numbers drawn from Gaussian distributions for N_{ex} and N_{ion}. Then, the number of escaping electrons is calculated with the field dependent recombination probability from the number of ions. On the recombination level, a skewness of the Gaussian distribution is introduced. This is explained in the following section. The number of photons is then obtained by adding the remaining, not-escaped ions to N_{ex}.

All these considerations deal with LXe as a scintillation medium, independent of the detection hardware. NEST would also allow the further modelling of a specific detector up to the resulting S1 and S2 signals, where effects like drift field inhomogeneity, electron lifetime and light collection efficiency can be considered. This is not required in the scope of this work, as only the yield calculation and the quanta generation are used.

6.2 Skew-Gaussian recombination model

With the recombination step, an additional statistical fluctuation is introduced (as in [69]). The number of escaping electrons is drawn from a skew-Gaussian distribution. The skew-Gaussian is defined by a Gaussian distribution that is modified by an error function, leading to

$$f(x,\xi,\omega,\alpha) = \frac{1}{\sqrt{2\pi}\omega} e^{-\frac{(x-\xi)^2}{2\omega^2}} \left(1 + \mathrm{erf}\left(\frac{\alpha(x-\xi)}{\sqrt{2}\omega}\right)\right). \tag{6.3}$$

The strength of the modification is defined by the skewness parameter α. Negative values of α shift the distribution towards lower values and the other way round. For the case $\alpha = 0$, an unmodified Gaussian distribution is obtained. In this limit case, the mean μ of the distribution is given by ξ and the standard deviation by

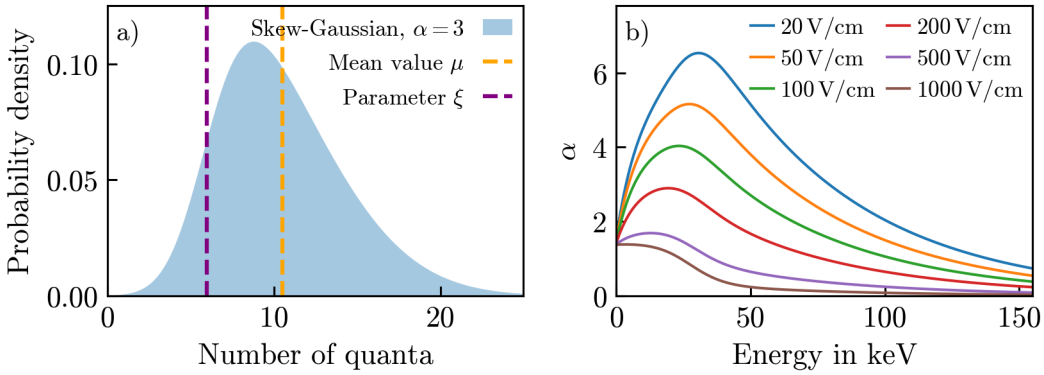

Figure 6.2: a) Probability density of a skew-Gaussian distribution against the obtained number of quanta with a mean value of $\mu = 10.5$ (orange dashed line) and a skewness parameter of $\alpha = 3$. The position of the parameter ξ is indicated by the purple dashed line. b) Skewness parameter α for the electron recombination versus the incident energy for different drift fields as implemented in the NEST model.

ω. For $\alpha \neq 0$, this changes to

$$\mu = \xi + \sqrt{\frac{2}{\pi}} \frac{\alpha\omega}{\sqrt{1+\alpha^2}} \qquad \text{and} \qquad \sigma^2 = \omega^2 \left(1 - \frac{2}{\pi} \frac{\alpha^2}{1+\alpha^2}\right) \tag{6.4}$$

for the mean value and the standard deviation. A visualisation of the function with $\alpha = 3$ and the mean value μ and the parameter ξ in comparison is shown in fig. 6.2a).

The actual number of electrons N_e is then obtained as a randomly drawn number from the skew-Gaussian distribution, centred at $(1-r)N_{ion}$. The number of photons is given by the remaining quanta:

$$N_{ph} = N_{ex} + N_{ion} - N_e. \tag{6.5}$$

In the NEST model, the skewness parameter α is not constant, but changes with the deposited energy and the drift field. For several field values, the energy dependent NEST skewness parameter, describing the skewness of the electron signal generation, is plotted in fig. 6.2b). As explained in [69], its modelling is based on the skewness of the log(S2/S1)-band of ^{3}H and ^{14}C calibration data from the Large Underground Xenon experiment (LUX). Here, the measured skewness in the log(S2/S1)-band is not the same as the modelled skewness from the recombination process, which is adapted during the calibration process in order to match the

skewness of the log(S2/S1)-band.

6.3 Application to ^{37}Ar data

The idea of a skewness introduced on the recombination level is adopted for the fit of the ^{37}Ar data. Both light and charge signal are considered in this section.

^{37}Ar S1 data

In NEST, the skewness is directly applied to the generation of escaping electrons. The distribution of photons is obtained as the remaining number of ions, added by the (usually small) amount of direct excitations. The number of primarily created ionisations (N_{ion}) fluctuates between the individual energy depositions and is taken from a Gaussian distribution. Now, the recombination model determines with a skew-Gaussian distribution which of these ions escape the recombination, leading to an accordingly skewed distribution of N_{e}. The remaining, recombining part of the primarily created ions (corresponding to a large percentage of N_{ph}) is also described by a skew-Gaussian shape, but with an anti-correlated skewness. The already determined number of photons is then added by N_{ex}, which itself originated from a Gaussian distribution without a skewness. Due to this summation, the resulting photon distribution as predicted by NEST can be shifted to a lower skewness value than the electrons are showing. Thus, the skew-Gaussian can also be used as a fit function for the light signal.

For an uncertainty estimation avoiding correlations, the fits are performed directly with the mean value μ from eq. (6.4) as an input parameter instead of ξ. Also, a scaling parameter is added in order to take different numbers of events per bin into account. The fits again use a maximum likelihood estimation with the accordingly constructed residuals. Also, the binning is chosen as in fig. 3.5. An example fit of bin 24 is plotted in fig. 6.3, yielding a reduced χ^2 of 1.13. As a goodness of fit measure for a Poisson likelihood fit, the reduced χ^2 value is calculated using the sum over the squared residuals normalised by the degrees of freedom.

When fitting the data in each bin separately, the skewness parameter α is allowed to vary throughout the TPC. The skewness is assumed to be created on the recombination level and should therefore not be position dependent.[1] The course of α over the TPC bins with an independent fit in each bin is plotted in green

[1] Under the assumption of a homogeneous electric field

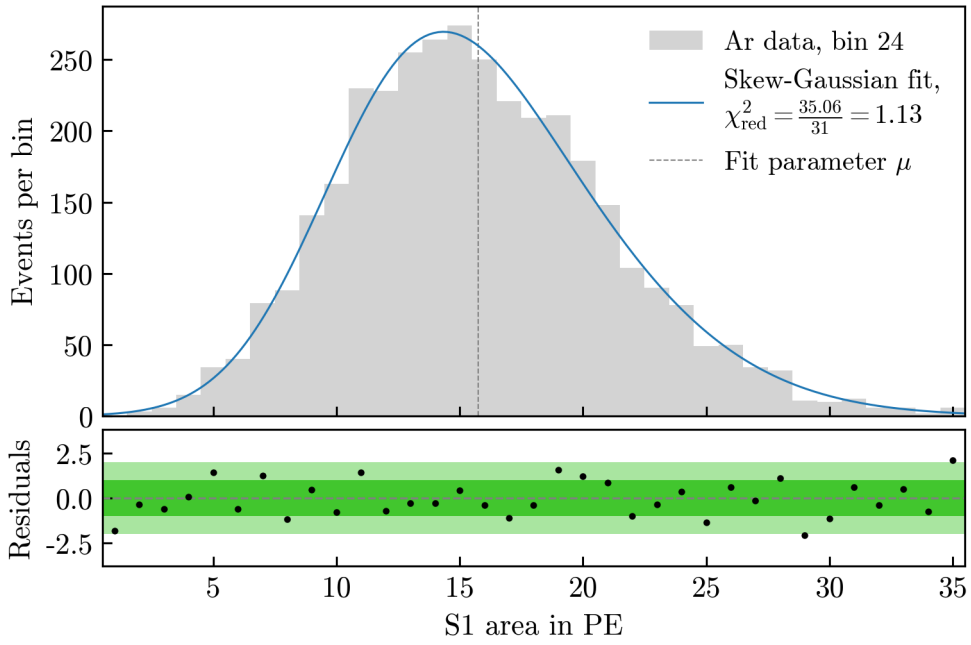

Figure 6.3: Data contained in bin 24 in a histogram of the S1 area (grey). A skew-Gaussian function (eq. (6.3)) is used for a fit of the data (blue). The lower panel shows the residuals of the fit.

in the lower panel of fig. 6.4. A slight trend towards higher values of α for bins further at the top of the TPC is visible. In order to avoid the position dependency of the skewness parameter, a simultaneous fit of all bins is performed. Instead of looping over the histograms of the TPC bins consecutively, all of them are passed to a single fit function. Each bin continues to have its own parameters for μ, ω and the scaling, while the skewness parameter α is shared among all bins. The resulting skewness of this shared fit is given by $\alpha = 1.89 \pm 0.02$ and is marked by the red line in the lower panel of fig. 6.4, where it takes a reasonable position as an average value. Summing over χ^2 in all 72 TPC bins involved in the simultaneous fit with its 217 fit parameters, a reduced χ^2-value of $\frac{2501}{2231} = 1.12$ is reached. Due to the high number of degrees of freedom, in this case $\chi^2_{\text{red}} = 1.12$ indicates a not optimal fit. This can be explained by the observed position dependence of α forced to be reduced to a single number.

A direct comparison between the NEST values for the skewness parameter and the value from the shared fit is not possible, since the NEST model directly describes the skewness of the electron recombination. The skewness of the light signal results from the relation in eq. (6.5). Following the NEST model, with this

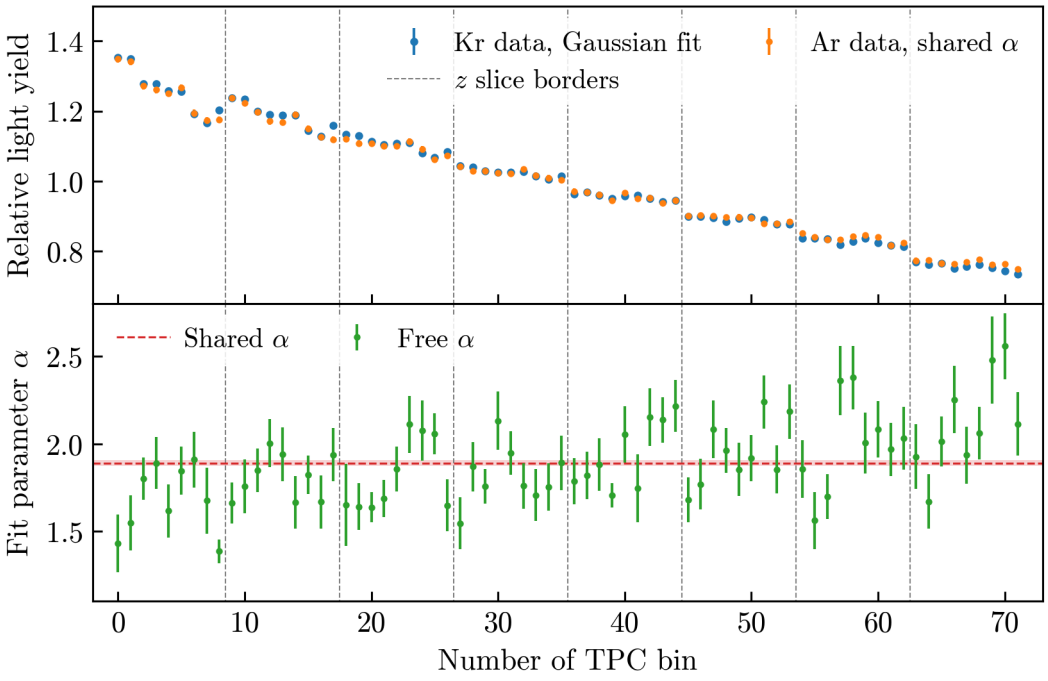

Figure 6.4: Upper panel: Relative light yield in each TPC bin, obtained by a skew-Gaussian fit in all TPC bins in parallel with a shared parameter α (orange). For comparison, the result from the krypton data with a Gaussian fit is given in blue. Lower panel: Comparison between independent fits in all bins with individual values for α (green) and the combined fit with a single, shared parameter α in all bins (red). The translation of the bin numbers to their positions is given in fig. 3.5.

relation the positive skewness for the electron recombination is linked to a negative skewness of the light signal's distribution, since light and charge signal add up to a normal distribution. The observed positive skewness for the light signal of the combined fit is therefore in contrast to the parameters of the NEST model.

Still, the mean value of each bin is used to calculate the relative light yield, which is plotted in the top panel of fig. 6.4. As a comparison, the result of the one-dimensional Gaussian fit of the krypton data is plotted. Those are the same data that were given for ^{83m}Kr in fig. 5.8. The ^{83m}Kr data show a larger overall variation of the LY, as already observed for the smeared Poisson function. Despite the disagreement with the NEST parameter prediction, the skew-Gaussian model fits well to the data and provides a relative LY that is comparable to the ^{83m}Kr data and to the smeared Poisson function. The graphical representation of the map is shown in the appendix (fig. A.7). After the application of the map, the

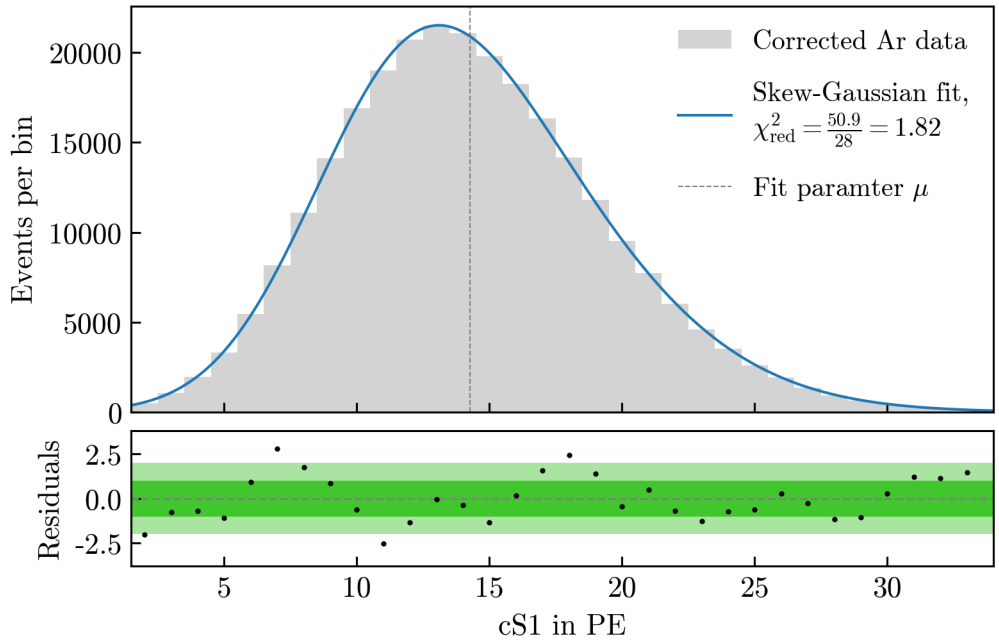

Figure 6.5: Histogram of the S1 ^{37}Ar data (grey) corrected with fig. A.7. Since the correction resolves the position dependency, the complete dataset is fitted by a skew-Gaussian (blue). 3 bins with less than 200 entries are not included in the χ^2-minimisation. The result of $\alpha = 1.85 \pm 0.02$ agrees with the shared fit of α on uncorrected data.

complete dataset is again fitted with a skew-Gaussian fit function. Since the S1 data are corrected to be at the same signal size, a division of the TPC into bins is not necessary anymore. With the exception of 3 bins at the border of the fit range, enough statistics is provided for a χ^2-minimisation. The fit is depicted in fig. 6.5 and has $\chi^2_{\text{red}} = 1.82$. As discussed for the simultaneous fit on the uncorrected data, this is possibly the result of an inhomogeneous electric field influencing the skewness. The fit yields a skewness of $\alpha = 1.85 \pm 0.02$, where the uncertainty interval still overlaps with the α from the shared fit. Therefore, it can be concluded that an application of the correction does not change the observed skewness to a large extent.

^{37}Ar S2 data

Since the evaluation of the light signals gave an unexpected sign for α, a follow-up check is done in order to investigate whether the skewness of the S2 data behaves like predicted by NEST. Before the shape of the charge signal can be examined, the

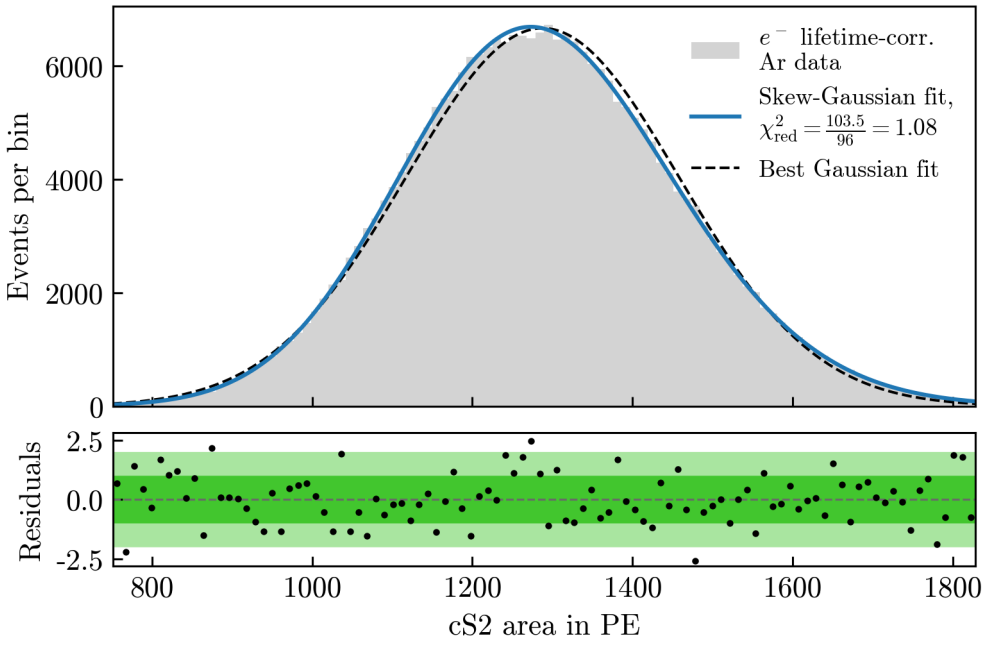

Figure 6.6: Histogram of the S2 area of the argon data, corrected according to section 5.4. The blue line is the skew-Gaussian fit with a skewness of $\alpha = 1.12 \pm 0.02$. In comparison, the best fit of a normal Gaussian fit is indicated by the black dashed line. The residuals of the skew-Gaussian fit are plotted in the lower panel.

electron lifetime correction needs to be applied. Without the correction, the data are smeared in their S2 area such that they cannot be fitted. Since the lifetime varies with time and a source dependency was detected, the correction derived in section 5.4 from the same dataset is used.

For the corrected dataset, the TPC is not resolved in various bins. Instead, a fit with the same skew-Gaussian function is performed on the complete dataset. Because of this, the statistic is high enough for a least squares minimisation. The fit range is set to 3 standard deviations from both sides of the mean value of the dataset. The fit is shown in fig. 6.6 and has a reduced χ^2-value of 1.08 at 96 degrees of freedom, suggesting a high goodness of fit. It results in $\alpha = 1.12 \pm 0.02$.

Since the skewness of the electron signal is directly described by the NEST model, it can be compared to fig. 6.2b). At small energies, the skewness parameter is modelled to be at around 1.5, which can be argued to be at least at a similar range with the same sign.

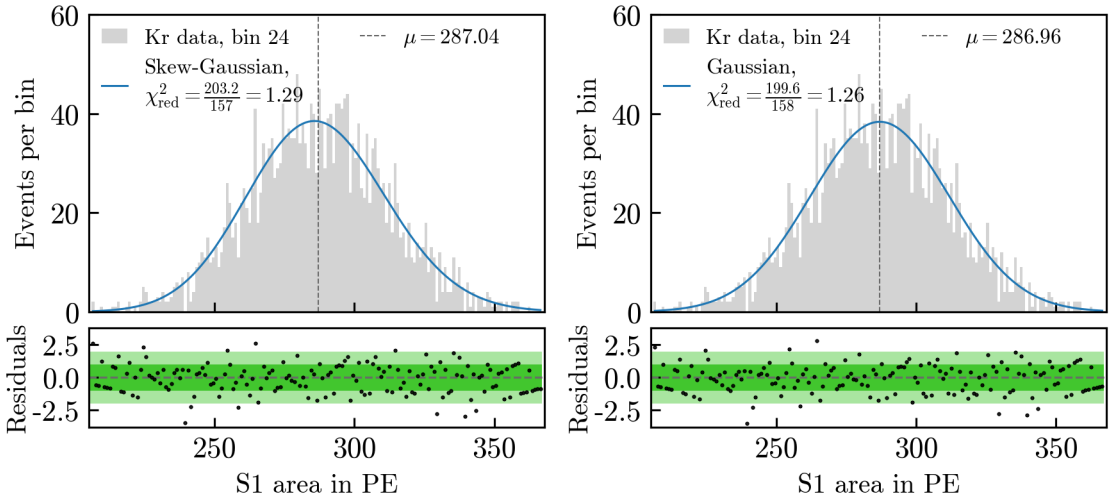

Figure 6.7: Histogram of the ^{83m}Kr data in bin 24 of the TPC (grey), see fig. 3.5. In the left panel, the corresponding parameters from the shared skew-Gaussian fit with $\alpha = 0.91$ are used for the plot of the fit function and the calculation of the residuals. The right panel shows the Gaussian fit in bin 24 with its mean value and the residuals.

6.4 Application to ^{83m}Kr S1 data

Until now, the modification of the Gaussian peak shape only was applied to the low-energetic argon data. Meanwhile, the krypton data were assumed to be Gaussian-distributed and their one-dimensional Gaussian fit to the S1 area was used as a reference. Following the NEST skewness model shown in fig. 6.2, the skewness of the charge signal and therefore also of the light signal is not restricted to appear at low energies. Thus, the krypton data should also exhibit a deviation from the plain normal distribution, making a simple Gaussian fit unjustified.

In order to test this, the same procedure as for the argon S1 data is applied to the krypton S1 data. The parameter α is shared among all bins, while the rest of the fit parameters are individually adjusted for each bin with a maximum likelihood fit. This results in a skewness for the ^{83m}Kr data of $\alpha = 0.91 \pm 0.28$. A comparison between the simple Gaussian and the skew-Gaussian fit is shown in fig. 6.7. The resulting mean values do not show a large deviation, and also the residuals are very similar. For the skew-Gaussian fit, this results in $\chi^2_{red} = 1.29$, while the normal Gaussian fit yields $\chi^2_{red} = 1.26$. This indicates a slight preference for the Gaussian fit, compared to the fit with a shared α.

For a skewness below 1, the skewness does not have a large influence on the mean

value. Also, the data can be described similarly good using the plain Gaussian fit. This small relevance of a skewness parameter below 1 might explain the large uncertainty that the shared fit assigns to the skewness parameter, since the possible skewness values in the uncertainty range could lead to a similarly good fit result.

The nearly vanishing skewness for the ^{83m}Kr double selection disagrees with the NEST prediction, which forecasts a maximum of the S2 skewness (and with that also for S1) at energies around 30 keV (see fig. 6.2), which is the energy range of the ^{83m}Kr transition. Therefore, the NEST parameter values do not fit to the results from comparisons to ^{83m}Kr and ^{37}Ar S1 data and ^{37}Ar S2 data. At least it legitimates the usage of a Gaussian fit function on the ^{83m}Kr S1 data. If the skewness of the ^{83m}Kr data behaved as predicted by NEST, the production of a relative LY correction map based on a Gaussian fit would not be justified.

6.5 Comparison with NEST simulations

Besides comparing the skewness of the fitted data to the underlying parameter from NEST in fig. 6.2b), the software can also be used to simulate the signal generation directly. For the simulation of the ^{37}Ar data, the ER interaction type simulating γ-rays is chosen. With an energy of 2.8 keV and the drift field value of SR1 (81 V/cm, [41]) the yields are calculated. The charge yield is at 34.90 electrons/keV, while for the light yield 39.48 photons/keV is obtained. For each of these yields, 4×10^5 interactions are simulated and the corresponding number of photons or electrons is stored. These numbers are then plotted in a histogram, which is depicted in fig. 6.8. On the binned values, a fit is performed with the same skew-Gaussian fit function as before. Due to the high number of simulated events, a χ^2 cost function is chosen. The skewness parameter for the electron distribution is at $\alpha = 2.28 \pm 0.01$ and for the photons at $\alpha = -1.90 \pm 0.01$. This is coherent with fig. 6.2b) and with the earlier expectation of a negative skewness for the light signal based on the NEST recombination model. The smaller absolute value of the skewness of the light signal can be explained by the Gaussian-distributed fraction of photons generated by the direct excitations processes.

The same approach is repeated for the 32.2 keV krypton line. For the yield calculation, the ^{83m}Kr model of NEST is applied. This model distinguishes between the 9.4 keV and the 32.2 keV transitions. For an energy of 9.4 keV, the time difference dt to the previous 32.2 keV decay is taken into account for the yield calculation. If no specific dt is passed, a time is drawn from an exponential distribution with the

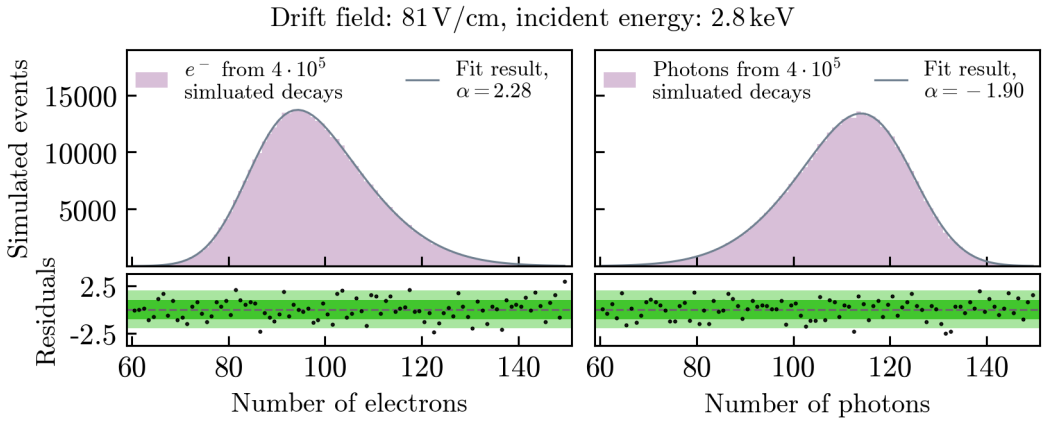

Figure 6.8: NEST simulation of electrons (left) and photons (right). For 4×10^5 decays, the resulting electrons and photons are plotted in a histogram (violet). The grey line is the skew-Gaussian fit of the histogram with a skewness of 2.28 ± 0.01 for the S2 and -1.90 ± 0.01 for the S1.

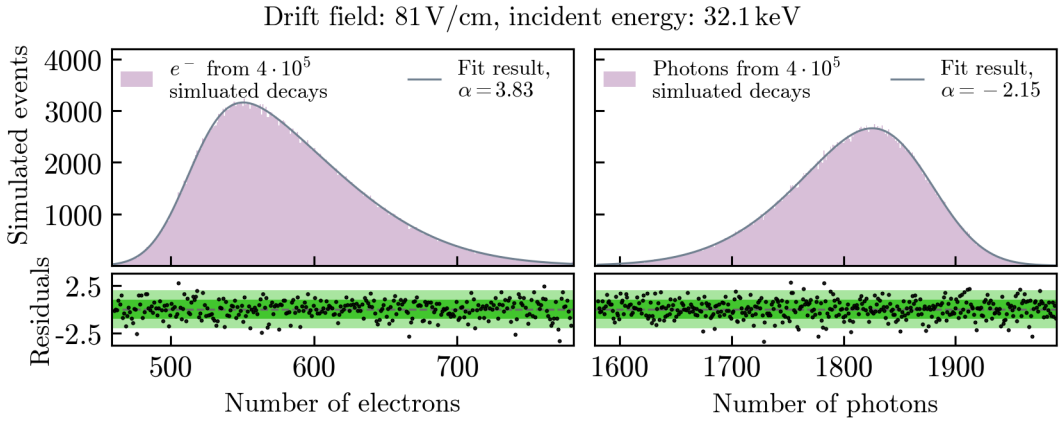

Figure 6.9: NEST simulation of electrons (left) and photons (right). For 4×10^5 decays, the resulting electrons and photons are plotted in a histogram (violet). The grey line is the skew-Gaussian fit of the histogram with a skewness of 3.83 ± 0.02 for the S2 and -2.15 ± 0.01 for the S1.

half-life of 154 ns. The yields of the 32.2 keV decay are independent of the choice of dt and only depend on the drift field. For the 81 V/cm of SR1, a photon yield of 56.29 photons/keV and an electron yield of 18.08 electrons/keV is obtained. With these yields, the same amount of simulated decays is generated as for the ^{37}Ar calibration energy. The histograms of the distributions and the skew-Gaussian fits are shown in fig. 6.9. With a skewness for the charge signal of 3.83 ± 0.02 and of -2.15 ± 0.01 for the light signal, the model in fig. 6.2 is reproduced. This

underlines the contradiction between the NEST model predicting a large negative skewness for a 32.2 keV light signal and the fit of a nearly vanishing skewness on the data.

In [70], the described skewness model is applied to a preliminary version of the argon calibration data from XENON1T. A fit to the 2.8 keV peak on a combined energy scale derived from both corrected light and charge signal results in a skewness of 1.51, with an estimated uncertainty of ± 0.2. This further confirms the observation of a positive skewness in XENON1T data also for the light signal.

The skewness parameter of the electron recombination probability was tuned on ^{3}H and ^{14}C LUX calibration data in order to fit the shape of the log(S2/S1)-band [69]. Apparently, the effect on the raw shape of the light signal was not sufficiently considered. Without any observed negative skewness in real data, the currently implemented recombination model cannot be supported. A possible alternative could be to include the skewness already on the lower level of ion generation, since the skewness of light and charge yield observed in the data is not symmetrical centred at zero. Since in NEST the quanta generation and the recombination model are built on top of the yield calculation, the observed discrepancy of the skewness does not allow to make conclusions on the implemented yield model.

Even though the NEST parameter values of α could not be confirmed in the data, the model itself turned out to be capable of describing the deformation of the argon data and can be used for the generation of a relative light yield map. A possibility not considered in this chapter is the combination of a skewness arising directly from the ^{37}Ar data and the effect of a low-energy detection threshold. Since section 5.3 concluded that such a threshold does not have a strong effect on the ^{37}Ar data, it cannot be used to explain the discrepancy between the NEST parameter α and the observation from ^{37}Ar. Also, the discrepancy for the ^{83m}Kr data would remain, since a detection threshold for low-energetic decays does not impact the 32.2 keV line.

7 Field inhomogeneities in the TPC

The model from chapter 6 includes a field dependency for the skewness parameter. By treating α as constant throughout the TPC, a homogeneous drift field is assumed. In this chapter, the homogeneity of the drift field in the XENON1T calibration data is investigated by its effect on the S1 areas of ^{83m}Kr signals.

7.1 S1b/S1a ratio of ^{83m}Kr

Since the variation of the electric field could not be measured directly, the local field strength has to be inferred from calibration data. One possibility is to take advantage of the fact that the effect of a varying drift field on the recombination probability is smaller for low-energetic electronic recoils than it is at higher energies. The border where the field starts to play a larger role for the recombination probability and therefore also on the S1 (and S2) yield has been determined to be at about 10 keV [71].

With the two decays of ^{83m}Kr placed above and below this border, the previously analysed calibration data can also be used for mapping the field strength. Since the relative LY map performs an effective correction of all effects that impact the S1 size, the effect of the field dependency is also impacted. Therefore, the investigation of possible inhomogeneities has to be done on a dataset without the LY correction. In order to filter out the geometric detector effects, the ratio of the two areas S1b/S1a is considered (as in [71, 72]). Since the S1a and S1b transitions from the same event take place at the same interaction site, considering the S1b/S1a ratio cancels these constant detector effects while maintaining the field dependency. The previously discussed field distortion correction aims to remove effects caused by a non-homogeneous field, but only acts on the reconstructed positions. Being a correction on the reconstructed event position in the TPC, it is required in order to base this analysis of the signal size on a correct position estimation.

The behaviour of S1a is dominant for S1b/S1a, since S1b is not strongly affected by field variations. Lowering the general S1 size, a stronger field therefore leads to a smaller S1a size and with that increases the ratio S1b/S1a. This can be

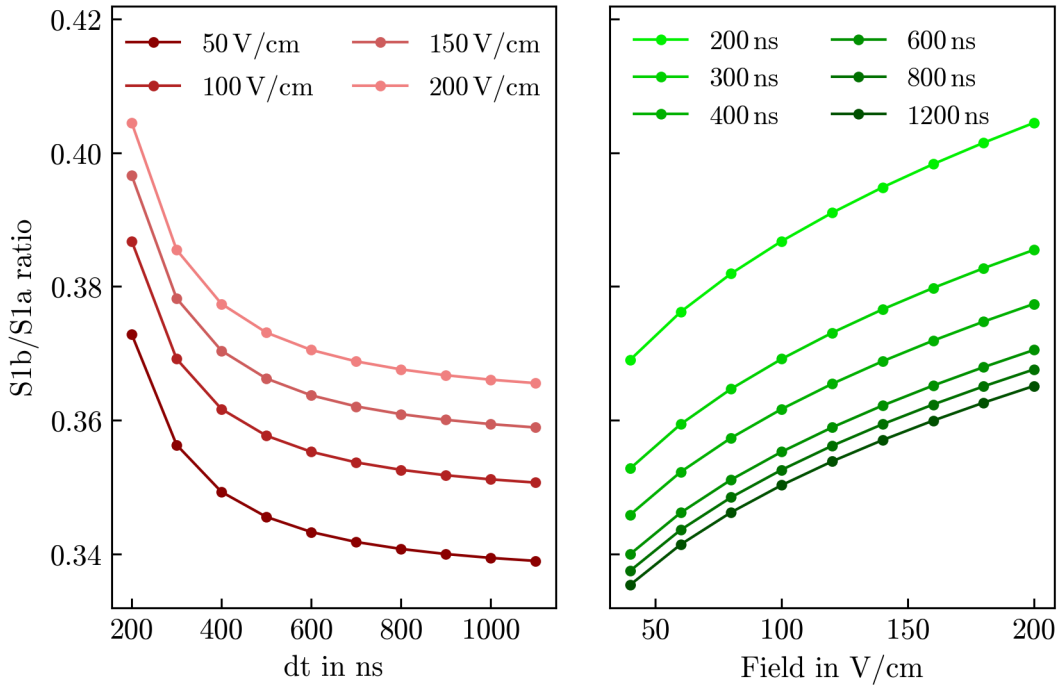

Figure 7.1: The ^{83m}Kr yield model from NEST is used for a prediction of the S1b/S1a ratio. The left panel shows its dependency on the delay time dt between S1a and S1b for four different field strengths, while the right panel shows the field dependency for several dt.

simulated on the photon yield level by the ^{83m}Kr model of NEST. These yields are calculated at a lower stage of the simulation, before the previously discussed generation of quanta. Thus, the detected discrepancy between skewness parameter and data does not affect the yield prediction. The NEST ^{83m}Kr yield model is adjusted to directly match ^{83m}Kr data measured by different experiments [63, 73]. Its prediction on the S1b/S1a ratio's field dependence is shown in the right panel of fig. 7.1. Next to the field, the ratio calculated by NEST depends on a second parameter. This is the time difference dt between S1a and S1b, which is shown in the left panel. It takes into account that the 9.4 keV signal size is affected by dt, as described consistently by [72, 74] and [75]. For dt $\lesssim$ 500 ns, an enhancement of S1b is observed, which is explained with more possibilities for recombination due to a leftover electron-ion cloud from the preceding S1a. Since this cloud diffuses as the time difference increases, the influence of this effect gets smaller for larger dt. The NESTpy implementation does not offer dt as a direct input parameter, thus the ^{83m}Kr yield model is reproduced for the creation of fig. 7.1 and the following analysis. As visible in fig. 7.1, the absolute impact on S1b/S1a is predicted to be

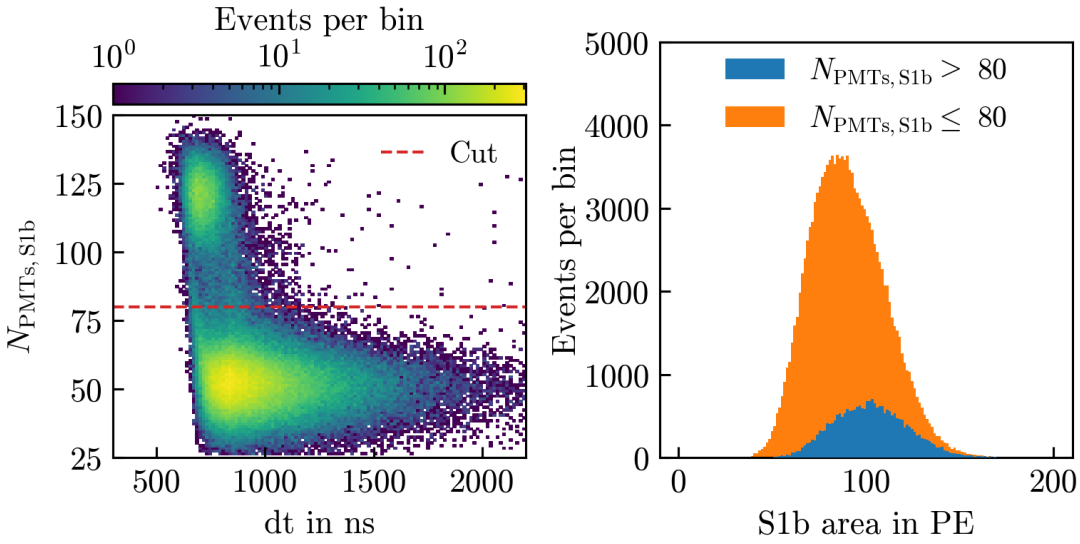

Figure 7.2: Left: Histogram of the strict selection obtained in section 4.1 in dt and $N_{\mathrm{PMTs,\,S1b}}$. Two populations are visible, indicating a not sufficient data selection for the S1b peak. Right: The observed populations are separated at $N_{\mathrm{PMTs,\,S1b}} = 80$ and plotted in a histogram of the S1b area. The population with a higher number of contributing PMTs is observed to have a higher S1b area. Thus, this observation influences the S1b/S1a ratio.

similarly large between dt = 200 ns and dt = 1000 ns at a constant drift field and between 50 V/cm and 200 V/cm at a constant dt. Since the influence on S1b/S1a is most sensitive for small dt, the previous rejection of events below 500 ns in fig. 4.7 contributes to the separation of these effects.

7.2 Data preparation

Until now, the S1b decay was only used for the selection and validation of the 32.2 keV events. Therefore, the S1b data themself were not checked for quality. A plot of the strict selection in $N_{\mathrm{PMTs,\,S1b}}$ and dt (see left panel of fig. 7.2) reveals the existence of two different populations. Since the calibration data are expected to have a homogeneous behaviour in all dimensions, this suggests that the data selection is not optimised for the S1b peaks. The influence on the S1b area is visualised in the right panel of fig. 7.2 by a separation at $N_{\mathrm{PMTs,\,S1b}} = 80$. Since the S1b/S1a ratio depends on the exact position of the S1b peaks, it is also affected by these two populations. In order to visualise this effect, for both populations S1b/S1a is calculated. This is done with the separation described above for different slices

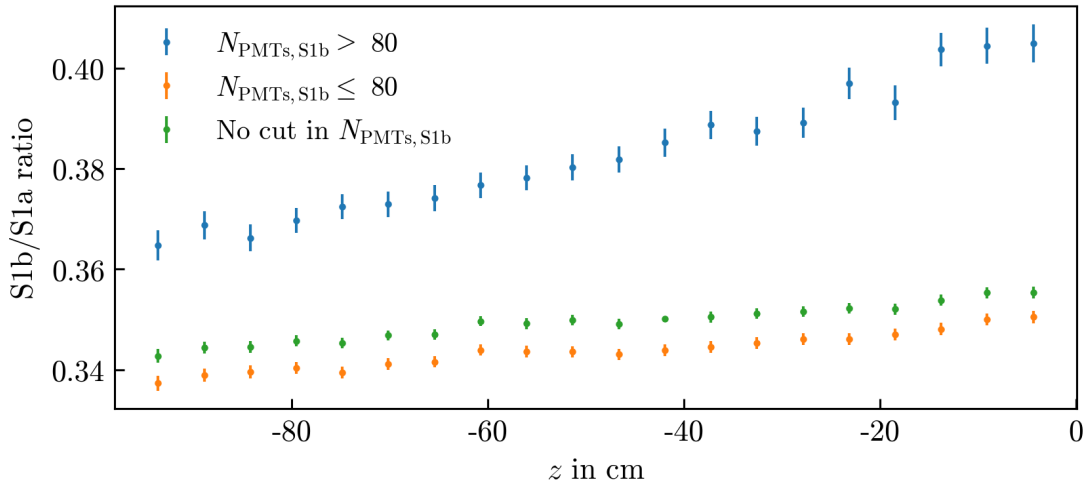

Figure 7.3: S1b/S1a ratios plotted against the central z-positions of axial slices of the TPC for three event populations. The separation discussed in fig. 7.2 is applied to the S1b/S1a calculation (blue and orange). The result of the two populations combined (green) lies closer to the population with less contributing PMTs, since it is weighted stronger due to the higher number of entries.

of the TPC. The events in each slice are fitted in S1a and S1b with a maximum likelihood estimation and the ratio in each slice is calculated. This is shown in fig. 7.3. The population with more than 80 contributing PMTs results in a higher ratio than the population below. Due to the larger population size of the lower population, the unseparated dataset lies closer to the ratio of the lower population. Compared to fig. 7.1, the S1b/S1a difference reached by the separation of the two populations is large and cannot be neglected.

The origin of the two populations is not clear, but it seems to be connected to the cut in fig. 4.7. This cut, utilising the number of distinct PMTs in S1b, was placed with the goal of rejecting PMT afterpulses. Since a 9.4 keV peak registered independently has no reason to be seen by roughly the same PMTs, setting a lower limit on this variable excludes events where the S1b peak was generated by afterpulses in the PMTs that were activated by the previous 32.2 keV decay. This approach would not remove events where both effects occur: the 9.4 keV peak itself is recorded correctly as the second largest peak in the event, but at the same time, afterpulses from the previous peak are recorded and added. Due to the enhancement of the number of distinct PMTs for S1b by the correctly identified 9.4 keV peak, the cut rejecting peaks based only on their distinct channels variable does not catch these events. Since an afterpulse adds only a small fraction of the

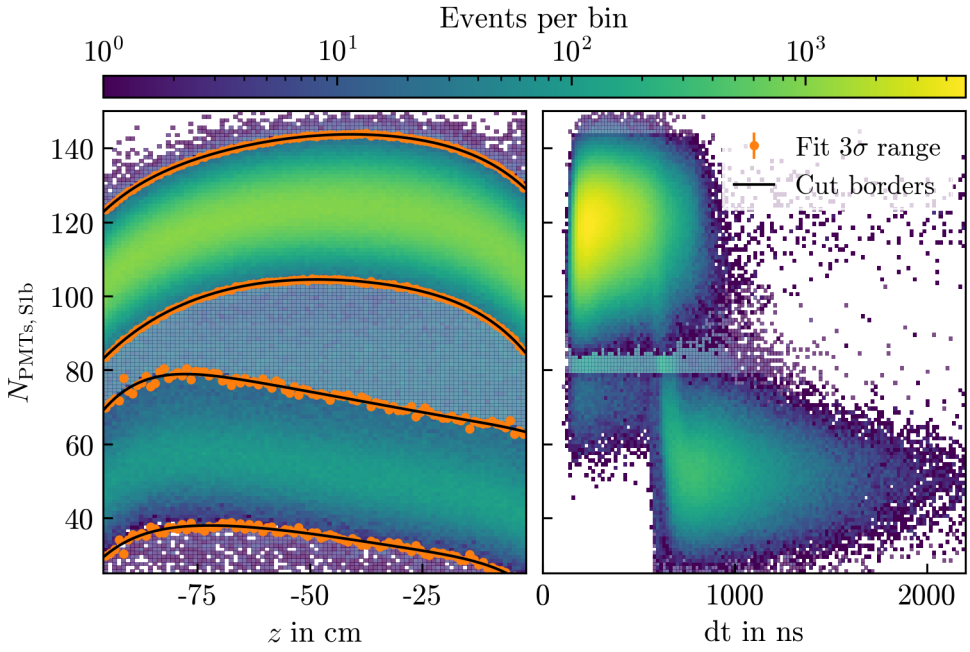

Figure 7.4: Left: Histogram in z and $N_{\mathrm{PMTs,\,S1b}}$ of the data selection with the S1b distinct channels cut from fig. 4.7 removed. The transparent dataset in the background is plotted without additional cuts. The orange points represent the 3σ range of a fitted sum of two Gaussian peaks for different z. The points are used for an interpolation with polynomials, which is depicted by black lines. Those are used as cut borders for the selection of the two populations with low and high numbers of contributing PMTs. Right: The populations selected in the left panel are plotted in a histogram of dt and $N_{\mathrm{PMTs,\,S1b}}$. The transparent events are the same dataset plotted in the background of the left panel.

overall area to the peak, it is not spotted directly in the distribution of the S1b area. Therefore, a possibility to reject these events is needed.

A possible alternative for the rejection was already indicated in fig. 7.2. Since the described effect is visible in the absolute number of PMTs contributing to S1b, this variable is examined more closely. The variable $N_{\mathrm{PMTs,\,S1b}}$ has a dependency on the height of the event in the TPC. At extreme values of z, the number of contributing PMTs is smaller. This can be explained with the effect of a larger solid-angle coverage of a single PMT close to the TPC edges. In order to directly see the effect of the afterpulses, the cut in fig. 4.7 is revoked. The so defined dataset is plotted in the background of the histograms in fig. 7.4. In the left panel, the distribution of $N_{\mathrm{PMTs,\,S1b}}$ over z is shown. This parameter space is used in order to define a cut separating the two populations while including the position dependence.

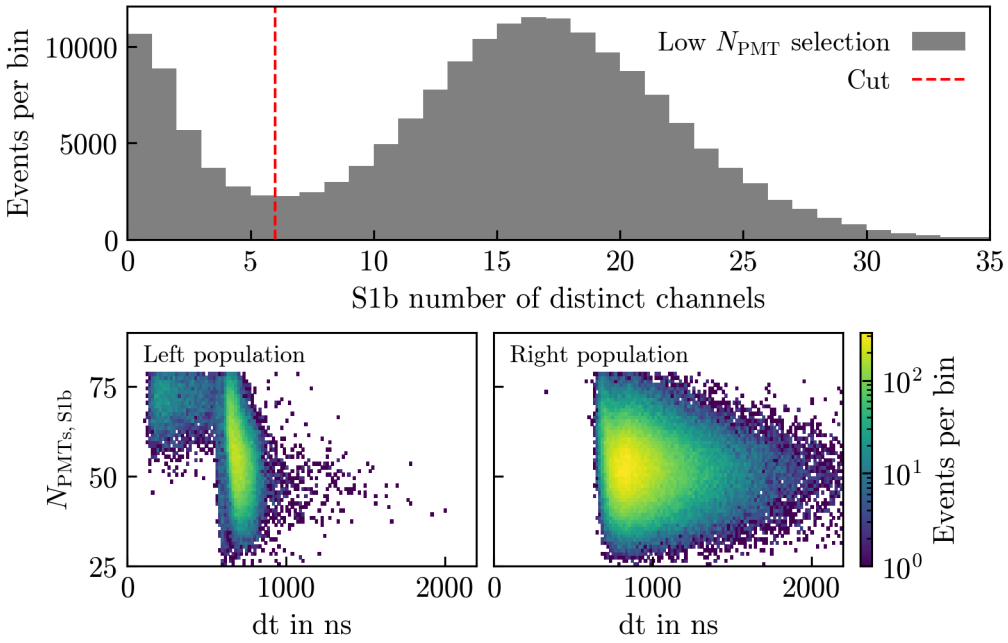

Figure 7.5: Top: After the cut in $N_{\mathrm{PMTs,\,S1b}}$, the lower population from fig. 7.4 is plotted in a histogram of the number of distinct channels for S1b. Again, two populations are observed. They are separated by the vertical dashed red line. Bottom: Histograms in dt and $N_{\mathrm{PMTs,\,S1b}}$. The two plotted populations are situated left and right of the separating red line in the top panel.

Therefore, the dataset is divided in z slices and each slice is fitted by a sum of two Gaussian peaks. A 3σ range around the respective mean values is marked by the orange points in fig. 7.4. In order to derive a continuous border for the placement of a cut, the fitted 3σ range is interpolated by polynomials. For this purpose, the numpy [37] functionality `polyfit` is used. The two resulting populations are plotted in the right panel of fig. 7.4, while the transparent background marks the rejected events. The lower population corresponds to a large part to the lower one in fig. 7.2, but with a successful rejection of the upper population. Thus, the cut presented here has a similar effect as the cut on the number of distinct PMTs given in fig. 4.7, while being more successful at rejecting those specific events.

Some events remain in the lower selection at low time differences. In order to check if they are removed by the cut in the number of distinct channels, this variable is examined again. A histogram of the lower population is plotted in fig. 7.5. Still, two distinguishable populations are observed. An additional cut in the number of distinct PMTs is performed in the minimum between them. As visible from the comparison of the two lower panels, this cut successfully removes

the remaining peaks at low time differences. With the combination of these two cuts, the revised population in the lower right panel of fig. 7.5 is obtained and used in the remainder of this chapter. This results in 145 311 usable events, which is comparable to the 165 032 events from the strict selection in section 4.1.

7.3 Field assignment

The obtained selection is again divided into z slices. For each slice, two Gaussian fits are performed in order to determine the means of S1b and S1a. An example of a fit of the 9.4 keV peak is given in the appendix in fig. A.8. The so obtained values are used for a calculation of S1b/S1a. Since the two values are obtained by independent fits and the detector geometry effects are cancelled out, the uncertainties of the mean values are considered uncorrelated. Thus, the combined uncertainty for the ratio is obtained by a Gaussian uncertainty propagation. The resulting ratios obtained by the fits are depicted by red dots and the scale on the left axis of fig. 7.6. The same increasing trend in z is observed as in fig. 7.3, indicating an increasing field strength towards the top of the TPC (cf. fig. 7.1). In a next step, the NEST model is used for converting each S1b/S1a value into an absolute field strength. As mentioned above, the ^{83m}Kr model depends on a specific value for the time difference, so that the enhancement of the S1b peak for low dt can be considered. Since the time difference distribution shows no position dependency, the arithmetic mean of the time difference distribution is used, yielding 969.8 ns. The converted values for the field strength are plotted in blue with the scale on the right axis. The highest and the lowest data point of both axes are aligned. Due to the nonlinear conversion which can be seen in the right panel of fig. 7.1, the data points do not directly correspond to the value at the same z-coordinate. For a better visibility, a distance of 0.5 cm on the z-position is inserted between both sets of points. The uncertainties of the field values are directly derived from the S1b/S1a uncertainties. Therefore, the maximum and the minimum ratio inside the uncertainty range are determined and the corresponding field range is calculated and used as the uncertainty. Since the transition between field and ratio is nonlinear, this would lead to asymmetric uncertainty bars. Since the uncertainties are used for a χ^2-fit in the next step, for simplicity the average of both deviations is calculated and used as a symmetric uncertainty.

A linear fit is employed, describing the field variation as a linear function of the z-coordinate with the parameters m as the slope and b as the intercept. For

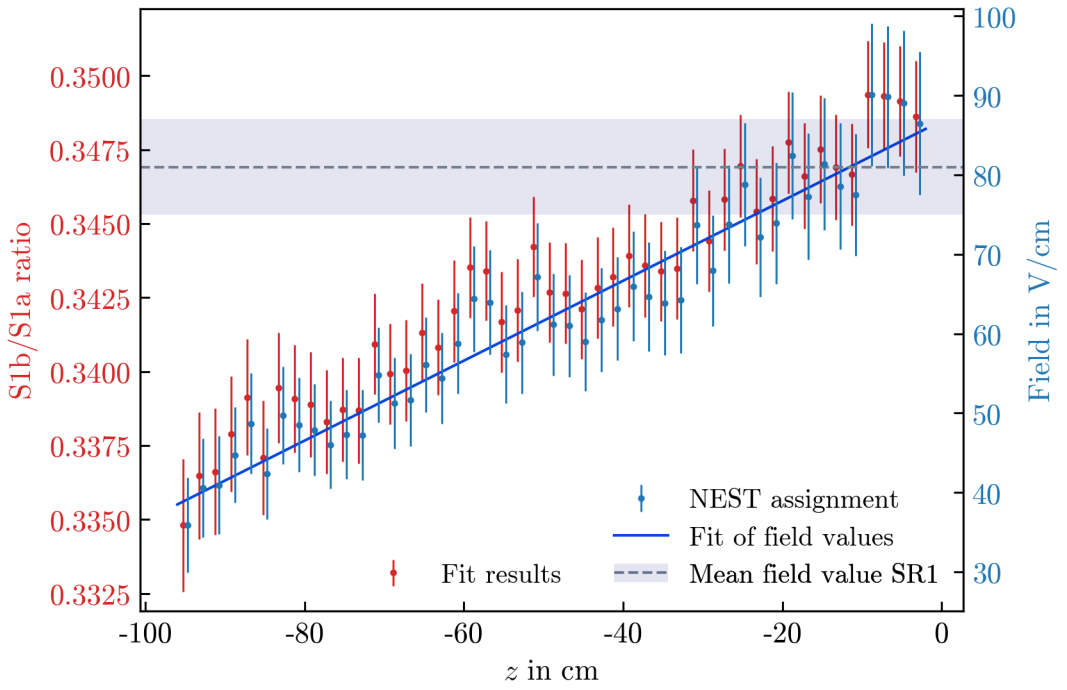

Figure 7.6: S1b/S1a ratio and assigned field strength against z on two vertical axes. The left axis with the red points shows the S1b/S1a ratio calculated from fits of S1a and S1b in various z slices. The blue points are connected to the right axis. They are obtained using NEST for the conversion of the ratio (red points) into field values and therefore share the z-coordinate. For a better visibility, a shift of 0.5 cm is inserted between each pair of points. Due to the nonlinear relation between ratio and field, red and blue points are not directly stacked, even though the axes are scaled in a way that the highest and the lowest red and blue points are aligned. The solid blue line is a linear fit of the field values with $\chi^2_{\text{red}} = \frac{12.7}{45} = 0.28$ and belongs to the right axis, as does the dashed horizontal line. For comparison, this line and the shaded region mark the mean field value in the fiducial volume of SR1 and its 1σ range.

the fit, a χ^2-minimisation is used, such that the previously calculated uncertainty on the field values are included. The fit yields $m = (0.49 \pm 0.04)\,\text{V/cm}^2$ for the slope, combined with an intercept of $b = (86.3 \pm 2.3)\,\text{V/cm}$. The result is plotted in fig. 7.6 as a solid blue line. Due to the large uncertainties on each data point, it has a reduced χ^2-value of $\frac{12.7}{45} = 0.28$. This indicates an overestimation of the uncertainties. With maximum likelihood fits used for the calculation of S1b/S1a, this is not expected, such that the reason for this overestimation is not clear. The dashed horizontal line marks the field value reported for SR1 of XENON1T [41] in the fiducial volume, while the shaded range describes its 1σ range.

With a function for the position dependent field strength, the overall potential difference causing the field can be analytically calculated compared to the voltage at the cathode of XENON1T during SR1. With the field strength as the function $E(z)$, the potential difference is given by

$$\Delta U = \int_{z_1=-96\,\mathrm{cm}}^{z_2=0\,\mathrm{cm}} E(z)\,\mathrm{d}z = \int_{z_1=-96\,\mathrm{cm}}^{z_2=0\,\mathrm{cm}} mz + b\,\mathrm{d}z = -\left(\frac{m}{2}z_1^2 + bz_1\right). \tag{7.1}$$

Since the gate is located at 0 cm, the upper integral border is dismissed in the last term of the equation. Inserting m and b together with the cathode position of $-96\,\mathrm{cm}$, the equation yields $\Delta U = 6.0\,\mathrm{kV}$. Since there are several factors contributing to the uncertainty of ΔU, its calculation is not trivial. Next to the uncertainty bars of the ratio fits, which are already considered by the minimisation, the uncertainty is affected by three additional factors. First, there is the general uncertainty of the NEST model itself. Since no reference was published for the estimation of general uncertainties on the NEST output, this contribution to the overall uncertainty cannot be quantified. The second contribution originates from the uncertainties of the parameters from the linear fit and can be directly calculated. Since they are correlated and both used in eq. (7.1) for the determination of ΔU, a correlated uncertainty propagation is necessary. For the calculation, the partial derivatives of ΔU with respect to m and b are required, as well as the error matrix ε. The overall uncertainty u is then given by

$$u(\Delta U) = \sqrt{\left(\frac{\partial(\Delta U)}{\partial m},\ \frac{\partial(\Delta U)}{\partial b}\right) \cdot \varepsilon \cdot \begin{pmatrix} \frac{\partial(\Delta U)}{\partial m} \\ \frac{\partial(\Delta U)}{\partial b} \end{pmatrix}}. \tag{7.2}$$

With iminuit, the error matrix for the two parameters is obtained. The partial derivates yield

$$\frac{\partial(\Delta U)}{\partial m} = -\frac{z_1}{2} \qquad \text{and} \qquad \frac{\partial(\Delta U)}{\partial b} = -z_1, \tag{7.3}$$

where $z_2 = 0\,\mathrm{cm}$ is already considered. Inserting the variables in eq. (7.2), a total uncertainty of $u(\Delta U) = 96\,\mathrm{V}$ is obtained from the fit. The third contribution originates from the choice of the time difference as a NEST input parameter. At the moment, the arithmetic mean of the distribution is chosen. Since the time difference has an asymmetric form, following an exponential decay after the initial

rise, the mean value does not coincide with the time value with the most entries. This value is at dt = 840 ns and is considered in the following as a comparison. Applying the same method as before, using the same fitted ratios with this alternative dt, the integrated potential difference is found to be at (5.7 ± 0.1) kV. Compared to (6.0 ± 0.1) kV with dt given by the arithmetic mean, the two values are slightly outside of each other's uncertainty range caused by the linear fit alone. This indicates that the choice of dt has a relevant uncertainty contribution, which is not included here. Compared to the actual potential difference between gate and cathode in SR1 of 8 kV [22], this result is clearly too small. Even with the additional consideration of dt as a source of uncertainties, such a large deviation cannot fully be explained.

Under the assumption of a trustworthy NEST field assignment, the discrepancy might be induced by the strong dependence of S1b/S1a on the data selection, as discussed earlier in this chapter. Since the assigned field strength and therefore the potential difference are strongly affected by slight changes of S1b/S1a, a large influence of the cut placement cannot be excluded. Another possible explanation is the accuracy of the NEST model itself, whose uncertainty was not considered due to the lack of information. Its parameters are fitted to measurements[1] of ^{83m}Kr decays, which are then extrapolated for the different field strengths. Since the measurements from various detectors are considered, it is possible that similar selection effects as observed in this chapter are incorporated into the model. However, with the current approach these two possible effects cannot be separated.

Leaving aside the absolute field strength, the information on the field inhomogeneity alone is still valid: at the top of the TPC, a larger field was observed than at the bottom. A possible usage could be a decoupling approach for the effects of the inhomogeneity on the calibration data and the light yield map, aiming for a "true" LY map, independent from the field effects. As previously discussed, the effect of a varying field strength depends on the deposited energy. Since the light yield map is used for corrections of data with all possible energy depositions, such a true LY map could enhance the detector's reconstruction performance. Another application could be observing the evolution of the field strength over time. With ^{83m}Kr datasets distributed over a complete SR period, time-dependent effects like indications of TPC wall charge-up behaviour [43] could be monitored. This is relevant for upcoming experiments, such as XENONnT.

[1] for example from the PIXeY detector [72]

8 Commissioning data of the XENONnT experiment

The following chapter exploits data from the commissioning phase of the XENONnT detector. Thus, all components of the experiment such as the detector data acquisition and data processing were not yet optimised. The data used in this work were acquired and analysed in order to tune the detector's operation parameters for a future science run. Thus, not all parameters were held constant during the calibration: the liquid level of the LXe between gate and anode was increased during the data taking from 5 mm to 5.5 mm above the gate. Further details on the detector's configuration are given in [76]. This chapter discusses the S1 correction derived from 88.7 h of ^{83m}Kr data that were taken during the commissioning calibration campaign at the end of January 2021.

Data selection

The data are stored in runs with half an hour duration. They are loaded with the previously used `event_info_double` plugin (see section 4.1). For computational reasons, the runs are loaded with the cuts already applied, using a pre-defined set of cuts [77]. Summarised, they are given by

- $80 \leq N_{\text{PMTs, S1a}} < 225$,
- $25 \leq N_{\text{PMTs, S1b}} < 125$,
- $0 \leq N_{\text{distinct PMTs}} < 60$,
- $750\,\text{ns} \leq \text{dt} < 2000\,\text{ns}$ and
- $2500\,\text{PE} < \text{S2a area} < 20\,000\,\text{PE}$.

The S1a and S1b parameter space of the resulting dataset is plotted in fig. 8.1. Its similarity to the loose selection in fig. 4.6 suggests that the cuts can still be improved, for example by a further selection in AFT_{S1a} or in the S2 width (as in [78]). After applying these cuts, the ^{83m}Kr dataset contains 318 620 events. This is an increase of about 90 % compared to the strict selection from the XENON1T ^{83m}Kr calibration data converted to straxen (see section 4.1).

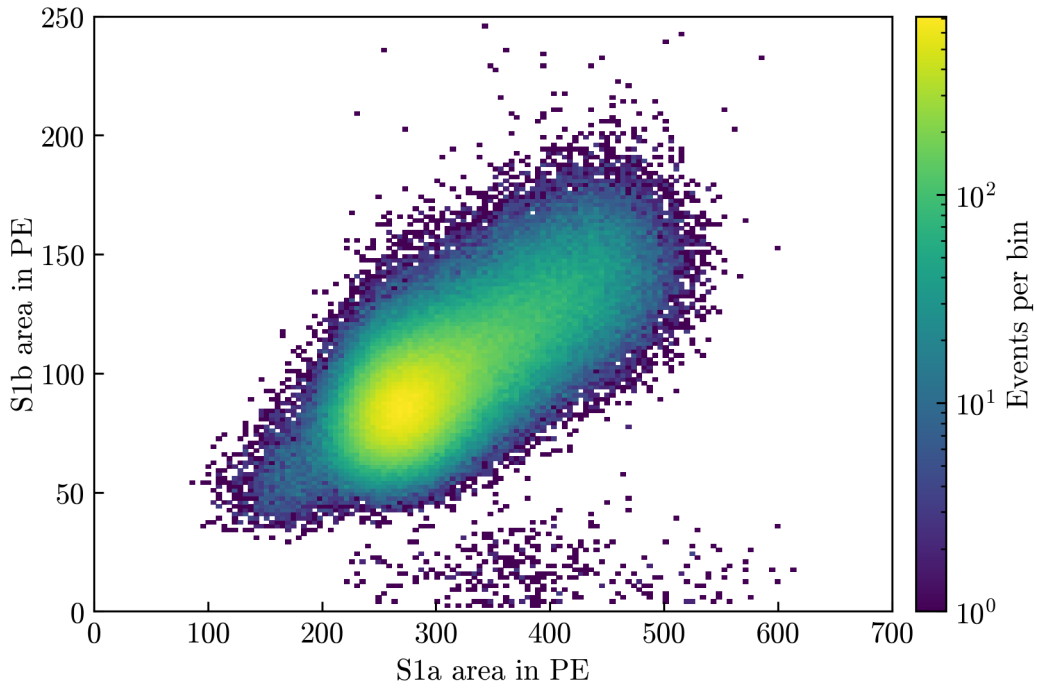

Figure 8.1: Two-dimensional histogram of the XENONnT ^{83m}Kr commissioning dataset from January 2021 in S1a area and S1b area. Cuts on $N_{\text{PMTs, S1a}}$, $N_{\text{PMTs, S1b}}$, $N_{\text{distinct PMTs}}$, dt and S2a area were applied.

Position reconstruction

Since the development and optimisation of configuration parameters of the analysis framework for XENONnT is still ongoing, not all required variables and corrections are implemented in straxen at the time of writing this thesis. Therefore, some parts of the data preparation and processing were performed manually using custom methods which are developed for future use within straxen after extensive testing. This includes the calculation of the z-, r- and φ-coordinates on which the S1 correction is calculated. As described in section 3.1, the z-coordinate is reconstructed based on the electron drift time. For this conversion, the constant drift velocity depending on the electric field is calculated, based on the detector's height and the maximally observed drift time. The XENONnT detector is operated with an electric field of approximately 18 V/cm [76] during this commissioning data acquisition, a lower field than for the XENON1T detector. The maximum drift time of 2381 ms corresponds to interactions directly at the cathode position and is marked in the left panel of fig. 8.2. With the TPC height of 148.5 cm, a drift

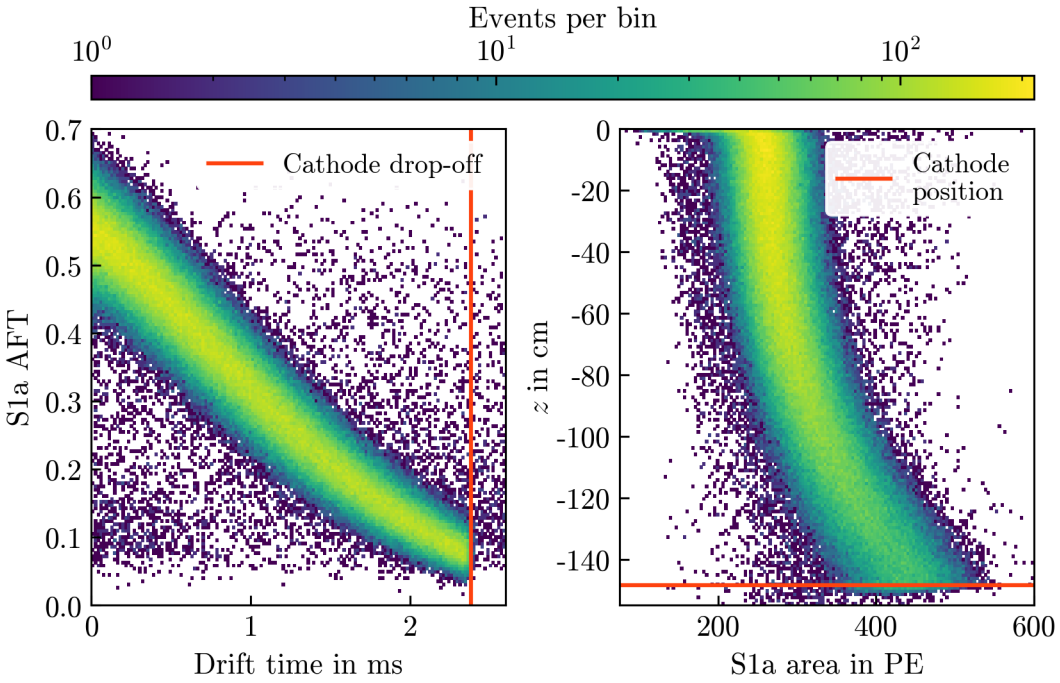

Figure 8.2: Two-dimensional histograms of the cut ^{83m}Kr dataset. Left: The AFT of the S1a signal and the drift time show an anti-correlated behaviour, since they are both coupled to the z-position. The maximum drift time is marked by the vertical red line as the cathode drop-off. Right: The reconstructed z-position is plotted against the S1a area, showing the effect to be corrected. The red line is plotted at the cathode position.

velocity of $v = 0.624\,\frac{\text{mm}}{\mu\text{s}}$ is derived.[1] The z-coordinate is obtained by multiplication of each event's individual drift time value with v, which is shown in the right panel of fig. 8.2. The r- and φ-coordinates are based on a position reconstruction method which was also used for x- and y-position of the XENON1T detector. At the moment, the actual x- and y-positions can be calculated by four different position reconstruction algorithms [79]. Three of them are based on neural network models and one uses a maximum likelihood method. The neural network models are a multilayer perceptron (MLP) model, which is similar to the position reconstruction used for XENON1T, a convolutional neural network (CNN) model and a graph convolutional network (GCN) model. Those models were trained and tested on simulated waveforms for the XENONnT detector. The differences between the algorithms on the positions do not affect the generation of the S1 correction map, since the estimated absolute difference is much smaller than the correction map's

[1] This approach does not yet consider the time the electrons spend between gate and anode.

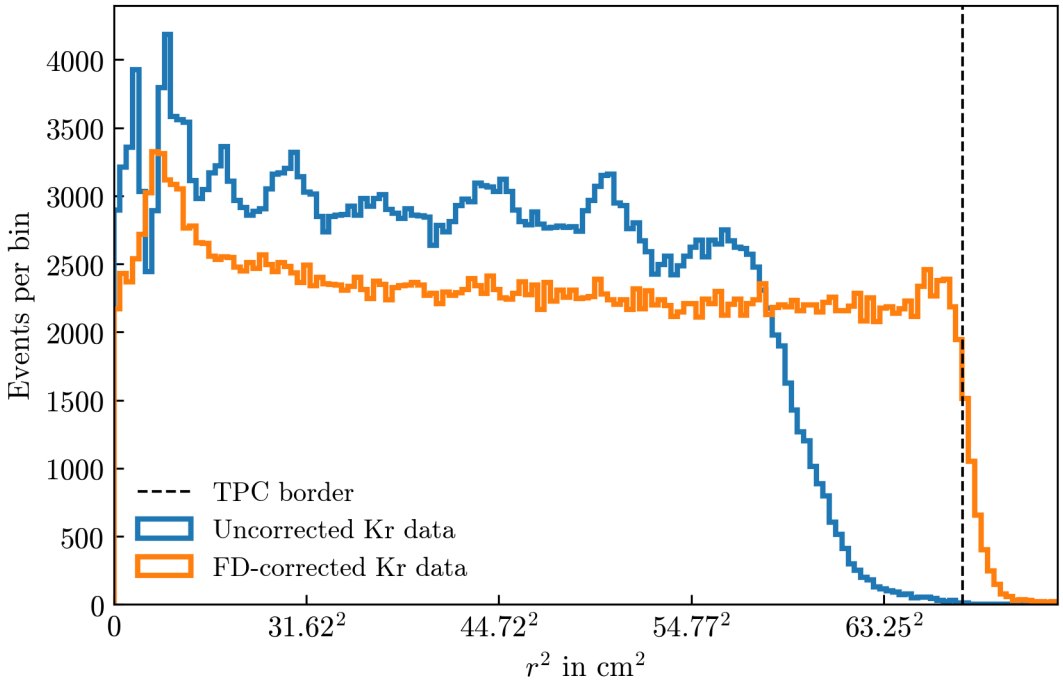

Figure 8.3: Histograms of the radial distribution before (blue) and after (orange) the application of the field distortion (FD) correction. The outer TPC border is marked by the vertical, dashed black line.

bin size. Thus, the fastest model (MLP) is chosen for this work. The MLP network consists of four dense neuron layers. The PMT channels are used as input connected to the two output parameters (x- and y-coordinate). Between those layers, a total of 57 842 parameters are adjusted during the training phase, such that the positions for the test data can be predicted accurately. In a next step, the x- and y-coordinates from the MLP model can be converted into r- and φ-coordinates.

Field distortion correction

A field distortion correction taken from [78] is applied, since corrected r-positions are crucial for a sufficient statistic in the outer TPC bins. The FDC correction was briefly introduced in section 4.1.1. The result of the position correction on the radial distribution is shown in fig. 8.3. Without the FDC, the binning of the S1 map is strongly limited by the low statistic at high radii. After the field-distortion correction, there remains a peak at low radii. This peak can be reduced by further cuts enhancing the data quality, as described at the data selection.

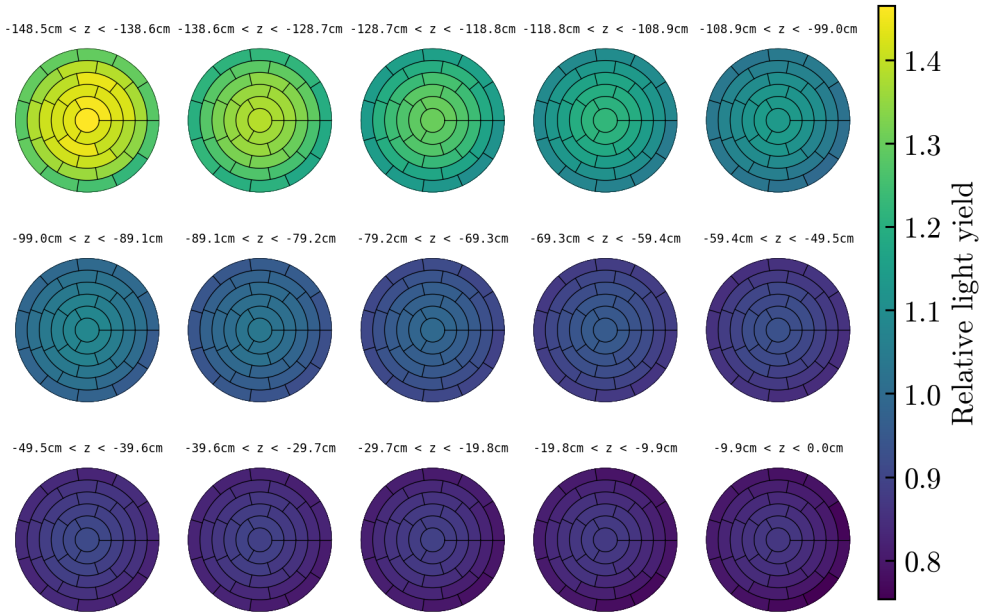

Figure 8.4: Relative light yield for XENONnT in each spatial TPC bin. The circles represent the slices in z. Due to the coarse binning and the detector commissioning data, the displayed map is at a preliminary stage. Further calibration campaigns under stable science run conditions with ^{83m}Kr in XENONnT will provide a higher statistic and thus allow a finer binning.

S1 map generation

After the FDC is applied, the data are prepared for the generation of a preliminary relative LY map. The approach is identical as discussed in section 3.4.1 and as already pursued for the XENON1T data in chapters 4 and 5. Therefore, the TPC is divided into bins. Since the XENONnT TPC is larger than its predecessor, the number of spatial bins needs to be increased in order to reach a similar absolute bin size. The higher number of events compared to the XENON1T dataset is used for an increase of slices in z from 8 in fig. 4.12 to 15 in fig. 8.4. This decreases the thickness of a single slice from 11.8 cm to 9.9 cm. Since the LY showed a stronger dependence on the z-position than on r or φ, the binning per slice is left unchanged, thus increasing the absolute width of a single annulus. The binning will be changed in future versions of the S1 correction map and optimised to the available statistic of the first ^{83m}Kr calibration data during commissioning. The fitting of the S1 peak in each bin is done with a maximum likelihood estimation (see section 4.2) and a one-dimensional Gaussian fit function (see eq. (4.11)). The resulting map is shown in fig. 8.4. Its properties are very similar to fig. 4.12.

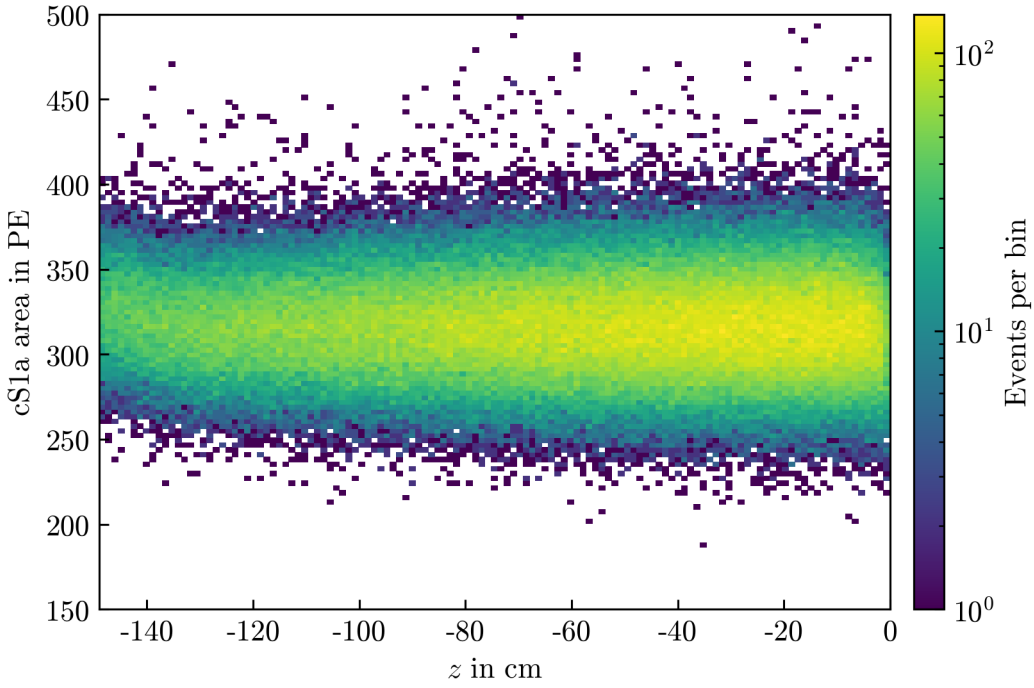

Figure 8.5: Two-dimensional histogram in cS1a and z of the first calibration data in the XENONnT detector with ^{83m}Kr. The correction is calculated with the relative light yield map in fig. 8.4.

With 1.47, the relative LY is highest in the middle of the bottom slice and with 0.75 lowest at the edge of the top slice. Close to the bottom of the TPC, there is a stronger radial variation of the light yield, while in the rest of the TPC, the variation is mainly observed between the slices. Also, the range of the relative light yield is similar to the result from XENON1T data. The S1 areas of the events after the application of the map are shown in fig. 8.5. On the lower end of the TPC, the cS1a area shows a deviation of its central value compared to other detector regions. This is caused by the the stronger variation of the S1a area at the bottom of the TPC compared to the top (see fig. 8.2) in combination with the coarse binning.

As mentioned above, this is a preliminary version of the S1 correction for a short ^{83m}Kr commissioning calibration campaign. During the first science run, the detector will be calibrated regularly with ^{83m}Kr. With the increased statistic and an additional selection in the S2 width and $\mathrm{AFT}_{\mathrm{S1a}}$, the analysis will be repeated with a finer binning in z, r and φ without a manual application of position reconstruction and corrections.

9 Conclusion and outlook

The XENON Dark Matter Project and its experiments XENON1T and the successor XENONnT are searching for WIMPs, employing a time projection chamber (TPC) filled with liquid xenon as a target material. XENONnT is equipped with the new data processor *streaming analysis for XENON* (straxen), which processes the light signal (S1) and the charge signal (S2) caused by an energy deposition in the detector. In this thesis, straxen is mainly used on reprocessed XENON1T data in order to prepare for upcoming XENONnT data.

The isotopes ^{83m}Kr and ^{37}Ar were used for calibration purposes. Both have the advantage that they can be mixed homogeneously within the liquid xenon target, while external sources do not reach the full active volume. After the calibration, both can be removed without leaving a remaining radioactivity in the TPC. With 2.8 keV, the argon decay has a lower energy deposition in the detector than ^{83m}Kr. The krypton source has a two-stepped decay with a 32.2 keV transition (S1a), followed by a 9.4 keV decay (S1b). The second decay has a lifetime of 154 ns, leading to a time difference in the order of a few 100 ns between both decays. Several signal corrections were performed on S1s and S2s. In general, the S1 needs to be corrected by a spatial signal size correction, compensating the varying light yield at different positions in the detector, caused by absorption and reflection of the light as well as different quantum efficiencies of the PMTs. The electrons generating the S2 are influenced by electronegative impurities in the xenon target. The resulting electron lifetime needs to be determined and corrected for.

The focus in this work was on the spatial S1 area correction, which was first derived from the ^{83m}Kr calibration data. The preparation of a ^{83m}Kr dataset showed that the correct reconstruction and selection of events with a time difference between S1a and S1b lower than $\sim$500 ns is not possible in most cases. Since a ^{83m}Kr S1 has a width of $\sim$200 ns, an improvement of the reconstruction should be possible. Thus, a further investigation could enhance the statistics available from the ^{83m}Kr source with the same amount of data. With the selected data, the spatially dependent relative S1 area correction was calculated and applied, assuming a Gaussian S1 signal size distribution of the events. The obtained correction

factor is between 0.72 and 1.37. Since the achievable spatial resolution of this correction depends on the available statistics, it would benefit from an improved ^{83m}Kr reconstruction.

Apart from this, an electron lifetime for both calibration sources was derived. For the ^{83m}Kr calibration data recorded in February 2017, the lifetime was at $(543.2 \pm 0.7)\,\mu s$, while the ^{37}Ar data which were recorded during October 2018 yielded $(1.13 \pm 0.01)\,ms$. Approximately 50 % of this discrepancy can be explained by the general lifetime difference between the two science runs of XENON1T. An additional source-dependent electron lifetime was observed by previous analyses. The most likely explanation for this is an inhomogeneous electric field causing an apparent lifetime difference between calibration sources. The electron lifetime was used as a correction factor for the S2 size depending on the z-position of an event.

A further emphasis was set on the distribution of the S1 peak size and the modelling of the low-energetic ^{37}Ar S1 signal, which showed a deformation of the Gaussian shape seen for the ^{83m}Kr S1. Therefore, two approaches were introduced: a phenomenological model using an additional threshold function and a model based on PMT gain calibration. With a simulation of photons reaching the individual PMTs, the hypothesis of the threshold behaviour being solely caused by geometric detector effects was rejected, favouring the latter description.

Further investigating this deformation of the S1 area distribution towards smaller areas, the noble element scintillation technique (NEST) software was introduced. NEST can predict electron and photon yields in liquid xenon for a given deposited energy and electric field value. Those yields were used for a prediction of the generated number of electrons and photons, employing a skew-Gaussian recombination model described by the skewness parameter α. This recombination model implemented in NEST was successfully employed in order to describe the deformation of the S1 ^{37}Ar data with $\alpha = 1.89 \pm 0.02$. Here, a contradiction between the application to the data and the NEST simulation was observed, since the simulation yielded a negative skewness of $\alpha = -1.90 \pm 0.01$. As a further investigation, a combined model of a detection threshold and the skewness model could be considered in future studies. However, also a comparison with ^{37}Ar S2 signals (simulated: $\alpha = 2.28 + -0.01$, observed: $\alpha = 1.12 \pm 0.02$) and the ^{83m}Kr S1 data (simulated: $\alpha = -2.15$, observed: $\alpha = 0.91 \pm 0.28$) showed that the parameter prediction obtained by a usage of NEST does not agree with the data.

In all previous chapters a homogeneous electric field was assumed. The homo-

geneity of the drift field was investigated with a method based on the ratio of the two S1s from ^{83m}Kr decays. The S1a is more sensitive to a varying electric field strength than the S1b. Thus, the ratio of the two signal sizes depends on the local field strength, while the geometric detector effects are cancelled. This allowed a qualitative observation of a varying field and resulted in the observation that the field strength increased towards the top of the TPC. With the NEST model, it was possible to obtain a prediction on the S1b/S1a ratio from the light yields at different field strengths. In this way, NEST was employed for an assignment of absolute field values to the measured S1b/S1a ratios. A minimum value of $(35.9 \pm 6.0)\,\mathrm{V/cm}$ and a maximum of $(90.1 \pm 8.9)\,\mathrm{V/cm}$ was found, which includes the $(81 \pm 6)\,\mathrm{V/cm}$ specified for the fiducial volume of XENON1T during the first science run. An integration of the electric field over the TPC led to an estimated total potential difference of $(6.0 \pm 0.1)\,\mathrm{kV}$ between cathode and gate, compared to an actual value of 8 kV. Those two values show a discrepancy not covered by the uncertainty interval. A possible explanation is the observed dependency of the S1b/S1a ratio on the data selection. Moreover, the NEST model could not be suitable for the description of such small field differences. However, these two possible effects cannot be separated with this method. Underestimated uncertainties could also lead to the discrepancy. One uncertainty factor is the choice of the time difference dt between S1a and S1b for the field assignment, since the enhancement of S1b especially for $\mathrm{dt} \lesssim 500\,\mathrm{ns}$ is considered by the NEST model. In a further investigation, this could be avoided by the choice of a more direct approach, where instead of employing an average time difference each event is evaluated separately. The evaluation of position dependent field variations would then be done directly with the individual events. This could also allow an easy variation of the considered spatial regions of the TPC, for example for the observation of radial effects. A further step could be the observation of the field strength over time. This could either be done with additional ^{83m}Kr runs from SR1 of XENON1T, or it could be planned for the upcoming calibration campaigns of XENONnT.

The last chapter utilises XENONnT commissioning data for a first application of straxen to the new data. The calculation and execution of the S1 correction demonstrates the successful preparation for the new experiment. The S1 correction for ^{83m}Kr data derived in this thesis will be implemented as a part of straxen's correction management system. Based on a larger dataset from further calibration campaigns, the binning will become finer with the increasing statistic as the

first science run is evaluated. Thus, the knowledge of the response to calibration data will be enhanced. The upcoming science data taking of XENONnT will be supported with the calculated S1 correction. The influence by the considered field inhomogeneities for different energies should be investigated further and reduced in the future.

A Additional plots

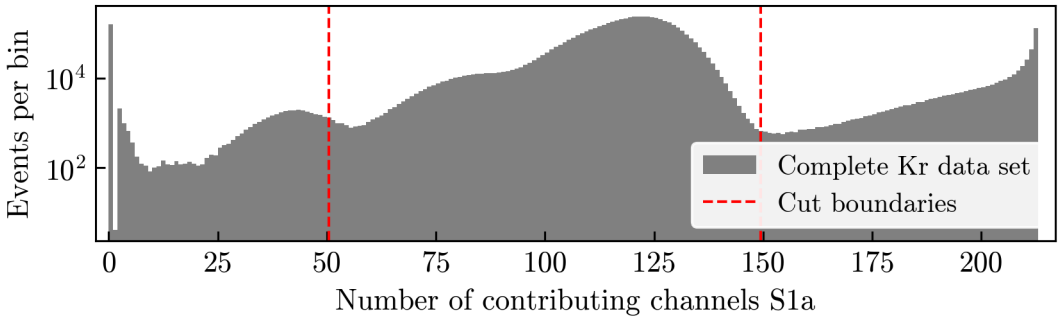

Figure A.1: The complete krypton dataset is plotted in a histogram in the number of channels contributing to S1a. The red lines mark the borders of the cut.

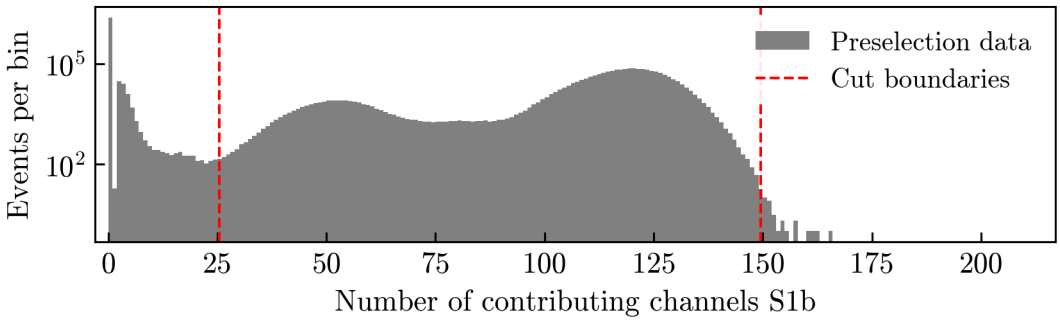

Figure A.2: The krypton preselection is plotted in a histogram in the number of channels contributing to S1b. The red lines mark the borders of the cut.

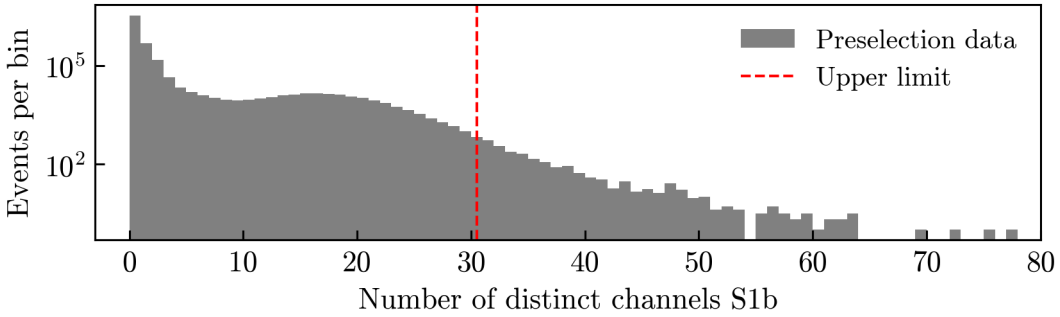

Figure A.3: The krypton preselection is plotted in a histogram in the number of distinct channels of S1b. The red line marks the upper limit.

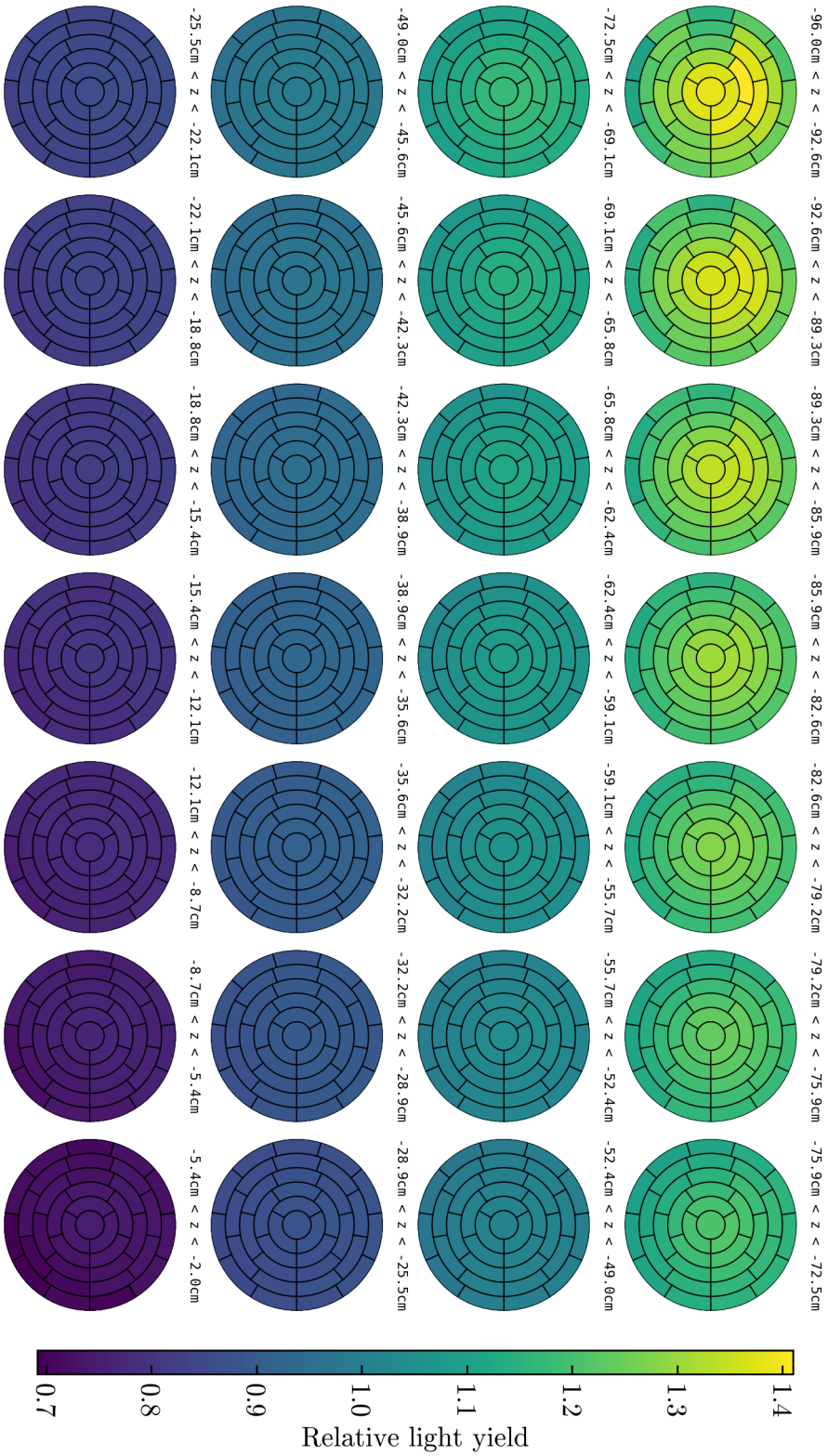

Figure A.4: Relative LY map based on the loose ^{83m}Kr selection. The higher statistic is utilised for a higher number of z slices and annuli.

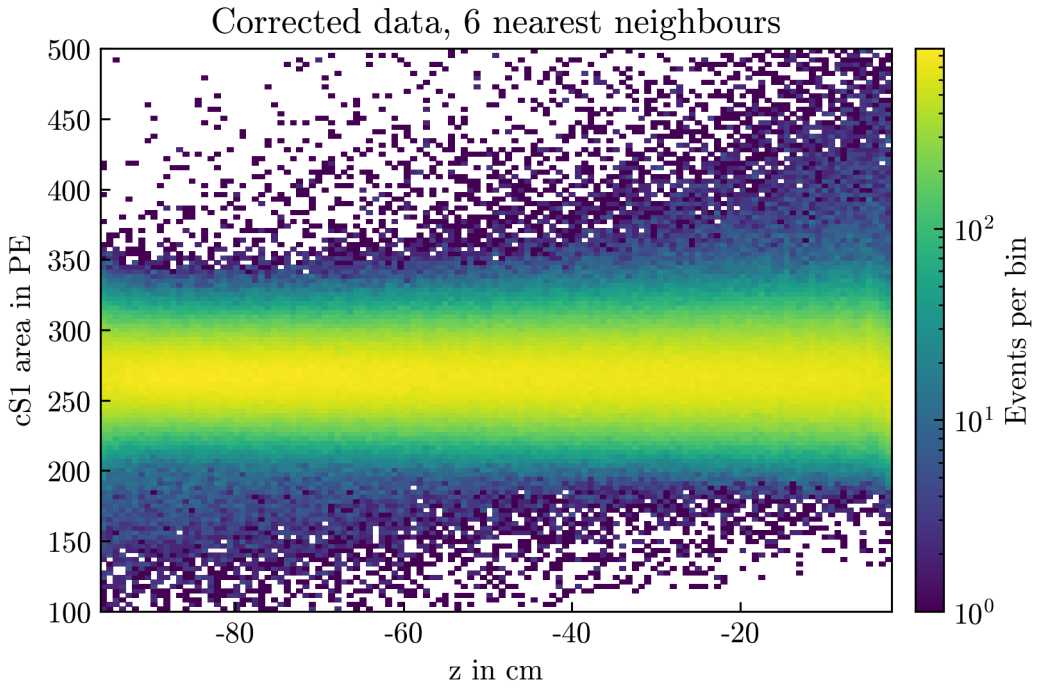

Figure A.5: Two-dimensional histogram of the loose ^{83m}Kr selection in z and corrected S1 area. The correction is done with fig. A.4 and with the default value of 6 nearest neighbours for the interpolation.

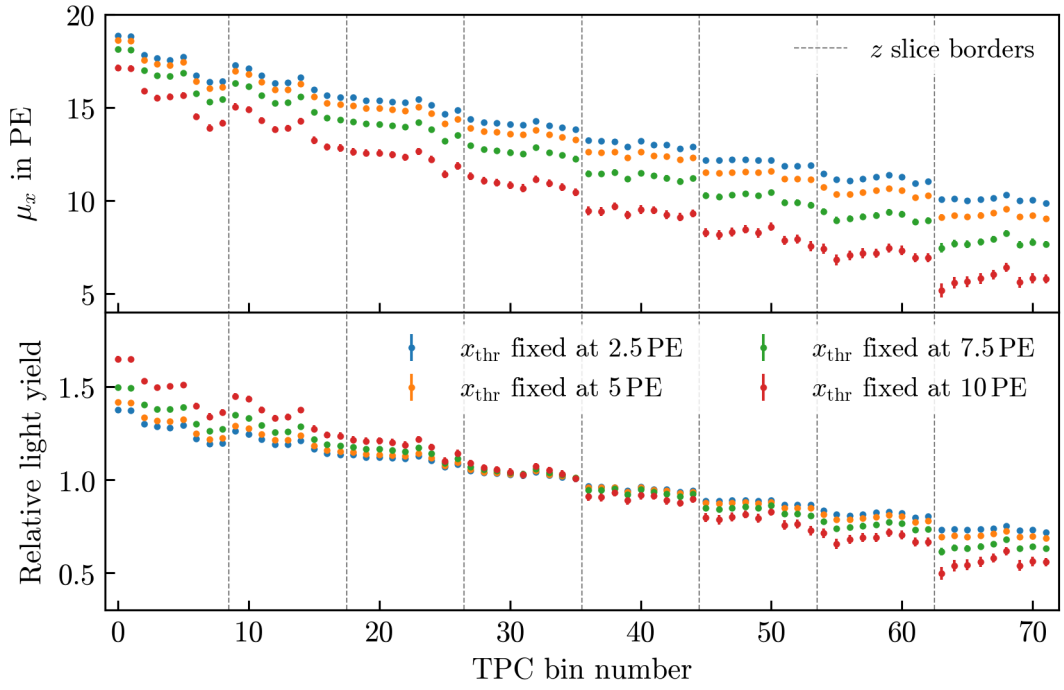

Figure A.6: Effect of different fixed thresholds in the fit of the ^{37}Ar data with eq. (5.1) on the mean value μ_x (upper panel) and the relative light yield (lower panel). The width of the threshold is in all cases fixed to $w_{\mathrm{thr}} = 3\,\mathrm{PE}$.

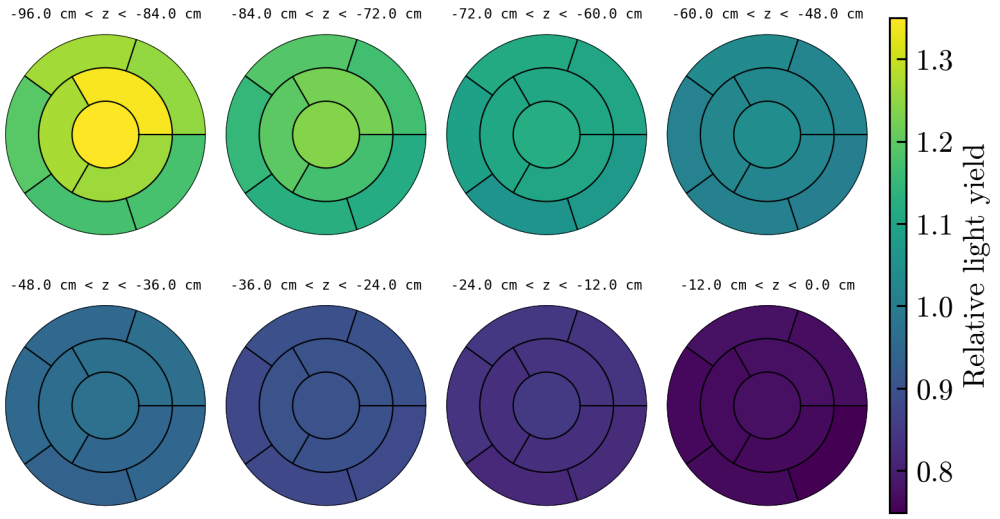

Figure A.7: Colour map of the relative LY as a result of the skew-Gaussian fit to the ^{37}Ar data. Each circle represents a z slice of the TPC.

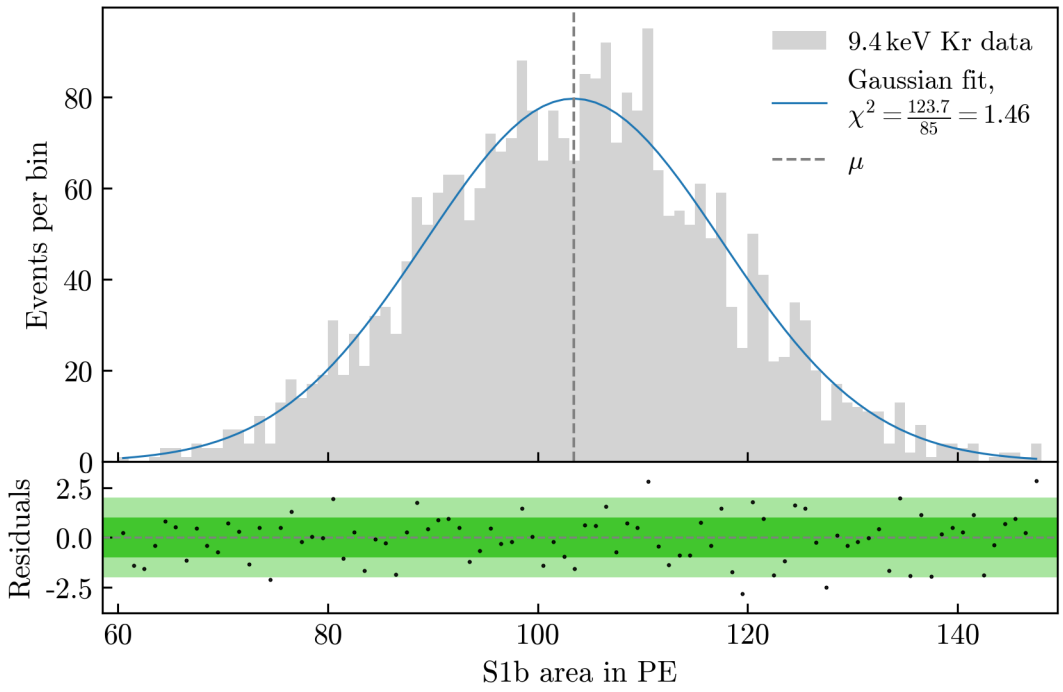

Figure A.8: Histogram of the S1b area in an example bin. A Gaussian fit (blue line) is done with a Poisson likelihood function. The residuals are plotted in the lower panel. The resulting mean value is marked with a vertical dashed line. The fit has a reconstructed reduced χ^2-value of 1.46 at 85 degrees of freedom.

B Bibliography

[1] H. Poincaré: *The milky way and the theory of gases.* Popular Astronomy, 14:475–488, 1906.

[2] J.C. Kapteyn: *First attempt at a theory of the arrangement and motion of the sidereal system.* The Astrophysical Journal, 55:302, 1922.

[3] J.H. Oort: *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems.* Bulletin of the Astronomical Institutes of the Netherlands, 6:249, 1932.

[4] F. Zwicky: *Die Rotverschiebung von extragalaktischen Nebeln.* Helvetica physica acta, 6:110–127, 1933.

[5] M.S. Roberts and A.H. Rots: *Comparison of rotation curves of different galaxy types.* Astronomy and Astrophysics, 26:483–485, 1973.

[6] V. Rubin, W.K. Ford Jr, and N. Thonnard: *Extended rotation curves of high-luminosity spiral galaxies. IV-Systematic dynamical properties, SA through SC.* The Astrophysical Journal, 225:L107–L111, 1978.

[7] T.S. van Albada, J.N. Bahcall, K. Begeman, and R. Sancisi: *Distribution of dark matter in the spiral galaxy NGC 3198.* The Astrophysical Journal, 295:305–313, 1985.

[8] A. Penzias and R.W. Wilson: *A measurement of excess antenna temperature at 4080 Mc/s.* The Astrophysical Journal, 142:419–421, 1965.

[9] G. Hinshaw, D. Larson, E. Komatsu, D.N. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, *et al.*: *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results.* The Astrophysical Journal Supplement Series, 208(2):19, 2013.

[10] Planck Collaboration: *Planck 2015 results – XIII. Cosmological parameters.* Astronomy & Astrophysics, 594:A13, 2016.

[11] A. Einstein: *Lens-like action of a star by the deviation of light in the gravitational field.* Science, 84(2188):506–507, 1936.

[12] F. Zwicky: *Nebulae as gravitational lenses.* Phys. Rev., 51:290–290, Feb 1937. `https://link.aps.org/doi/10.1103/PhysRev.51.290`.

[13] R. Massey, T. Kitching, and J. Richard: *The dark matter of gravitational lensing.* Reports on Progress in Physics, 73(8):086901, 2010.

[14] D. Clowe, M. Bradač, A.H. Gonzalez, M. Markevitch, S.W. Randall, C. Jones, and D. Zaritsky: *A direct empirical proof of the existence of dark matter.* The Astrophysical Journal Letters, 648(2):L109, 2006.

[15] X-ray: NASA/CXC/CfA/M.Markevitch et al.; Optical: NASA/STScI; Magellan/U.Arizona/D.Clowe et al.; Lensing Map: NASA/STScI; ESO WFI; Magellan/U.Arizona/D.Clowe et al. `https://chandra.harvard.edu/photo/2006/1e0657/more.html`. [accessed 15-March-2021].

[16] M. Schumann: *Direct detection of WIMP dark matter: concepts and status.* Journal of Physics G: Nuclear and Particle Physics, 46(10):103003, 2019.

[17] L. Covi, H.B. Kim, J.E. Kim, and L. Roszkowski: *Axinos as dark matter.* Journal of High Energy Physics, 2001(05):033, 2001.

[18] S. Dodelson and L. M. Widrow: *Sterile neutrinos as dark matter.* Physical Review Letters, 72(1):17, 1994.

[19] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrinin, M. Pierre, S. Profumo, and F.S. Queiroz: *The waning of the WIMP? A review of models, searches, and constraints.* The European Physical Journal C, 78(3), Mar 2018, ISSN 1434-6052. `http://dx.doi.org/10.1140/epjc/s10052-018-5662-y`.

[20] T. Marrodán Undagoitia and L. Rauch: *Dark matter direct-detection experiments.* Journal of Physics G: Nuclear and Particle Physics, 43(1):013001, Dec 2015. `https://doi.org/10.1088/0954-3899/43/1/013001`.

[21] J.D. Lewin and P.F. Smith: *Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil.* Astroparticle Physics, 6(1):87–112, 1996, ISSN 0927-6505. `https://www.sciencedirect.com/science/article/pii/S0927650596000473`.

[22] E. Aprile and the XENON Collaboration: *Dark matter search results from a one ton-year exposure of XENON1T.* Phys. Rev. Lett., 121:111302, Sep 2018. `https://link.aps.org/doi/10.1103/PhysRevLett.121.111302`.

[23] E. Aprile and the XENON Collaboration: *The XENON1T dark matter experiment.* Eur. Phys. J. C, 77(12):881, 2017. `https://doi.org/10.1140/epjc/s10052-017-5326-3`.

[24] E. Aprile and T. Doke: *Liquid xenon detectors for particle physics and astrophysics.* Rev. Mod. Phys., 82:2053–2097, Jul 2010. `https://link.aps.org/doi/10.1103/RevModPhys.82.2053`.

[25] A. Bideau-Mehu, Y. Guern, R. Abjean, and A. Johannin-Gilles: *Measurement of refractive indices of neon, argon, krypton and xenon in the 253.7–140.4 nm wavelength range. Dispersion relations and estimated oscillator strengths of the resonance lines.* Journal of

Quantitative Spectroscopy and Radiative Transfer, 25(5):395–402, 1981, ISSN 0022-4073. https://www.sciencedirect.com/science/article/pii/0022407381900571.

[26] V.N Solovov, V. Chepel, M.I Lopes, A. Hitachi, R. Ferreira Marques, and A.J.P.L Policarpo: *Measurement of the refractive index and attenuation length of liquid xenon for its scintillation light.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 516(2):462–474, 2004, ISSN 0168-9002. https://www.sciencedirect.com/science/article/pii/S0168900203024331.

[27] E. Aprile and the XENON Collaboration: *First dark matter search results from the XENON1T experiment.* Phys. Rev. Lett., 119:181301, Oct 2017. https://link.aps.org/doi/10.1103/PhysRevLett.119.181301.

[28] L. Althüser for the XENON Collaboration. https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:althueser, Nov 2020. [accessed 25-January-2021].

[29] E. Aprile and the XENON Collaboration: *Energy resolution and linearity of XENON1T in the MeV energy range.* The European Physical Journal C, 80, Aug 2020.

[30] E. Aprile and the XENON Collaboration: *Projected WIMP sensitivity of the XENONnT dark matter experiment.* Journal of Cosmology and Astroparticle Physics, 2020(11):031–031, Nov 2020. https://doi.org/10.1088/1475-7516/2020/11/031.

[31] E. Aprile and the XENON Collaboration: *Lowering the radioactivity of the photomultiplier tubes for the XENON1T dark matter experiment.* The European Physical Journal C, 75(11), Nov 2015, ISSN 1434-6052. http://dx.doi.org/10.1140/epjc/s10052-015-3657-5.

[32] E. Aprile and the XENON Collaboration: *Material screening and selection for XENON100.* Astroparticle Physics, 35(2):43–49, 2011, ISSN 0927-6505. https://www.sciencedirect.com/science/article/pii/S0927650511000971.

[33] E. Aprile and the XENON Collaboration: *Conceptual design and simulation of a water cherenkov muon veto for the XENON1T experiment.* Journal of Instrumentation, 9(11):P11006–P11006, Nov 2014. https://doi.org/10.1088/1748-0221/9/11/p11006.

[34] M. Murra: *Intrinsic background reduction by cryogenic distillation for the XENON1T dark matter experiment.* PhD thesis, Westfälische Wilhelms-Universität Münster, 2018.

[35] J. Aalbers and other: *Streaming analysis for XENON.* https://github.com/XENONnT/straxen, 2021.

[36] Aalbers, J. and other: *Axfoundation/strax: v0.13.7*, Jan 2021. https://doi.org/10.5281/zenodo.4479898.

[37] C.R. Harris and other: *Array programming with NumPy.* Nature, 585(7825):357–362, Sep 2020. https://doi.org/10.1038/s41586-020-2649-2.

[38] *Straxen datastructure.* https://straxen.readthedocs.io/en/latest/reference/datastructure.html#events, 2020. [accessed 29-December-2020].

[39] J. Aalbers: *Jenks/Fischer Natural breaks splitting.* https://github.com/AxFoundation/strax/pull/225, Dec 2019.

[40] J. Long: *Single electron pre-study under straxen framework.* https://xe1t-wiki.lngs.infn.it/doku.php?id=jlong:single_electron_pre-study_under_straxen_framework, Mar 2020. [accessed 19-March-2021].

[41] E. Aprile and the XENON Collaboration: *XENON1T dark matter data analysis: Signal reconstruction, calibration, and event selection.* Phys. Rev. D, 100:052014, Sep 2019. https://link.aps.org/doi/10.1103/PhysRevD.100.052014.

[42] *Straxen package: InterpolatingMap.* https://github.com/XENONnT/straxen/blob/master/straxen/itp_map.py, 2021. [accessed 17-February-2021].

[43] M. Vargas: *Data analysis in the XENON1T dark matter experiment.* PhD thesis, Westfälische Wilhelms-Universität Münster, 2019.

[44] S. Rosendahl and other: *Determination of the separation efficiencies of a single-stage cryogenic distillation setup to remove krypton out of xenon by using a ^{83m}Kr tracer method.* Review of Scientific Instruments, 86:115104, Nov 2015.

[45] Brookhaven National Laboratory: *Nuclear chart.* https://www.nndc.bnl.gov/nudat2/. [accessed 19-March-2021].

[46] V. Hannen, E. Aprile, F. Arneodo, L. Baudis, M. Beck, K. Bokeloh, A. D. Ferella, K. Giboni, R. F. Lang, O. Lebeda, H W. Ortjohann, M. Schumann, A. Spalek, D. Venos, and C. Weinheimer: *Limits on the release of Rb isotopes from a zeolite based ^{83m}Kr calibration source for the XENON project.* Journal of Instrumentation, 6(10):P10013–P10013, Oct 2011. https://doi.org/10.1088/1748-0221/6/10/p10013.

[47] V. Barsanov and other: *Artificial neutrino source based on the ^{37}Ar isotope.* Physics of Atomic Nuclei, 70:300–310, Feb 2007.

[48] M. Murra: *Ar-37 removal model and fit.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:murra:ar37_evolution_model_fit, Dec 2018. [accessed 24-March-2021].

[49] M. Pierre: *Preliminary study for Kr83m events selection with strax.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:maximepierre:double_s1_scatter_strax, Dec 2019. [accessed 20-March-2021].

[50] P. Barrow, L. Baudis, D. Cichon, M. Danisch, D. Franco, F. Kaether, A. Kish, M. Lindner, T. Marrodán Undagoitia, D. Mayani, L. Rauch, Y. Wei, and J. Wulf: *Qualification tests of the R11410-21 photomultiplier tubes for the XENON1T detector.* Journal of Instrumentation, 12(01):P01024, Jan 2017. https://doi.org/10.1088/1748-0221/12/01/p01024.

[51] L. Baudis, A. Behrens, A. Ferella, A. Kish, T. Marrodán Undagoitia, D. Mayani, and M. Schumann: *Performance of the Hamamatsu R11410 photomultiplier tube in cryogenic xenon environments.* Journal of Instrumentation, 8(04):P04026, Apr 2013. https://doi.org/10.1088/1748-0221/8/04/p04026.

[52] M. Pierre: *KrDoubleS1 population event selection.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:mpierre:krdoubles1_cuts_definition_time_ordering, Apr 2020. [accessed 20-March-2021].

[53] H. Dembinski, P. Ongmongkolkul, and other: *scikit-hep/iminuit.* Nov 2020. https://doi.org/10.5281/zenodo.4283509.

[54] S. Baker and R.D. Cousins: *Clarification of the use of chi-square and likelihood functions in fits to histograms.* Nuclear Instruments and Methods in Physics Research, 221(2):437–442, 1984, ISSN 0167-5087. https://www.sciencedirect.com/science/article/pii/0167508784900164.

[55] J. Wulf, S. Kazama, and E. Shockley: *PMT status post SR1.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenon1t:pmt:pmts_status_post_sr1, Aug 2019. [accessed 03-March-2021].

[56] C. Hills: *Spatial response of the TPC at 2.8 keV.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenon1t:ar37calibration:analysis:2019:analysis_status, Jul 2019. [accessed 26-March-2021].

[57] E.M. Boulton, E. Bernard, N. Destefano, B.N.V. Edwards, M. Gai, S.A. Hertel, M. Horn, N.A. Larsen, B.P. Tennyson, C. Wahl, and D.N. McKinsey: *Calibration of a two-phase xenon time projection chamber with a ^{37}Ar source.* Journal of Instrumentation, 12(08):P08004–P08004, aug 2017. https://doi.org/10.1088/1748-0221/12/08/p08004.

[58] L. Althüser: *Light collection efficiency simulations of the XENON1T experiment and comparison to data.* Master's thesis, Westfälische Wilhelms-Universität Münster, Nov 2017.

[59] C. Hills, E. Angelino, A. Molinario, M. Alfonsi, and E. Shockley: *Anomalous behaviour of ^{37}Ar data on top of the TPC.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenon1t:dsg:muonveto:torino:s2only_cutssr2:ardatamc:discrepancy_top_tpc_note, October 2020. [accessed 15-February-2021].

[60] E. Brown, A. Buss, A. Fieguth, C. Huhmann, M. Murra, H. W. Ortjohann, S. Rosendahl, A. Schubert, D. Schulte, D. Tosi, G. Gratta, and C. Weinheimer: *Magnetically-coupled piston pump for high-purity gas applications.* The European Physical Journal C, 78(7), Jul 2018, ISSN 1434-6052. http://dx.doi.org/10.1140/epjc/s10052-018-6062-z.

[61] J. Palacio, S. Diglio, and C. Therreau: *Electron lifetime from SR1 and post-SR1.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenon1t:suba:electronlifetime_sr1_postsr1, Nov 2018. [accessed 22-March-2021].

[62] J. Brodsky, C. Tunnell, and other: *NESTCollaboration/nest*, Jan 2019. https://doi.org/10.5281/zenodo.1314499.

[63] M. Szydagis, N. Barry, K. Kazkaz, J. Mock, D. Stolp, M. Sweany, M. Tripathi, S. Uvarov, N. Walsh, and M. Woods: *NEST: A comprehensive model for scintillation yield in liquid xenon.* Journal of Instrumentation, 6(10):P10002–P10002, Oct 2011. https://doi.org/10.1088/1748-0221/6/10/p10002.

[64] T. Takahashi, S. Konno, T. Hamada, M. Miyajima, S. Kubota, A. Nakamoto, A. Hitachi, E. Shibamura, and T. Doke: *Average energy expended per ion pair in liquid xenon.* Phys. Rev. A, 12:1771–1775, Nov 1975. https://link.aps.org/doi/10.1103/PhysRevA.12.1771.

[65] T. Doke and other: *Absolute scintillation yields in liquid argon and xenon for various particles.* Japanese Journal of Applied Physics, 41(Part 1, No. 3A):1538–1545, Mar 2002. https://doi.org/10.1143/jjap.41.1538.

[66] J. Thomas and A.D. Imel: *Recombination of electron-ion pairs in liquid argon and liquid xenon.* Phys. Rev. A, 36:614–616, Jul 1987. https://link.aps.org/doi/10.1103/PhysRevA.36.614.

[67] T. Doke and other: *LET dependence of scintillation yields in liquid argon.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 269(1):291–296, 1988, ISSN 0168-9002. https://www.sciencedirect.com/science/article/pii/0168900288908923.

[68] J.B. Birks, year=1964: *The theory and practice of scintillation counting.* Pergamon Press.

[69] D.S. Akerib and the LUX Collaboration: *Discrimination of electronic recoils from nuclear recoils in two-phase xenon time projection chambers.* Phys. Rev. D, 102:112002, Dec 2020. https://link.aps.org/doi/10.1103/PhysRevD.102.112002.

[70] M. Szydagis, C. Levy, G.M. Blockinger, A. Kamaha, N. Parveen, and G.R.C. Rischbieter: *Investigating the XENON1T low-energy electronic recoil excess using NEST.* Phys. Rev. D, 103:012002, Jan 2021. https://link.aps.org/doi/10.1103/PhysRevD.103.012002.

[71] D.S. Akerib and the LUX Collaboration: *^{83m}Kr calibration of the 2013 LUX dark matter search.* Phys. Rev. D, 96:112009, Dec 2017. https://link.aps.org/doi/10.1103/PhysRevD.96.112009.

[72] A.G. Singh, E.P. Bernard, A. Biekert, E.M. Boulton, S.B. Cahn, N. Destefano, B.N.V. Edwards, M. Gai, M. Horn, N. Larsen, Q. Riffard, B. Tennyson, V. Velan, C. Wahl, and D.N. McKinsey: *Analysis of ^{83m}Kr prompt scintillation signals in the PIXeY detector.* Journal of Instrumentation, 15(01):P01023–P01023, Jan 2020. https://doi.org/10.1088/1748-0221/15/01/p01023.

[73] M. Szydagis, A. Fyhrie, D. Thorngren, and M. Tripathi: *Enhancement of NEST capabilities for simulating low-energy recoils in liquid xenon.* Journal of Instrumentation, 8(10):C10003–C10003, Oct 2013. https://doi.org/10.1088/1748-0221/8/10/c10003.

[74] E. Aprile, R. Budnik, B. Choi, H.A. Contreras, K. L. Giboni, L.W. Goetzke, J.E. Koglin, R.F. Lang, K.E. Lim, A.J. Melgarejo Fernandez, R. Persiani, G. Plante, and A. Rizzo: *Measurement of the scintillation yield of low-energy electrons in liquid xenon.* Phys. Rev. D, 86:112004, Dec 2012. https://link.aps.org/doi/10.1103/PhysRevD.86.112004.

[75] L. Baudis, H. Dujmovic, C. Geis, A. James, A. Kish, A. Manalaysay, T. Marrodán Undagoitia, and M. Schumann: *Response of liquid xenon to compton electrons down to 1.5 keV.* Phys. Rev. D, 87:115015, Jun 2013. https://link.aps.org/doi/10.1103/PhysRevD.87.115015.

[76] J. Ye: *Brief summary of XENONnT commissioning data in S2 era.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:data_taking_jan_commissioning, Feb 2021.

[77] M. Pierre: *^{83m}Kr selection at the event level during XENONnT commissioning phase - summary note.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:mpierre:kr83m_events_selection_xenonnt_commissioning_phase, Feb 2021.

[78] F. Toschi: *Field distortion correction (FDC) during commissioning.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:analysis:reconstruction_team:fdc_commissioning, Feb 2021.

[79] S. Lian and A. Terliuk: *Current status of S2 position reconstruction.* https://xe1t-wiki.lngs.infn.it/doku.php?id=xenon:xenonnt:analysis:reconstruction_team:posrec_summary, Nov 2020.

Declaration of Academic Integrity

I hereby confirm that this thesis on **Signal calibration for XENON1T with the new data reconstruction framework straxen** is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

Münster, March 31, 2021

Johanna Jakob

I agree to have my thesis checked in order to rule out potential similarities with other works and to have my thesis stored in a database for this purpose.

Münster, March 31, 2021

Johanna Jakob

Danksagung

An dieser Stelle möchte ich mich bei den Menschen bedanken, die mich bei der Entstehung dieser Arbeit unterstützt haben oder sie überhaupt erst möglich gemacht haben. Zunächst gilt mein Dank für die Betreuung dieser Arbeit Prof. Weinheimer, ohne dessen Ermunterung ich seine großartige Arbeitsgruppe nicht kennengelernt hätte.

Aus der Arbeitsgruppe möchte ich mich vor allem bei Christian bedanken. Als mein direkter Ansprechpartner hat er mir viele Dinge erklärt und gezeigt und auch inhaltlich sehr hilfreiche Ideen beigesteuert. Neben Christian, Michael, Kevin und Lutz möchte ich für die Korrektur der Arbeit vor allem Henning danken, der das Korrekturlesen in ihrer ersten (und vermutlich anstrengendsten) Form übernommen hat. Weiterhin bin ich Lutz überaus dankbar für die schnelle Erreichbarkeit zu nahezu jeder Tages- und Nachtzeit sowie für die unkomplizierte, fähige Hilfe bei vielen kleinen Problemchen und größeren technischen Fragen.

Dann, natürlich, geht ein großes Dankeschön an Lukas für die Geduld mit mir in den letzten Monaten und die liebevolle Unterstützung. Ein schönes Beispiel ist das Übernehmen des Kuchenbackens oder des ein oder anderen Lebensmitteleinkaufs.

Abschließend möchte ich mich an dieser Stelle bei meinen Eltern bedanken. Durch ihre volle Unterstützung während des gesamten Studiums haben sie mir eine schöne und sorgenfreie Studienzeit ermöglicht. Vielen Dank!