TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

ModeLimit - A Python Package for Computing Exclusion Limits Using Data with Unknown Background

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

im Rahmen des Studiums

Technische Physik

eingereicht von

Fatih Okçu, BSc
Matrikelnummer 01425617

Ausgeführt am Atominstitut
der Fakultät für Physik der Technischen Universität Wien

Betreuung
Betreuer/in: Assistant Prof. Dipl.-Phys. Dr.rer.nat. Florian Reindl
Mitwirkung: Univ.Ass.in Leonie Einfalt, MSc

Wien, 11.12.2023

_____   _____
(Unterschrift Verfasser/in)   (Unterschrift Betreuer/in)

Eidesstattliche Erklärung

Hiermit versichere ich eidesstattlich, dass ich die vorliegende Arbeit eigenständig und ausschließlich unter Verwendung der im Quellen- und Literaturverzeichnis aufgeführten Werke angefertigt habe.

Wien, 11.12.2023

_____
(Unterschrift Verfasser/in)

Abstract

Direct detection dark matter searches look for the characteristic signal signature of dark matter particles scattering off standard model nuclei in their measured recoil energy spectra. The elusive nature of a dark matter signal introduces difficulties in its distinction from a background, especially when the aspects of the background are not fully grasped or determined. Even in instances where a signal is not detected, the experimenters can still establish an upper limit on the signal parameter. Due to the potential presence of unknown background processes, the experimenters cannot rule out the possibility that these unidentified factors might be significant enough to explain all observed events. By an appropriate choice of an interval containing fewer events than would be expected from the dark matter signal, it is possible to make the influence of the unknown background the least disruptive. Yellin's methods provide this region for the calculation of upper limits on the interaction strength of dark matter with standard model particles.

This thesis presents the development and application of the *ModeLimit* Python package, a versatile tool designed for modelling one-dimensional and two-dimensional expected signals in the context of dark matter research. Leveraging methods proposed by S. Yellin, including the maximum gap method, optimum interval method, and maximum patch method, *ModeLimit* allows the determination of upper limits for dark matter scattering cross sections. Notably, the package introduces an approximation to the optimum interval method, providing a resource-efficient alternative without compromising accuracy. The parallelisation of data tabulation processes significantly reduces run-time, particularly for experiments with a high expected number of events, resulting in an expanded size of tabulated data.

Results demonstrate *ModeLimit's* compatibility with published results and showcase its role in providing open-source, user-friendly solutions for dark matter analysis in a Python environment. The package not only aligns with the results of Yellin's original Fortran-based implementation but also allows applying the maximum patch method to two-dimensional data, which is the most important contribution of this thesis.

In cases where an experiment is collecting data from more than one channel, thus having multiple observables, the utilization of the maximum patch method becomes relevant and advantageous. Unlike conventional approaches relying on predetermined acceptance regions, the maximum patch method algorithmically identifies the optimal area for analysis. Even though the acceptance regions set by researchers by leveraging their knowledge of the data allow stricter limits on the signal parameter than without any acceptance region at all, they are inherently biased. Bias in experimental setups poses a formidable challenge in the pursuit of scientific accuracy. The inherent subjectivity introduced by manually set acceptance regions can skew results, potentially leading to misleading conclusions. By discarding the need for acceptance regions and finding the most optimal area for analysis algorithmically, *ModeLimit* significantly enhances the objectivity of the analysis, offering a more transparent and unbiased evaluation of dark matter interactions. *ModeLimit* emerges as a promising contribution to the field, offering researchers an accessible and efficient means to navigate the complexities of dark matter research.

Zusammenfassung

Bei der Erforschung von Dunkler Materie im Direktnachweis analysiert man die Rückstoßenergiespektren auf der Suche nach der charakteristischen Signatur von Teilchen der Dunklen Materie, die an Kernen des Standardmodells gestreut werden. Die schwer zu erfassende Natur eines Dunkle-Materie-Signals erschwert die Unterscheidung von Hintergrundrauschen, insbesondere wenn die Eigenschaften des Hintergrunds nicht vollständig verstanden oder festgelegt sind. Auch wenn kein Signal erkannt wird, haben die Forscher die Möglichkeit, eine obere Grenze für die Signalparameter anzugeben. Aufgrund potenziell unbekannter Hintergrundprozesse können die Forscher nicht ausschließen, dass der unbekannte Hintergrund hinreichend signifikant ist, um alle beobachteten Ereignisse zu erklären. Durch die geschickte Auswahl eines Intervalls mit weniger Ereignissen, als für das erwartete Signal der Dunklen Materie zu erwarten wären, kann der Einfluss des unbekannten Hintergrunds minimiert werden. Yellins Methoden bieten diesen Ansatz zur Berechnung von Obergrenzen für die Wechselwirkungsstärke der Dunklen Materie mit Teilchen des Standardmodells.

Diese Arbeit beschäftigt sich mit der Entwicklung und Anwendung des Python-Pakets *ModeLimit*, einem vielseitigen Werkzeug, das speziell für die Modellierung von ein- und zweidimensionalen erwarteten Signalen im Kontext der Dunkle-Materie-Forschung entwickelt wurde. *ModeLimit* greift auf Methoden zurück, die von S. Yellin vorgeschlagen wurden, darunter die Maximum-Gap-Methode, die Optimum-Intervall-Methode und die Maximum-Patch-Methode. Es ermöglicht die Ermittlung von Obergrenzen für Dunkle-Materie-Wirkungsquerschnitte. Insbesondere führt das Paket eine Annäherung an die optimale Intervallmethode ein, die eine ressourcenef-

fiziente Alternative darstellt, ohne die Genauigkeit zu beeinträchtigen. Die Parallelisierung der Datentabellierungsprozesse reduziert die Laufzeit erheblich, insbesondere bei Experimenten mit einer hohen erwarteten Anzahl von Ereignissen, was zu einer Erweiterung der tabellierten Daten führt.

Die Ergebnisse zeigen die Kompatibilität von *ModeLimit* mit veröffentlichten Ergebnissen und unterstreichen seine Rolle bei der Bereitstellung von benutzerfreundlichen Open-Source-Lösungen für die Analyse dunkler Materie in einer Python-Umgebung. Das Paket stimmt nicht nur mit den Ergebnissen der ursprünglichen Fortran-basierten Umsetzung von Yellin überein, sondern ermöglicht auch die Anwendung der Maximum-Patch-Methode auf zweidimensionale Daten, was als wesentlicher Beitrag dieser Arbeit herausgestellt wird.

Bei den Experimenten, die Daten aus mehr als einem Kanal sammeln und somit über mehrere Observablen verfügen, erweist sich der Einsatz der Maximum-Patch-Methode als relevant und vorteilhaft. Im Unterschied zu herkömmlichen Ansätzen, die auf vordefinierten Akzeptanzbereichen basieren, ermittelt die Maximum-Patch-Methode algorithmisch den optimalen Analysebereich. Obwohl die Akzeptanzbereiche, die Forscher basierend auf ihrem Wissen über die Daten festlegen, strengere Grenzen für den Signalparameter ermöglichen als ein fehlender Akzeptanzbereich, sind sie von Natur aus verzerrt. Verzerrungen in Versuchsanordnungen stellen eine große Herausforderung bei der Suche nach wissenschaftlicher Genauigkeit dar. Die inhärente Subjektivität, die durch manuell festgelegte Akzeptanzbereiche eingeführt wird, kann zu verfälschten Ergebnissen und potenziell verzerrten Schlussfolgerungen führen. *ModeLimit* verbessert die Objektivität der Analyse erheblich, indem es auf Akzeptanzbereiche verzichtet

und stattdessen den optimalen Bereich für die Analyse algorithmisch sucht. Dies bietet eine transparentere und unvoreingenommenere Bewertung der Wechselwirkungen mit Dunkler Materie. *ModeLimit* erweist sich als vielversprechender Beitrag in diesem Bereich, indem es Forschern ein zugängliches und effizientes Mittel bereitstellt, um die Komplexität der Dunkle-Materie-Erforschung zu bewältigen.

Contents

1 Introduction

In the pursuit of unravelling cosmic enigmas, the study of dark matter stands as a formidable challenge. Dark matter has not yet been detected. However, there are signs in the observations of galaxies or the cosmic microwave background radiation pointing to its existence. Hence, there are many theories for explaining the nature of dark matter particles and also many experiments, such as CRESST, aiming for the detection of dark matter particles.

The expected signal from nuclear scatterings of weakly interacting massive particles, one of the particle candidates for dark matter, is discussed in section 2. At first, the differential rate is described as a one-dimensional energy-dependent function. Then, using the one-dimensional differential rate, the two-dimensional probability density function is constructed in the energy-light yield space. This generalisation allows the usage of two-dimensional data in the analysis.

Now, even if the signal the researchers are trying to detect is below the sensitivity level of the experiment and if a positive detection eludes the researchers, it is still possible to set upper limits using the expected dark matter signal and the sensitivity of the experiment. However, such an experiment might be contaminated with an unknown background, which might be hard to remove. But, if the distribution of this unknown background were to be different than the distribution of the expected signal, this difference can be used to set a stronger upper limit using the methods proposed by S. Yellin [1, 2, 3]. These methods work with algorithmically set intervals where the influence of the unknown background is expected to be the least disruptive.

Three methods are discussed in this work: the maximum gap method, the optimum interval method and the maximum patch method. Apart from the maximum gap method, these analysis methods cannot be computed analytically and rely on conducting many Monte Carlo experiments. Thus, computers are essential for handling the large amounts of tabulated data. A resource-efficient approximation of the optimum interval method is introduced in section 3.3. Additionally, the parallelisation of various operations enabled further expediting of the process.

In addition to the optimum interval method, the more advanced but computationally demanding maximum patch method is also implemented in this work. Instead of working with a one-dimensional expected signal, it uses a two-dimensional density function and randomly tabulated data in a two-dimensional plane. While the maximum patch method does not necessarily yield much stricter limits than the optimum interval method, it helps to remove bias in the acceptance region applied to experimental data before applying the one-dimensional optimum interval method.

The original script for e.g. optimum interval calculations was written in Fortran in 2002 [3]. Fortran is currently not as widely used as many other programming

languages and thus few researchers can fully understand the script. This is where the need for an updated version of the implementation arises. For example, an implementation of the optimum interval method has already been realised before [4]. The Python package *ModeLimit* is the new open-source implementation of Yellin's methods in Python which not only allows users to model the expected signal of the dark matter particles depending on their mass but also to tabulate data with custom tabulation parameters and apply Yellin's methods to set upper limits on the cross sections of the dark matter. Exclusion charts as in Fig. 30 are composed of these upper limits.

Python is currently one of the most widely used programming languages in the field of data analysis. In comparison to Fortran, it is much easier to learn and code in Python, which is the primary advantage of implementing the methods in a Python package. It is modern and allows many more researchers to make use of the methods. The package will be available for download on GitHub.

The package can be utilised in three main ways. If the researcher is just interested in the expected dark matter signal as described in 2, *ModeLimit* can be used to calculate one- and two-dimensional probability density functions, and cumulative density functions, draw random variates stemming from these functions and determine the expected number of events. If, however, the researcher is also interested in setting upper limits on the cross section of dark matter particles, the researcher may use these generated expected signals in combination with Yellin's methods. The third case is where the researcher would like to utilise Yellin's methods, but use different formulae for the generation of the expected signal. In this case, the researcher may combine the expected signals calculated by other assumptions and means with a *ModeLimit* method to set limits on the cross section.

Almost all of the instance and class variables are changeable, which is of importance for flexibility reasons. Great care is taken in writing the docstrings as readable and understandable as possible to create a user-friendly experience when using the package. There will also be some examples provided, covering all the above-mentioned main usages, so that the researchers can almost directly use the package without needing to write even a single line of code but only change some required parameters fitting their experiment. The purpose of this study is that the ease of access brought forth by this new implementation helps researchers in their endeavour to understand and set stricter limits on the cross section of dark matter.

Section 5 delves into the outcomes and the implications of employing the *ModeLimit* package. The one-dimensional and two-dimensional density functions are discussed, providing a comprehensive understanding of the expected spin-independent dark matter signal. The discussion extends to the construction of exclusion charts, offering a visual representation of upper limits on the cross sec-

2

tions of dark matter particles. Subsequently, the examination of the advantages and drawbacks of the approximation of the optimum interval method serves as a pragmatic evaluation, ensuring that users can make informed decisions tailored to the specific nuances of their experimental context. Finally, the description of a recommended workflow acts as a practical guide, facilitating a seamless integration of *ModeLimit* into the researcher's toolkit.

It's important to note that the results discussed in section 5 stem from a blend of both real and simulated CRESST data. While the focus is on CRESST, it's crucial to emphasise that *ModeLimit* isn't confined to this experiment alone. Yellin's upper limit setting methods, the backbone of *ModeLimit*, have already been successfully employed in the analysis of data from the CDMS experiment. This experiment uses a different detector concept, utilising low-temperature germanium and silicon crystals [5].

The applicability of *ModeLimit* extends beyond these examples. Experiments such as XENONnT, LZ, and PandaX, which harness the distinctive properties of liquid xenon in their search for dark matter, can also leverage Yellin's methods implemented in *ModeLimit* [6, 7, 8]. These experiments detect signals labelled as S1 and S2, triggered by various events. While these signals differ from those in the CRESST experiment, the strength of *ModeLimit* lies in its adaptability. The methods can be effectively applied to the one- or two-dimensional data produced by these experiments, enabling the setting of upper limits and enhancing the analytical toolkit for a broader spectrum of dark matter searches.

1.1 A Brief Introduction to Dark Matter

Dark matter is a hypothetical form of matter and has not been experimentally detected as of yet. According to models of galaxy formation and evolution, as well as cosmological models, it is thought to be at least five times more abundant than baryonic matter, making up roughly 27% of the total energy density of the observable Universe [9]. One of the earliest documented observations of dark matter, if not the foremost, was conducted by Swiss astronomer Fritz Zwicky in 1933 during his investigation of the Coma Cluster [10]. However, Zwicky's "dark matter" was not thought of as some new hypothetical form of matter. He aimed to calculate the total dynamical mass of the Coma Cluster using its velocity dispersion in combination with the virial theorem. This dynamical mass he then compared to the luminous mass, deduced from the amount of light emitted. He found that the dynamic mass is some 400 times larger than the luminous mass in the Coma cluster. The gravitational effects observed could not solely be explained by the mass of the luminous matter in the galaxy cluster. Hence, as an explanation for this shortage of mass, "dark matter" as in "non-luminous matter" was proposed. Since it was thought to be just non-luminous matter causing the discrepancy in

results, its significance was not duly recognised.

Some 40 years later, the nature of this non-luminous matter became popular. In the 1970s, the American astronomers Vera Rubin and Kent Ford were studying the rotation of the neighbouring spiral galaxy Andromeda using the analysis of galaxy spectra [11]. Similar to Zwicky's findings, a discrepancy emerged between the mass of luminous matter and dynamic mass. Subsequent observations conducted by Rubin in the 1980s substantiated and extended the evidence supporting this discrepancy [12]. Analysing the spectrum data of spiral galaxies allows for determining the rotation curve, which describes the rotation speed of the galaxy as a function of the distance to the centre. Assuming a Keplerian behaviour, one would expect a decrease in the rotation speed for larger distances from the centre. This, however, was not observed. There was a flattening of the rotation curve. Rubin and Ford observed that the stars at the outer edge of the Andromeda galaxy, the closest galaxy to the Milky Way and a spiral galaxy, were rotating too fast. One possible explanation for this strange behaviour was the existence of a halo of non-visible matter surrounding the galaxies. This halo might constitute a large portion of the mass of the galaxy to explain the behaviour of the rotation curve. The flattening of the rotation curve is illustrated in Fig. 1.
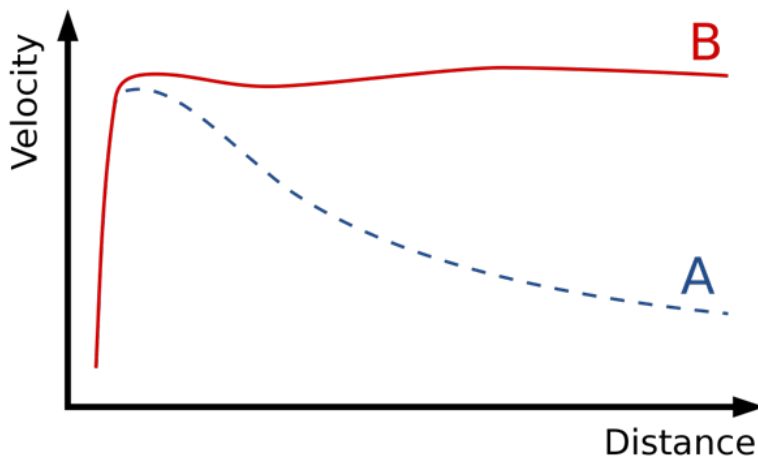

Figure 1: The expected and the observed rotation curves of a common spiral galaxy. The blue line, labelled "A", represents the expected rotation curve following Keplerian laws. The red line, labelled "B", represents the observed rotational curve of the galaxy. The resulting huge discrepancy and the flattening of the observed rotation curve may be explained with the help of a dark matter halo that constitutes a large part of the galaxy's mass [13].

The Bullet Cluster is one of the more recent observations that strongly supports the existence of dark matter [14, 15, 16]. The gravitational lensing caused by

the colliding clusters cannot be explained purely by the baryonic matter and the spatial offset of the centre of the mass from the baryonic mass is too large for a modification of the laws of gravity to be able to explain. An X-ray image of the collision is shown in Fig. 2. As can be seen from the image, the mass centres inferred from the gravitational lensing and the X-ray imagery do not align. The gaseous section, housing the majority of luminous matter, is slowed down due to friction, consequently trailing behind the dark matter.
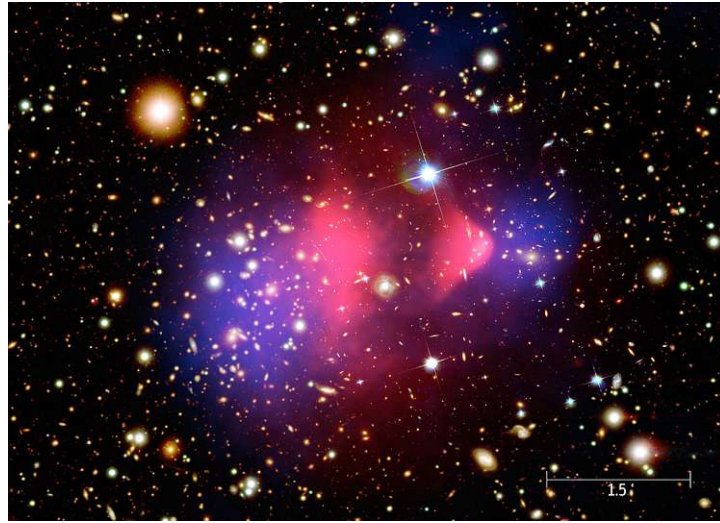
Figure 2: Superimposition of three images. The galaxies seen are from the visible light image. The pink image represents the X-ray image. The blue image represents the matter distribution inferred from the gravitational lensing. The discrepancy between the blue and pink image is visible [17].

The Cosmic Microwave Background (CMB) is the earliest snapshot of the universe, and one can see the existence of hot and cold spots in this early snapshot, as seen in Fig. 3. Two main forces are acting on matter, one being the gravitational force pulling matter together and the other one being the force exerted by photonic pressure pushing matter outward. Since dark matter does not interact with light at all, its existence, or non-existence, would radically change the patterns observed in CMB, because it would clump into dense regions by the gravitational pull of regular and dark matter. Accurate measurements provided by the Wilkinson Microwave Anisotropy Probe and ESA's Planck of the CMB fluctuations showed in their power spectrum that the existence of dark matter is more likely than not [18, 19]. Hence, in addition to the galactic rotation curves and the gravitational lensing of the Bullet Cluster, CMB is also in favour of the existence of dark matter.
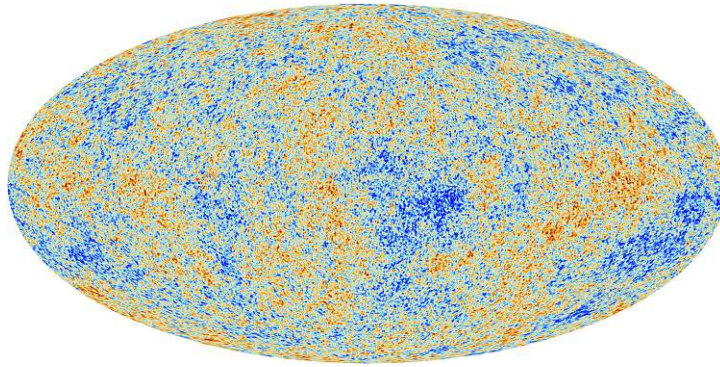
Figure 3: The anisotropies, the hot and cold spots, in the CMB can be seen in the CMB-plot calculated with the data provided by ESA's Planck spacecraft [20].

As to the true nature of dark matter, there are many hypotheses. Some fall under the category of the so-called hot dark matter, meaning they are fast, relativistic and relatively lighter, some fall under the category of cold dark matter, meaning they are slow, non-relativistic and heavier in comparison. The focus of this work will be **W**eakly **I**nteracting **M**assive **P**articles (WIMPs). The originally proposed WIMPs are not the same as the currently sought-after WIMPs, which usually differ in their mass and are usually called generalised WIMPs, being lighter than the original WIMPs. Nonetheless, they are non-baryonic matter candidates and are considered cold dark matter. The term "WIMPs" will be used interchangeably with "generalised WIMPs" in this study.

In general, there are three categories of dark matter experiments and corresponding ways of observing dark matter particles:

1. Dark matter particles can be found indirectly by looking for secondary particles created when they decay or annihilate. There are dozens of such studies currently underway or planned, particularly for gamma rays as secondary products [21, 22, 23, 24, 25].

2. Direct searches are the other kinds of experiments, which are of relevance to the CRESST Experiment. Through their interactions with the regular matter in a detector, they hope to directly detect dark matter particles [26, 6, 27, 28, 8, 29, 7]. The entire discussion in section 2 is completely devoted to such direct detection, especially for WIMPs.

3. There are also experiments being conducted, which could produce dark matter particles through high-energy collisions [30, 31, 32, 33, 34]. For example, if dark matter particles were created at the Large Hadron Collider (LHC), physicists could infer their production and existence from the amount of energy and momentum not observed after a collision.

1.2 CRESST Experiment

In Gran Sasso, Italy, the **C**ryogenic **R**are **E**vent **S**earch with **S**uperconducting **T**hermometers (CRESST) experiment searches for dark matter utilising detectors that operate at extremely low temperatures. Particle interactions generate heat in the detectors and this local rise in temperature allows events to be detected. Presented herein is a brief overview of the fundamental principles underlying cryogenic detection, the CRESST facility and its detector concepts using the information provided in [35].

At exceedingly low temperatures, the components of ordinary matter enter a state of inactivity, where even minute amounts of energy can give rise to significant consequences. A deposit of a few hundred electronvolts (eV) from a single particle interaction inside a CRESST crystal with a mass of $\sim \mathcal{O}(100 \text{ g})$, causes a measurable temperature spike in the attached superconducting film thermometer.

In order to investigate the interaction of dark matter particles in the laboratory, two competing prerequisites must be met. On the one side, given the extremely low interaction probability that dark matter particles are expected to have, a large exposure is required to get a quantifiable interaction rate. Large exposure can be achieved by increasing the crystal size. On the other side, the capacity to see minuscule energy deposits is required because the recoil energy of the impacted nucleus is expected to be low, with most events occurring in the keV range or lower. Given that a small crystal size is more adept at detecting minuscule energy deposits, it is crucial to find the right balance between these two competing prerequisites. Especially for these purposes cryodetectors are one of the frequently used detector technologies and produce leading limits on dark matter interactions in the low mass range. The full deposited energy in the particle recoil is detected by the cryodetectors, acting either as bolometers or calorimeters.

1.2.1 CRESST Facility

The low expected event rate for dark matter particle-nucleus elastic scattering necessitates a low background environment. In the detector, in addition to the dark matter particles, muons, neutrons, electrons, photons, and alpha particles will also interact. Cosmic rays, along with natural and induced radioactivity near the detector, may cause these interactions. If not suppressed, these background signals would be substantially more common than the signals expected from dark matter particles. As a result, the equipment is situated in a deep subterranean site under the Gran Sasso massif in Italy, which is covered on average by 1400 meters of granite.

Secondly, various passive shielding layers as shown in Fig. 4 shield ambient radiation emanating from the surroundings to the greatest extent possible. To

prevent radon from penetrating the shielding, the entire shielding is encased in an airtight aluminium container (the radon-box) that is constantly flushed with nitrogen gas and maintained at a minor overpressure. With the polyethylene neutron moderator in place, neutrons induced by muons in the lead of the shielding would dominate the remaining neutron flux. The muon veto system fitted inside the neutron moderator suppresses such a background.
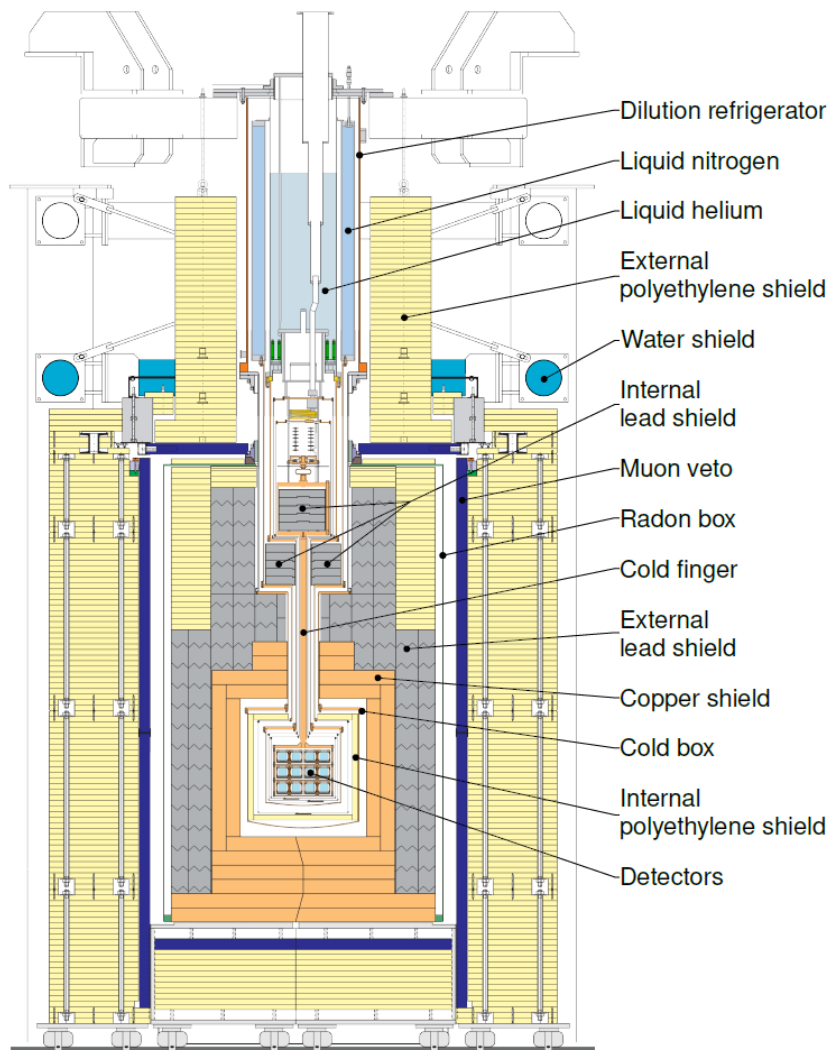

Figure 4: The CRESST experiment with its different shielding layers. These shielding layers are of paramount importance for the reduction of background noise, shielding the in-built absorber from radiation as much as possible [36].

1.2.2 Detector Concepts

The nuclei are in a cryogenic detector's absorber, which can detect the minuscule energy of a recoiling nucleus that has been hit by a dark matter particle. In the pursuit of detecting rare interactions with low energy deposits, a sensitive detector and highly effective noise reduction are essential. This is because various particles, not just dark matter particles, interact within the detector, and distinguishing them from the signals of interest is necessary.

An absorber and a temperature sensor in thermal contact, loosely coupled to a heat bath, make up the cryogenic calorimeter as shown in Fig. 5. Since the detector can be described as an absorber with a temperature-dependent heat capacity of C in a very simplified model, the temperature rise is a direct measure of the deposited energy. The energy deposition due to a particle interaction results in a detectable temperature increase at mK temperatures due to the T^3 dependence of the heat capacity in the phonon system.
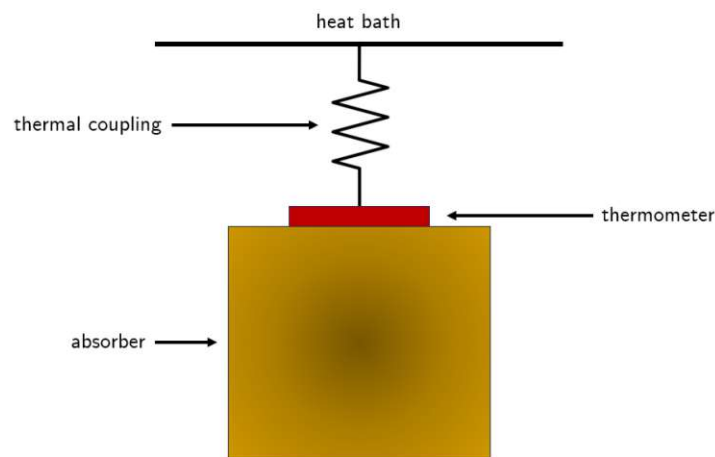
Figure 5: Schematic drawing of a CRESST calorimeter element. An absorber (brown) and a thermometer (red) are coupled to a heat bath.

The temperature sensors used by CRESST are superconducting phase transition thermometers, also known as transition edge sensors (TES). The thermometers are stabilised in the transition from the normal conducting to the superconducting phase, where a tiny temperature increase causes a noticeable increase in resistance, making them exceptionally sensitive thermometers. This is illustrated in Fig. 6. CRESST uses absorbers connected to such thermometers as cryogenic calorimetric detectors, which are extremely sensitive and can measure the total energy deposited by an interacting particle.
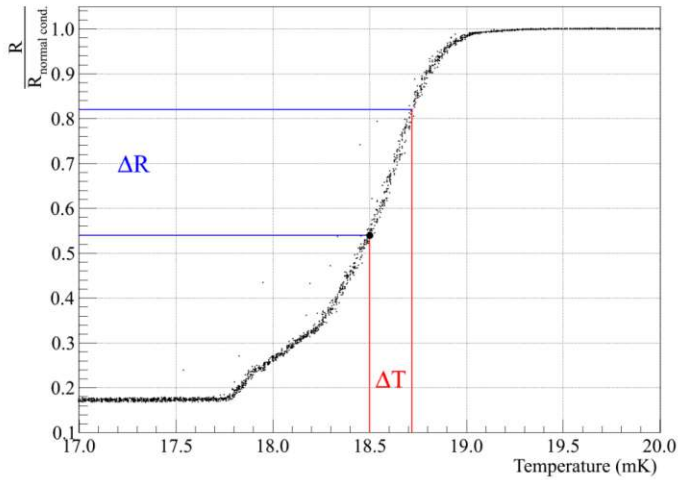
Figure 6: An example of a superconductor transition, showing the dependency of the normalised resistance on temperature. Normalisation is performed by dividing the resistance by the resistance in the normal conducting phase. Due to the steepness of the curve, even a minor change in temperature causes a noticeable change in resistance [37].

The detector modules for CRESST take advantage of the fact that most backgrounds create some light in a scintillating material, whereas the sought-after dark matter induced recoils produce little to no light. As a consequence, detectors based on scintillating $CaWO_4$ crystals as absorbers, shown in Fig. 7, were created. Particle interaction produces primarily heat in the form of phonons in this crystal. A small quantity of the deposited energy is also radiated as scintillation light.
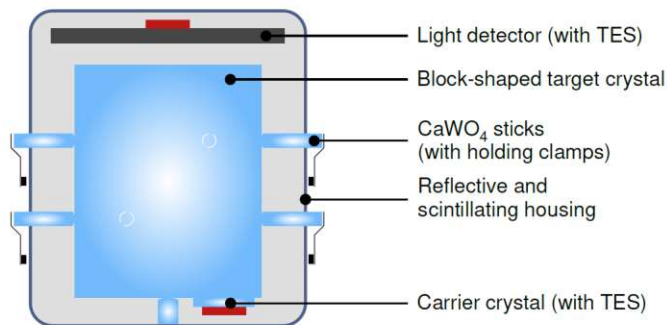
Figure 7: Schematic drawing of a CRESST-II detector module. In a reflection and scintillation housing (grey), the target crystal (blue) and the light detector (black) are built. Both are equipped with TES (red) that are thermally coupled to a heat bath [36].

10

Most common backgrounds (e.g. electron recoils, alpha events, etc.) can be separated from potential dark matter signals because the amount of light produced varies depending on the type of particle. Note that this is only true under the assumption that dark matter particles interact with standard model nuclei and are thus suitable for direct detection. The so-called light yield, $LY = E_l/E_p$, of an event is defined as the ratio of the energies deposited in the light and phonon channels and can be used to distinguish between various particle interactions, see Figs. 8 and 9.
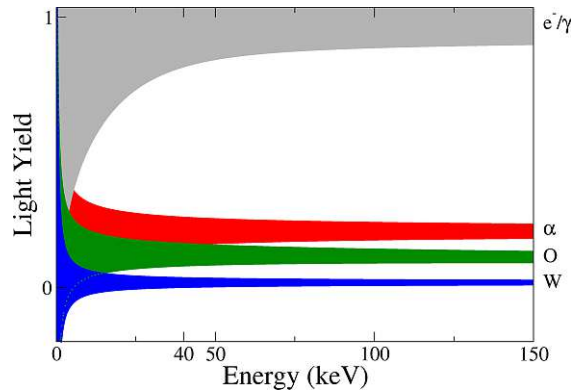
Figure 8: Light yield for different kinds of particles and recoil types. The electron/gamma band is represented by the grey area, the alpha particles band by the red area, the oxygen band by the green area, and the tungsten band by the blue area [35].

The horizontal bands originating from various event types are illustrated in the light yield-energy plane in Fig. 8. As aforementioned, the light yield of different event types can differ substantially. The finite detector resolution allows the partial separation of these bands.

Now events from a specific detector will be closely examined. The data from this detector is then later analysed using the methods from *ModeLimit*, which will be discussed in later sections. The events in the Americium-Beryllium neutron calibration data that passed the selection criteria, with all the relevant cuts applied, are shown in Fig. 9. The 90% upper and lower bounds of the e^-/γ-band are indicated by solid blue lines. The expected bands for recoils off oxygen and tungsten are indicated by the red and green lines, respectively. The calcium band, which is not drawn for clarity, is located between the oxygen and tungsten bands.
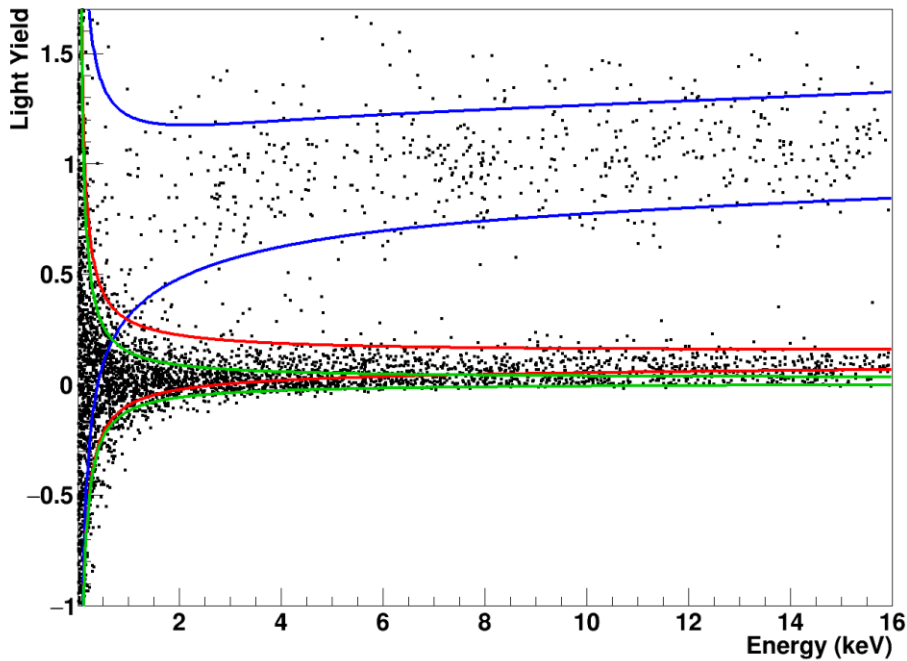
11

Figure 9: Neutron calibration data for an exemplary CRESST-III detector in the light yield versus energy plane. The data are fit to determine the bands for e^-/γ-events (blue), nuclear recoils off oxygen (red), and tungsten (green), where the respective lines correspond to the upper and lower 90% boundaries of the respective band [26].

Following the application of all the cuts, Fig. 10 displays the background data in the light yield versus energy plane. As in Fig. 9, the blue, red, and green bands represent the e^-/γ-events and nuclear recoils off tungsten and oxygen, respectively. The mean of the oxygen band is depicted by the red dashed line, which simultaneously marks the upper boundary of the acceptance region, shaded in yellow. The tungsten band's 99.5% lower limit, with an energy range of 30.1 eV to 16.0 keV, serves as the lower bound of the acceptance region. Red-highlighted events in the acceptance region are regarded as possible dark matter candidate events.

The acceptance region refers to the parameter space or set of conditions within which a detector or experiment is sensitive to and capable of detecting potential signals from dark matter interactions. The goal is to establish a well-defined set of conditions where the experiment can effectively distinguish between potential dark matter interactions and other sources of noise or background signals. The specific characteristics of the acceptance region depend on the design and capabilities of the particular dark matter experiment or detector being employed. In CRESST's

case, this parameter space is defined along the light yield axis by the mean of the oxygen band and the 99.5% lower limit of the tungsten band.
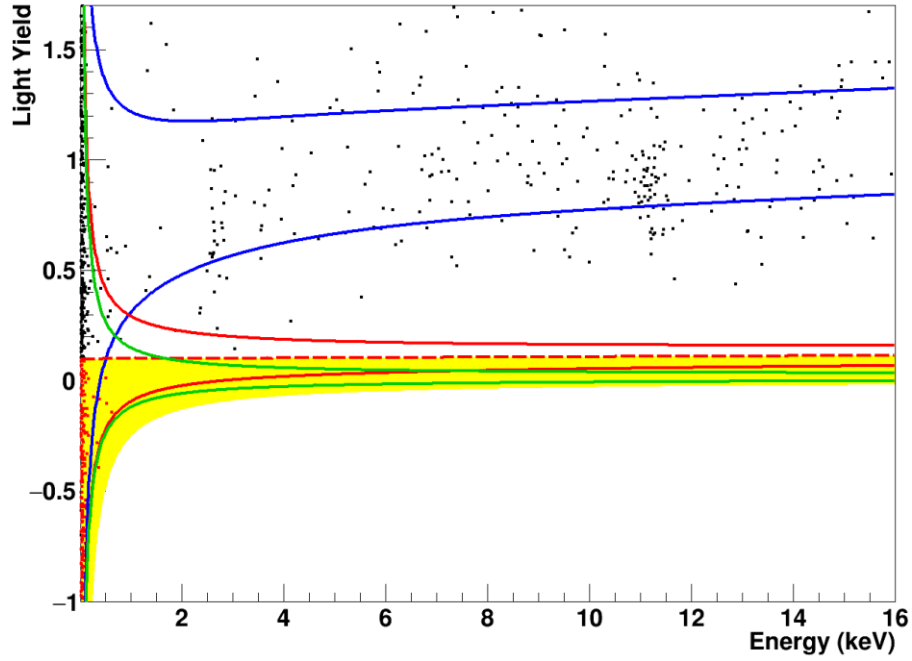

Figure 10: The light yield versus the energy of events in the dark matter data set, following the application of selection criteria. The 90% upper and lower bounds of the e^-/γ-band are shown in blue, while oxygen and tungsten are represented by red and green bands, respectively. The red dashed line is the mean of the oxygen band, and the yellow area represents the acceptance region extending from that point down to the lower 99.5% border of the tungsten band. Events in the acceptance region are highlighted in red. As in Fig. 9, the bands' positions were derived from the neutron calibration data [26].

13

2 Description of Expected DM Signals

WIMPs, alongside Axions, are currently the non-baryonic dark matter candidates that are being studied the most in-depth and are thought to have the best detection potential. Therefore, like with the CRESST experiment, many dark matter searches focus specifically on WIMPs. The foundational ideas and characteristics of WIMP searches will be covered in this section.

According to the dark matter density at Earth's position in the Milky Way ρ_χ

$$\rho_\chi := 0.3 \frac{\text{GeV}/c^2}{\text{cm}^3}, \tag{1}$$

one might anticipate that due to its speed through the galaxy, the Earth will experience a WIMP flux of the order of 10^6 cm^{-2}s^{-1} for an assumed WIMP mass of $(100\,\text{GeV}/c^2)$ [38]. A small but detectable portion of these particles may scatter off ordinary matter in a detector. These interactions do not have to be standard model interactions for them to be detected and WIMPs can transfer considerable energy, primarily to nuclei, through these interactions in the standard scenario elastic scattering scenario. Direct search experiments, as described in 1.1, record the rate and energies of nuclear recoils in dedicated detectors. The expected event rate for such WIMP-induced nuclear recoils can be estimated and such recoil events may be distinguished from background events from other sources. The following sections 2.1 and 2.2 rigorously adhere to the notation and formulae established by Jens Schmaler, as described in [39]. Furthermore, in the context of employing a non-rotating model, the formulations and notation described in [40] will be followed.

2.1 Modelling the Expected DM Signal

The one-dimensional projection of the expected dark matter signal, as a function of the recoil energy, necessitates a comprehensive examination of four pivotal subtopics. The initial considerations involve the definitions of relevant cross sections integral to the description of the WIMP-nucleon scattering and the introduction of the form factor to accommodate the substructure inherent in nuclei. Subsequently, attention is directed towards the total interaction rate and the differential rate. Lastly, the evaluation of these rates requires a detailed examination of insights derived from astrophysics.

2.1.1 Cross Section

The differential WIMP-nucleus cross section $\frac{d\sigma}{dE_R}$ incorporates all of the inputs from nuclear and particle physics, particularly the WIMP interaction features. In general, the WIMP-nucleus cross section is made up of two constituents: a

14

spin-dependent cross section describing a potential coupling of the dark matter particles to the net spin of the target nuclei and a spin-independent cross section summarising the scalar interactions.

These two contributions are independent of each other and can thus be summed together to give the total cross section. The spin-dependent component, however, can be disregarded if the target nuclei have no or minimal net spin, as is the case in the CRESST experiment with $CaWO_4$ as the target material. Thus, the spin-independent interaction in the following is the main topic of discussion.

Most theories generally assume a similarity in the coupling of WIMPs with various nucleons. However, the formulae presented below will utilise quantities related to protons rather than neutrons. The cross section for this coupling can be formulated by considering the nuclear mass m_N, the mass number A, the strength of the interaction between the protons and WIMPs f^p, the form factor F and the velocity v of WIMPs [41]:

$$\frac{d\sigma}{dE_R} = \frac{2m_N A^2 (f^p)^2}{\pi v^2} F^2(E_R) \tag{2}$$

As seen in Eq. (2), the differential cross section is not dependent on the direction of the velocity, but solely on its absolute value. The factor A^2 represents the outcome of the coherent interaction with the target nuclei as a whole, where the scattering amplitudes with all A nucleons add up in phase. The form factor that describes the partial loss of coherence at greater momentum transfers is the function $F(E_R)$ and is discussed in more detail in section 2.1.2.

The relative speed of WIMPs is expected to be below the galactic escape velocity at the order of 10^2 km s^{-1}, because otherwise, they could not be tied to the Milky Way and their density in the galaxy would thus be negligibly small. If the density is negligibly small, so is the rate of interactions with these WIMPs. Therefore, the focus can be directed towards WIMPs with non-relativistic relative speeds. In this case, the recoil energy is dependent on the scattering angle Θ in the centre of the mass frame and can be expressed in a straightforward manner

$$E_R = \frac{\mu_N^2 v^2 (1 - \cos\Theta)}{m_N}, \tag{3}$$

where

$$\mu_N := \frac{m_\chi m_N}{m_\chi + m_N} \tag{4}$$

is defined as the reduced mass of the WIMP-nucleus system, with the dark matter mass m_χ. Eq. 4 reaches its maximum value at the largest equal m_χ and m_N values and $\cos\Theta$ equals 1 for $\Theta = \pi$. With $v \approx 10^{-3}c$, maximum expected recoil energies are at the order of $\mathcal{O}(10 \text{ keV})$. Another outcome of the non-relativistic limit is that

15

WIMP's energy is not high enough to probe the target nuclei's substructure and will instead mostly interact with the nucleus as a whole.

Eq. (3) can be used to introduce $E_{R,max}$ as the highest possible energy transfer of a WIMP with a velocity of v for $\cos\Theta = 1$:

$$E_{R,max} = \frac{2v^2\mu_N^2}{m_N} \tag{5}$$

Using Eq. (5), Eq. (2) can be reformulated to define the total point-like WIMP-nucleus scattering cross section σ_0 via

$$\frac{d\sigma}{dE_R} =: \frac{\sigma_0}{E_{R,max}(v)}F^2(E_R). \tag{6}$$

By comparing Eqs. (2) and (6), it can be seen that the quantity σ_0 is dependent on the mass number of the target nucleus. Therefore, it is not the best measure to use for comparing results from different experiments, as different experiments employ different nuclei. For example, for xenon experiments, the relevant nucleus is that of Xe, for CRESST the relevant nuclei are those of Ca, W and O. For this reason, it is important to introduce a new normalised WIMP-nucleon cross section σ_{WN}, where the dependence of the specific target is avoided [40]

$$\sigma_{WN} = \left(\frac{1 + m_\chi/m_N}{1 + m_\chi/m_p}\right)^2 \frac{\sigma_0}{A^2}. \tag{7}$$

Here, m_p denotes the proton mass. In direct dark matter searches, σ_{WN} is the quantity that is most frequently discussed and compared. σ_{WN} is also the quantity for which *ModeLimit* determines the upper limits and thus constrains the interaction strength of WIMPs.

2.1.2 Form Factor

The form factor, denoted as $F^2(E_R)$, is essential for the description of the cross section as introduced in Eq. (2). It accounts for the non-point-like substructure of nuclei, introducing a loss of coherence in WIMP-nucleus interactions at higher momentum transfers. This phenomenon is a consequence of the nuclei having internal structure rather than being treated as point masses.

In conventional dark matter searches, the form factor is often parameterised using a model initially proposed by Helm and further refined by Engel [42, 43]. This model conceptualises the nucleus as a convolution of a sphere with constant density and a skin function

$$\rho(\vec{r}) = \int d^3r' \Theta(R_0^2 - r'^2) \cdot const. \cdot \exp\left(-\frac{(\vec{r} - \vec{r}\,')^2}{2s^2}\right). \tag{8}$$

16

R_0^2 can be defined by $(R^2 - 5s^2)$ with the nuclear radius $R \simeq 1.2A^{1/3}$ fm and skin thickness $s \simeq 1$ fm. The skin function, modelled by a Gaussian, accounts for the gradual loss of density near the sphere's edge, providing a representation of the distribution of the scattering centres within the nucleus. Applying the Fourier transformation to this density yields the form factor

$$F(q) = 3\frac{j_1(qR_0)}{qR_0} \exp\left(-\frac{1}{2}q^2s^2\right) \qquad (9)$$

with the momentum transferred in the scattering process $q = \sqrt{2m_N E_R}$. j_1 denotes the first spherical Bessel function.

The analytical calculability of this form factor for any target nucleus, coupled with its strong approximation capabilities for numerous nuclei, especially those with low mass numbers, makes it favourable [43]. Although the Helm model provides a straightforward approach, its simplicity might limit its ability to accurately capture the true behaviour of nuclear density. Notably, parameters such as the nuclear radius R_0^2 are often determined by fitting to a broad range of nuclei, making them less precise representations of individual elements.

To address this limitation, more sophisticated formulae for estimating the nuclear radius, like the one proposed by Lewin and Smith [44], can be employed. It has been observed that using altered formulations for R_0, such as

$$R_0 := \sqrt{c^2 + \frac{7}{3}\pi^2 a^2 - 5s^2}, \qquad (10)$$

can lead to better agreement with scattering data obtained from various nuclei when employing a Helm-type form factor as in (9). The parameters a, s, and c are defined as follows:

$$a := 0.52 \text{ fm}, \ s := 0.9 \text{ fm}, \ c := 1.23 \cdot A^{1/3} - 0.6 \text{ fm}. \qquad (11)$$

Fig. 11 depicts the form factor for different nuclei present in the CRESST experiment, namely Ca, W and O. As anticipated by Eq. 9, the form factors corresponding to distinct nuclei result in different values at identical energy levels. The effect of the spherical Bessel functions is best observed in the form factor of tungsten.
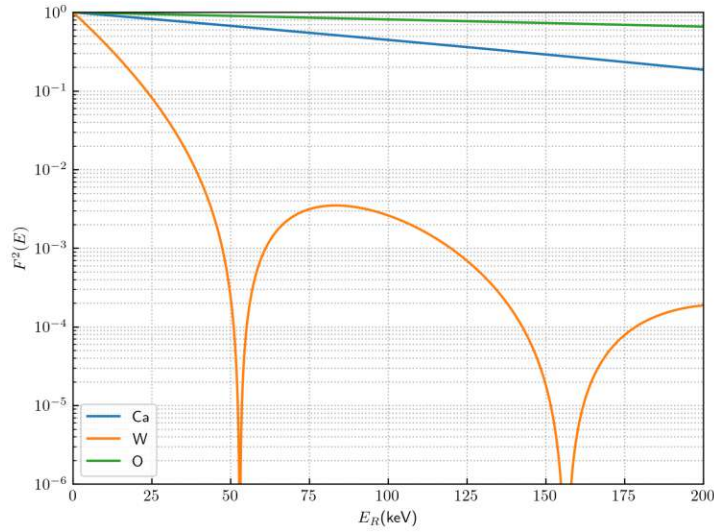
Figure 11: The squared form factor $F^2(E)$ for Ca, W and O as a function of the transferred energy E_R.

2.1.3 Interaction Rates

The total interaction rate, denoted as Γ, in a detector penetrated by a flux of WIMPs, denoted as Φ, can be generally written as a function dependent on the WIMP-nucleus scattering cross section σ:

$$\Gamma = n_{target}\Phi\sigma(v) \tag{12}$$

n_{target} stands for the density of the target nuclei with a nuclear mass of m_N in the detector with a total mass of the detector M_{target} and can be calculated by $n_{target} = M_{target}/m_N$. The flux of WIMPs can be written as:

$$\Phi = \frac{\rho_\chi}{m_\chi}v \tag{13}$$

ρ_χ is the aforementioned mass density of the WIMPs at Earth's location in the Milky Way (1) and v is their relative speed. Combining Eqs. (12) and (13) results in:

$$\Gamma = \frac{M_{target}}{m_N}\frac{\rho_\chi}{m_\chi}v\sigma(v) \tag{14}$$

The rate that will be modelled and is vital for setting upper limits on cross section is not the total rate, but the differential event rate, denoted by $\frac{d\Gamma}{dE_R}$ and often defined in terms of counts per kilogram of target material, per day, and per keV of recoil energy. To determine the differential rate, the expression in (14) must be differentiated with regard to the recoil energy. Additionally, the fact that not

18

all WIMPs have the same relative speed v has to be taken into consideration; it necessitates the incorporation of a velocity distribution $f(\vec{v})$ leading to

$$\frac{d\Gamma}{dE_R} = \frac{\rho_\chi}{m_N m_\chi} \int_{v_{min}}^{\infty} d^3v\, f(\vec{v})\, v \frac{d\sigma(\vec{v}, E_R)}{dE_R},$$

(15)

where v_{min} is the lowest possible speed of a WIMP which can induce a nuclear recoil of energy E_R. By reformulating Eq. (3) with $\cos\Theta = 1$, v_{min} can be expressed as:

$$v_{min} = \sqrt{\frac{E_R m_N}{2\mu_N^2}}$$

(16)

Although the velocity integral's upper limit in Eq. (15) is formally infinite, it is crucial to bear in mind that, as aforementioned, WIMPs with velocities greater than the galaxy's escape velocity cannot be bound to the galaxy and will therefore have very low densities. Therefore, the integral can be truncated at the escape velocity.

The derivative of the cross section in regard to the recoil energy in Eq. (15) has already been discussed in section 2.1.1. By using Eqs. (6) and (7), it is possible to reformulate Eq. (15) to

$$\frac{d\Gamma}{dE_R} = \frac{\rho_\chi}{2m_\chi \mu_p^2} A^2 \sigma_{\text{WN}} F^2(E_R) \underbrace{\int_{v_{min}}^{\infty} d^3v \frac{f(\vec{v})}{v}}_{=:I(v_{min})},$$

(17)

where the reduced mass μ_p is defined as

$$\mu_p := \frac{m_\chi m_p}{m_\chi + m_p}.$$

(18)

Thus, the differential rate becomes dependent on the normalised WIMP-nucleon cross section σ_{WN}. This final expression of the differential rate is the one that will be modelled in this thesis and *ModeLimit*. However, the calculation of this rate necessitates the evaluation of the final integral term by considering the astrophysical inputs.

19

2.1.4 Astrophysical Inputs

Now, attention can be directed to the integral $I(v_{min})$ in Eq. (17). In addition to the WIMP density ρ_χ already provided in Eq. (1), the integral $I(v_{min})$ over the velocity distribution is the other astrophysical input required to evaluate the differential rate. The most basic assumption in this case is that the WIMP velocities $\vec{v}_G$ in the rest frame of the galaxy follow a Maxwell-Boltzmann distribution

$$f_G(\vec{v}_G) = \frac{1}{\mathcal{N}}\left(\frac{3}{2\pi w^2}\right)^{3/2}\exp\left(-\frac{3v_G^2}{2w^2}\right), \tag{19}$$

where w is the root mean square velocity of the dark matter particles and $\mathcal{N}$ is a constant normalisation factor. The distribution function in the differential rate Eq. (17) would be exactly f_G, if the WIMP velocities followed this distribution up to infinity, and, if Earth was at rest in relation to the galaxy. The integral $I(v_{min})$ would have the following shape under the oversimplified assumptions and through the usage of the definition of v_{min} in Eq. (16):

$$I(v_{min}) \propto \exp\left(-\frac{v_{min}^2}{w^2}\right) \propto \exp\left(-E_R\right) \tag{20}$$

Nonetheless, a general notion of the shape of the differential recoil rate as a function of recoil energy can already be derived from this simplified case: It is effectively the product of an exponentially decaying function and the form factor, which is energy-dependent, see Fig. 11.

The Earth, however, is not at rest in relation to the galaxy and its average velocity, denoted by $v_\oplus$, will be taken as the constant value of

$$v_\oplus = 220 \cdot 1.05 = 231 \text{ km s}^{-1}. \tag{21}$$

The periodic effect of Earth's rotation around the Sun to its velocity is disregarded and the main component of the velocity comes from the Sun's movement through the Milky Way.

Furthermore, as aforementioned, the WIMP velocity distribution must be truncated at the Milky Way's escape velocity. This can be achieved by using a modified normalisation factor [44]

$$\mathcal{N} = \text{erf}(z) - \frac{2}{\sqrt{\pi}}z\exp\left(-z^2\right) \tag{22}$$

with z being defined as

$$z^2 := \frac{3v_{esc}^2}{2w^2} \tag{23}$$

20

and setting v_{esc} as the integral's upper limit. In the case of a non-rotating model, as described in [40], this integral takes the form of a piece-wise defined function where one has to differentiate between three cases:

$$I(v_{min}) = \kappa \cdot \begin{cases} \chi(x_{min} - \eta, x_{min} + \eta) - 2\eta \exp\left(-z^2\right) & x_{min} < z - \eta \\ \chi(x_{min} - \eta, z) - \exp\left(-z^2\right)(z + \eta - x_{min}) & z - \eta \leq x_{min} < z + \eta \\ 0 & x_{min} \geq z + \eta \end{cases} \tag{24}$$

with

$$\eta^2 := \frac{3v_\oplus^2}{2w^2}, \quad \kappa = \frac{1}{\mathcal{N}\eta}\left(\frac{3}{2\pi w^2}\right)^{1/2}, \quad x_{min}^2 := \frac{3v_{min}^2}{2w^2} = \frac{3m_N E_R}{4\mu_N^2 w^2}, \tag{25}$$

and the function

$$\chi(x, y) = \frac{\sqrt{\pi}}{2}[\mathrm{erf}(y) - \mathrm{erf}(x)]. \tag{26}$$

The constants w and v_{esc} used in the above expressions can now be defined. Under the assumption of the halo model of an isothermal sphere, the root mean square velocity w in the galactic halo is dependent on the asymptotic value of the rotational velocities v_∞

$$w = \sqrt{\frac{3}{2}}v_\infty. \tag{27}$$

Thus, using the International Astronomical Union standard value of $v_\infty \approx 220$ km s^{-1}, this yields

$$w = 270 \text{ km s}^{-1}. \tag{28}$$

The galactic escape velocity will be set to the typically used value of [45]:

$$v_{esc} = 550 \text{ km s}^{-1} \tag{29}$$

Using all the formulae discussed in section 2.1, it is now possible to model the one-dimensional expected dark matter signal. Note that the detector-specific quantities such as cut efficiency etc. are not discussed and implemented in this section. As mentioned at the beginning of section 2.1, this whole chapter closely adhered to the notation and formulae described in [40] and [39].

2.2 Transition to Two-Dimensional Density Function

The differential rate formulated in (17) is only dependent on the total nucleon-WIMP scattering energy and thus one-dimensional. As described in section 1.2, however, it is seen that the CRESST experiment not only collects energy but also light yield data of scattering events. Therefore, to be able to make use of these secondary data points, it is mandatory to transition from a one-dimensional differential rate to a two-dimensional density function in the energy-light yield plane

$$\rho(E_R, L) := \frac{\partial^2 \Gamma(E_R, L)}{\partial L \partial E_R} \text{ or } \rho(E_R, LY) := \frac{\partial^2 \Gamma(E_R, LY)}{\partial LY \partial E_R}, \tag{30}$$

with L being the light energy and LY the light yield.

Since the light yield data is also an important factor in deciding whether an event is more likely to be a signal event or not, it is essential to not limit the analysis to only one dimension. The need for a dimensional expansion can also be seen when looking at Fig. 9. If the focus were to be solely on the energies of events, distinguishing between events in the e^-/γ band and those in the nuclear recoil bands becomes challenging. The latter, however, is more likely to be a signal event. So, by limiting the analysis to one dimension, the weights of different data points by their light yield values would be lost and the limits set would be weaker.

The following formulation of the two-dimensional density function will make use of formulae that can be found in [39] and [46] with modified notation matching the notation from section 2.1.

The relationship between light yield, light energy and recoil energy can be expressed as

$$LY = L/E_R. \tag{31}$$

The mean light yield of the e^-/γ band, from hereafter denoted LY_e with e for electron, can be phenomenologically parameterised as

$$LY_e(E_R) = (L_0 + L_1 E_R)(1 - L_2 \exp(-E_R/L_3)), \tag{32}$$

with L_0, L_1, L_2 and L_3 being parameters that are derived for each detector module through the usage of maximum-likelihood methods. The latter two parameters account for the scintillator non-proportionality [47].

The energy-dependent quenching factor, $QF(E_R)$, for a certain nucleus N can be written as

$$QF_N(E_R) = LY_N(E_R)/LY_{e,norm}(E_R). \tag{33}$$

$LY_N(E_R)$ is the mean light yield of a nuclear recoil occurring at nucleus N and $LY_{e,norm}$ is a normalisation factor characteristic to the detector which embodies the mean light yield of e^-/γ events. Conventionally, this normalisation factor is

22

given by the mean light yield of the e^-/γ band (32) without the detector-specific non-proportionality factors and is thus expressed as:

$$LY_{e,norm}(E_R) = L_0 + L_1 E_R \tag{34}$$

Using the expressions (33) and (34) the mean light yield of a nucleus can be rewritten as

$$LY_N(E_R) = (L_0 + L_1 E_R)QF_N(E_R). \tag{35}$$

The scattering cross section of neutrons interacting with nuclei is dependent on mass, particularly evident at higher energies where nuclear recoil events for heavier nuclei become less probable. Notably, beyond an energy threshold, such as 350 keV, a predominant proportion of nuclear recoil events involve scatterings with oxygen nuclei. For lower recoil energies, the nuclear recoil events involving calcium nuclei become more frequent, with the emergence of the tungsten recoil band only observable for energies below approximately 240 keV. Given the considerable overlap in nuclear recoil bands at recoil energies below 350 keV, the phenomenological parametrisation in Eq. 36 is deemed more appropriate for accurate representation [46]. Note that the variation in the mean light yield is influenced by the energy dependence associated with the nuclear mass A [46]. Consequently, the light yield means for Ca and O exhibit more changes in this energy region compared to W, which can be considered nearly constant. The new phenomenological parametrisation is similar in structure to the (32)

$$LY_N(E_R) = LY_N^\infty(1 + a_N \exp{(-E_R/d_N)}), \tag{36}$$

where LY_N^∞ stands for the light yield of the nucleus for infinite recoil energy, a_N for the fraction of energy-dependent component and d_N is responsible for the rate of exponential decay with energy.

Now, the next step involves examining the quenching pattern of the crystals in greater detail. Different $CaWO_4$ crystals show varying quenching behaviour [47]. This variation seems to be dependent on the crystal's optical quality. The relative quantity QF exhibits a lower value when a crystal possesses a reduced defect density, resulting in a higher overall light output. In other words, an increased defect density has a lesser impact on the light yield of nuclear recoils. This quenching variation can be simply accounted for by introducing a scaling factor $\epsilon_i = \overline{QF_{N,i*}}/\overline{QF_N}$ with index i standing for different detector modules. The quenching factor of the nucleus N can be calculated for each module individually and by rearranging the expression the quenching factor $QF_{N,i*}$ can be expressed as

$$QF_{N,i*}(E_R) = \epsilon_i QF_N(E_R). \tag{37}$$

23

Thus, through combining the parametrisations and taking the variations in the quenching factors into account, Eqs. (35), (36) and (37) can be distilled into one equation describing the mean light yield of a nucleus N:

$$LY_N(E_R) = (L_0 + L_1 E_R) \cdot \epsilon_i \cdot QF_N(E_R) \cdot (1 + a_N \exp(-E_R/d_N)) \qquad (38)$$

Finally, by utilising the one-dimensional differential rate and assuming that events follow Gaussian distributions in the light yield plane, centred around the mean values of the bands with varying standard deviations at each energy, the two-dimensional density function can be formulated as:

$$\rho_N(E_R, LY) = \frac{d\Gamma_N}{dE_R}(E_R) \frac{1}{\sqrt{2\pi\sigma_N^2(E_R)}} \exp\left(-\frac{(LY - LY_N(E_R))^2}{2\sigma_N^2(E_R)}\right) \qquad (39)$$

The energy-dependent standard deviation can be calculated using the resolutions of the light and phonon detectors with

$$\sigma_N(E_R) = \sqrt{(\sigma_L(L_N(E_R))^2 + \left(\frac{dL_N}{dE_R}\sigma_P(E_R)\right)^2}, \qquad (40)$$

where

$$\sigma_L(L_N(E_R)) = \sqrt{\sigma_{L,0}^2 + \sigma_{L,1} L_N(E_R) + \sigma_{L,2} L_N(E_R)^2} \qquad (41)$$

and

$$\sigma_P(E_R) = \sqrt{\sigma_{P,0}^2 + \sigma_{P,1}^2(E_R^2 - E_{thr}^2)}. \qquad (42)$$

Parameters $\sigma_{L,0}$ and $\sigma_{P,0}$ are given by the baseline resolution of the light and phonon detectors respectively. Parameters $\sigma_{L,1}$, $\sigma_{L,2}$ and $\sigma_{P,1}$ are derived from calibration data. Note that the Eq. (39) is dependent on light yield but the $\sigma_N(E_R)$, $\sigma_L(E_R)$ and $\sigma_P(E_R)$ are light-dependent. For these expressions to be compatible with each other, convert standard deviation values from light to light yield adhering to the principle outlined in expression (31).

3 Yellin's Methods for Setting Limits

The methods discussed in this section can be used to determine upper limits from experiments with very low backgrounds; nevertheless, they cannot be applied to identify a positive detection. Prior to delving into Yellin's methods, it is beneficial to take a closer look at why this method for analysing the experimental data is useful and what Yellin's maximum gap method is, which can be thought of as the first step before the optimum interval method. Afterwards, an examination of a potential approximation of the optimum interval method will follow, which can be utilised for getting rough estimates of the limit results with much shorter runtimes and this approximation is not as computationally expensive as the optimum interval method. Finally, an explanation of how the maximum patch method operates will be provided, highlighting its utilisation of two-dimensional data in contrast to the maximum gap and optimum interval methods.

Consider the events in the acceptance region of the CRESST experiment as shown in Fig. 14. There may be known and/or unknown background processes with some degree of uncertainty regarding their rate or shape generating the events. Known backgrounds can be accounted for and their effect on the results can be thus minimised. If the experimenters are unable to rule out the possibility that the unknown background is big enough to explain all the observed events, they are only able to report an upper limit on the signal variable, e.g. σ with a significant certainty. Even though if they were to make a signal claim, the uncertainty bounds would be even larger in comparison to the limits.

The methods covered in this section adopt the following strategy: Using the expected dark matter signals described in Eqs. (17) and (39), the algorithms determine the intervals containing fewer events than would be anticipated from the expected dark matter signal. Thus, the algorithm renders the unknown background the least disruptive. This is of importance, especially for lower energy regions, where the number of events can reach several hundred, as in Figs. 14 and 10. Furthermore, the automatic selection of the interval makes these methods unbiased.

For a more comprehensive understanding of these methods and detailed derivations of the formulae, please refer to Yellin's papers [1, 2].

3.1 Maximum Gap Method

Fig. 12 depicts the maximum gap method for CRESST-III data in the acceptance region for a dark matter mass of 1 GeV. The black ticks along the horizontal axis indicate events, and their location on the axis corresponds to the measured recoil energy E_R. The curve displays the normalised expected dark matter probability density function, denoted by $\widetilde{\Gamma}_n$. This function can be calculated by normalising

25

Eq. (59). The maximum gap can be found by calculating the integral of the normalised probability density function between each neighbouring event along the energy axis and determining the largest integral. If there are N events, then there are a total of $N+1$ gaps to be calculated, accounting for the two gaps at the boundaries corresponding to the minimum and maximum energies. Integrating the normalised probability density function between two event energies can be substituted by subtracting the values of the normalised cumulative density function, $\widetilde{I}_n$, evaluated at these event energies:

$$x = \max\left\{ x_i = \int_{E_i}^{E_{i+1}} \frac{d\widetilde{\Gamma}_n}{dE_R} : i = 0, 1, \ldots, N-1 \right\} \tag{43}$$

$$x = \max\left\{ x_i = \widetilde{I}_{n,i+1} - \widetilde{I}_{n,i} : i = 0, 1, \ldots, N-1 \right\} \tag{44}$$

Note that x_i is an invariant term when performing bijective coordinate transformations. Eq. (43) describes the calculation of the gaps by using events defined in the recoil energy space. However, Eq. (44) results in the same gaps by using the transformed values, which are calculated by evaluating the normalised cumulative density function at the recoil energies of the events. These transformed values are confined to $[0, 1]$. In contrast, recoil energies do not have such a confinement.
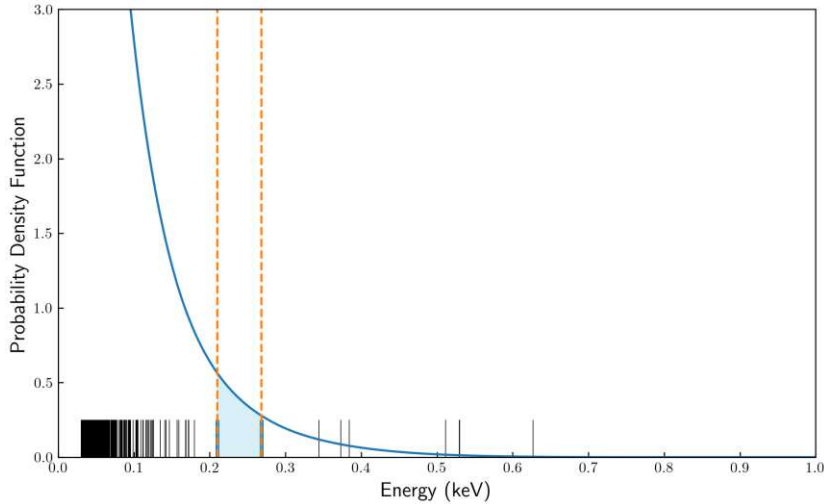
Figure 12: The maximum gap for CRESST-III data in the acceptance region for a dark matter mass of 1 GeV. The normalised expected dark matter probability density function is represented by the blue curve. The black ticks along the horizontal axis are the measured recoil energies in the acceptance region and they represent the possible signal events, known background events, and unknown background events. The maximum gap (43) is depicted as the blue-filled area confined by the energy axis, the expected dark matter signal and the two neighbouring events.

26

Because the dark matter nuclear scattering described in section 2 is dependent on σ, the magnitude of the maximum gap also depends directly on σ. This dependence is crucial in setting the limit. If desired, a certain σ can be chosen which results in a myriad of events in the maximum gap, where none are observed. However, such a large cross section would be experimentally ruled out because it is nearly impossible to detect zero events where a myriad of them are expected unless there is a fundamental flaw with the experimental setting. This feature can be utilised to set an upper limit on σ.

An assumed value of σ is rejected as being too high with confidence level C_0 if the x-values of random experiments consisting of random numbers drawn from a uniform distribution of unit density are lower than the actual maximum gap size with probability C_0. This comparison between the event distribution and a uniform distribution of unit density can be made because of the invariance of x. Thus, C_0 can be defined as the probability that the maximum gap size would be lower than a given value of x and can be analytically calculated with [1]:

$$C_0(x, \mu(\sigma)) = \sum_{k=0}^{m} \frac{(kx - \mu(\sigma))^k \exp^{-kx}}{k!} \left(1 + \frac{k}{\mu(\sigma) - kx}\right) \tag{45}$$

μ is the total number of expected events calculated by integrating the expected dark matter signal in the region of interest and it is dependent on σ. m is the largest integer, for which the condition $m \leq \mu/x$ applies. In order to determine the upper limit, vary σ until the observed x-values and μ are such that C_0 in Eq. (45) reaches the desired confidence level.

3.2 Optimum Interval Method

The optimum interval method is a generalisation of the maximum gap method. In the context of the maximum gap method, it is a fundamental premise that intervals invariably lack the occurrence of any event. The fact that the maximum gap method disregards all other intervals containing one or more events is a choice. The removal of this choice by the algorithmic determination of the most optimal interval not only leads to the removal of this bias but may also result in stronger limits. It is not guaranteed that the interval with zero events will yield the strongest limit, especially when there are many events in question. For instance, in the case of CRESST-III, there are several hundred events in the acceptance region. The optimal interval may contain e.g. 20 events, as depicted in Fig. 13. The algorithm calculates the maximum interval value for each n number of events in an interval, with n going from 0 to N. This results in considering a total of $(N + 1)(N + 2)/2$ intervals. Subsequently, the algorithm selects the most optimal interval that will yield the strongest of upper limits.
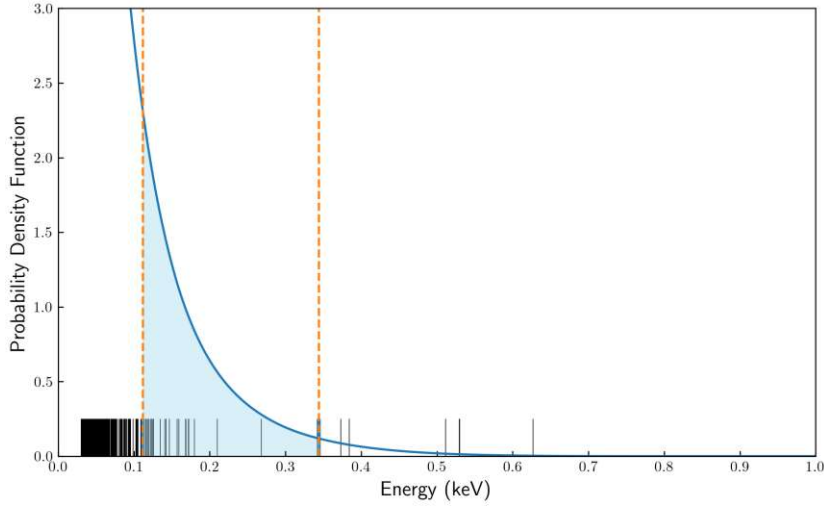
Figure 13: The maximum interval value containing 20 events for CRESST-III data in the acceptance region for a dark matter mass of 1 GeV. The normalised expected dark matter probability density function is represented by the blue curve. The black ticks along the horizontal axis are the measured recoil energies in the acceptance region and they represent the possible signal events, known background events, and unknown background events. The maximum interval value, representing the largest integral of the signal between two events with 20 events in between, is depicted as the blue-filled area confined by the energy axis, the expected dark matter signal and the two neighbouring events.

Define $C_n(x, \mu)$ as the probability that an interval containing $\leq n$ events has an expected number of $\leq x$ events for a given cross section. $n = 0$ represents the case mentioned in the maximum gap method and can be analytically solved. In contrast to C_0, $C_n(x, \mu)$ for $n \geq 1$ cannot be determined analytically and Monte Carlo tabulations need to be performed. Then, once the optimal n has been set by the algorithm, C_n may be used to determine an upper limit in the same manner as C_0: with x equal to the maximum expected number of events taken across all intervals with $\leq n$ events, $C_n(x, \mu)$ is the confidence level at which the assumed cross section is rejected as being too large.

The details of the Monte Carlo tabulations and the setting of upper limit are discussed in section 4.3.

3.3 Approximation of the Optimum Interval Method

As explained in 3.2, for the optimum interval method to be utilised, a lot of tabulations and computational calculations need to be processed. These tabulations and calculations can extend run-times from minutes to hours and even days, depending on tabulation and evaluation parameters set. If a researcher, however, is interested in quick estimates of the upper limits a certain signal model yields, the approximation is much more useful and time-saving.

The approximation of the optimum interval method implemented in this work relies on two key assumptions. The first assumption is that the tabulated data for a certain amount of the expected number of events can be approximated by pre-calculated tabulated data of two neighbouring μ. The second assumption, namely that the distribution of the tabulated data is roughly Gaussian distributed complements the first one and makes predictions feasible for μ exceeding that of pre-calculated tabulated data. However, it is important to remember that these assumptions have their limits and could affect how accurate the results are.

A more detailed explanation of the implementation can be found in section 4.4. The advantages and disadvantages of this approximation are described in section 5.4.

3.4 Maximum Patch Method

The one-dimensional Yellin methods discussed in sections 3.1 and 3.2 have a significant drawback when applied to CRESST data analysis. This drawback arises from their inability to consider the two-dimensional nature of CRESST data, which involves both energy and light yield values. These methods focus solely on event energies and disregard the light yield information once an event is classified as either within or outside the pre-defined acceptance region.

This drawback means that these methods cannot distinguish whether the light yields of observed events match the expected characteristics of a signal. This issue becomes evident when considering the background events from the e^-/γ. They would be treated the same as events located within the nuclear bands at the same energy levels, even though the latter are more likely to be genuine signal events.

Hence, as already briefly discussed in section 2.2, there is a clear need for a more comprehensive approach to the data analysis of dark matter experiments that takes into account both energy and light yield parameters. This approach would enable a more refined differentiation between potential signals and background events. This is exactly where the maximum patch method comes in.

The working principle of this method is similar to the principle of the maximum gap method, plus one dimension. The description of the maximum patch method in this section resembles Jens Schmaler's more in-depth discussion in [39]. The

29

maximum gap can be thought of as the largest difference between two neighbouring cumulative density function values corresponding to the energies of the measured events. The range of a cumulative density function is by definition between 0 and 1. The energy values E of the data are thus converted into new values x_1

$$E \to x_1 \tag{46}$$

with $0 \leq x_1 \leq 1$.

The same logic can be applied to the two-dimensional density function in Eq. (39) through a coordinate transformation, converting energy-light yield values to newly defined x_1 and x_2 values

$$(E, LY) \to (x_1, x_2) \tag{47}$$

with $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 1$. The transformed distribution function's value then corresponds to the expected number of events

$$\rho(x_1, x_2) = \mu \tag{48}$$

where μ denotes again the expected number of events. Since the expected number of events for a certain experiment is not dependent on other variables, the statement

$$\rho(x_1, x_2) = const. \tag{49}$$

is valid. Yellin shows a possible way of transforming the multidimensional density function in [2]. Since the density function is two-dimensional, two quantities shall be defined,

$$\rho_1(E) := \int_{-\infty}^{+\infty} dLY' \rho(E, LY') \tag{50}$$

and

$$\rho_0 := \int_{-\infty}^{+\infty} dE' \rho_1(E). \tag{51}$$

Since ρ_0 is calculated by simply integrating the two-dimensional density function along both of its axes, ρ_0 equals μ. Using these two ρ_i quantities, the transformed coordinates x_1 and x_2 can now be calculated using the following relations:

$$x_1(E) := \frac{1}{\rho_0} \int_{-\infty}^{E} dE' \rho_1(E') \tag{52}$$

and

$$x_2(E, LY) := \frac{1}{\rho_1(E)} \int_{-\infty}^{LY} dLY' \rho(E, LY'). \tag{53}$$

30

Using the coordinate transformation described above the two-dimensional data provided by an experiment searching for dark matter can be transformed into x_1 and x_2 values. Just as in the maximum gap method, the next step is identifying the maximum patch, the largest rectangle with no data points inside it. Analogous to Eq. (43), the size of these patches is expressed as

$$x_i := \int_{x_{2,1,i}}^{x_{2,2,i}} \int_{x_{1,1,i}}^{x_{1,2,i}} \rho(x_1, x_2) dx_1 dx_2. \tag{54}$$

Because the density function is constant, see Eq. (49), Eq. (54) can be reformulated into

$$x_i := \mu \int_{x_{2,1,i}}^{x_{2,2,i}} \int_{x_{1,1,i}}^{x_{1,2,i}} dx_1 dx_2. \tag{55}$$

The integral part of Eq. (55) represents the geometrical area A in the x_1-x_2 plane, which leads to

$$x_i = \mu \cdot A_i. \tag{56}$$

The maximum patch is then the rectangle with the largest area of all possible rectangles in the x_1-x_2 plane

$$x = \mu \cdot \max\{A_i : i = 0, 1, \ldots, n_i\} \tag{57}$$

with $A = x/\mu$. The fact that μ, which can be calculated by the differential rate of the dark matter scattering described in section 2, is dependent on the normalised WIMP-nucleon cross section σ_{WN} results in maximum patch x also being dependent on σ_{WN}. Exactly this dependency is needed for setting the upper limit using the maximum patch method. Similarly, $C_0(x, \mu)$ represents the likelihood that, when considering an expected number of events μ in the experiment, the observed maximum patch size is less than or equal to x. In contrast to the maximum gap method, this time the analogous function $C_0(x, \mu)$ cannot be expressed analytically, but is expressed through a Poisson process

$$C_0(x, \mu) = \sum_{N=0}^{\infty} C_0'(A, N) \left(\frac{\mu^N}{N!} e^{-\mu} \right) \tag{58}$$

where $C_0'(A, N)$ represents the probability that the maximum patch size is less or equal to x, given N observed events. The sum in Eq. (58) can be truncated at a certain N-value, after which the absolute values of the summands become negligibly small. In order to determine this probability a Monte Carlo simulation is required, which is explained in detail in section 4.5.

Finally, as is the procedure with the maximum gap method, to find the upper limit on the cross section with a certain confidence level, for example, 90% confidence level, identify the value of σ_{WN} at which the function $C_0(x, \mu)$ (58) equals 0.9.

4 Implementation

In this work, the expected signal of WIMPs in the standard scenario and Yellin's analysis methods have been implemented in a package based on the programming language Python. Initially, there were two separate classes, one for modelling the signal, namely *SignalModel*, and one for analysing the data with Yellin's methods and setting upper limits on the cross section, namely *Limit*. Later a third class was created, the *ModeLimit*, which inherited both of these classes and combined their workflow, resulting in a more user-friendly experience. The code mainly relies on two fundamental libraries to accelerate the run-time and make the code run smoothly: NumPy and SciPy [48, 49]. NumPy helps in processing large amounts of data stored in multidimensional arrays efficiently. SciPy makes it possible to utilise fundamental mathematical operations like integrating discrete functions and defining standard functions like normal distributions, error functions and natural constants. Ultimately, the library Matplotlib is used to provide most of the plots in this study [50].

A closer look at the CRESST-III and CRESST-II data is necessary before explaining what the package classes do. CRESST-III and CRESST-II data are used for the calculation of upper limits on the cross sections of dark matter using the maximum gap and optimum interval methods. Additionally, the maximum patch method uses two-dimensional simulated data that resembles that of CRESST-III.

4.1 Description of CRESST Data

Certain detector-related information is provided in the descriptions of CRESST-III and CRESST-II data [26, 51].

Starting with CRESST-III, the light yield graphs of this detector have already been briefly discussed in Fig. 9 and in Fig. 10. The energy range is restricted to events below 16 keV, which determines the upper integral limit, E_{Upper}, for calculating the expected number of events μ. Exposure before data selection is 5.594 kg d. In the context of the CRESST experiment, adjustments to the signal model are imperative to account for various experimental factors, ensuring its comparability to the energy distribution observed by the detector. The corrected signal model, denoted as $\frac{d\widetilde{\Gamma}}{dE_R}(E_R)$, incorporates considerations for finite energy resolution, energy threshold, and the probability of survival after imposed cuts. This model is formulated through an integral involving the convolution of the differential rate $\frac{d\Gamma}{dE_R}(E)$ with a normal distribution $\mathcal{N}(E_R - E, \sigma_{Res}^2)$, where σ_{Res} represents the resolution of the phonon detector:

$$\frac{d\widetilde{\Gamma}}{dE_R}(E_R) = \Theta(E_R - E_{thr}) \cdot \varepsilon_{Cut} \cdot \varepsilon_{AR} \cdot \int_0^\infty \frac{d\Gamma}{dE_R}(E) \cdot \mathcal{N}(E_R - E, \sigma_{Res}^2)dE \quad (59)$$

The convolution term for the finite energy resolution of the experiment. Due to the convolution term, events with energies just under the threshold energy are also taken into consideration. Therefore, it is crucial for the integral to begin at 0 rather than the threshold energy. The integral can be truncated at the upper limit of the region of interest, 16 keV. Not every event survives the data selection criteria, hence the convoluted differential rate is multiplied by the probability of surviving ε_{Cut}. Additionally, the likelihood of a signal event falling within the acceptance region is represented by ε_{AR}. Lastly, the condition for an event triggering a detection in reconstructed energy sets a sharp threshold, represented with a Heaviside step function $\Theta(E_R - E_{thr})$, with E_{thr} specified as 0.0301 keV. For CRESST-III, the resolution σ_{Res} is 0.0046 keV. ε_{Cut} and the material-specific ε_{AR} are provided by the CRESST Collaboration [26]. The histograms of the data can be seen in Fig. 14.
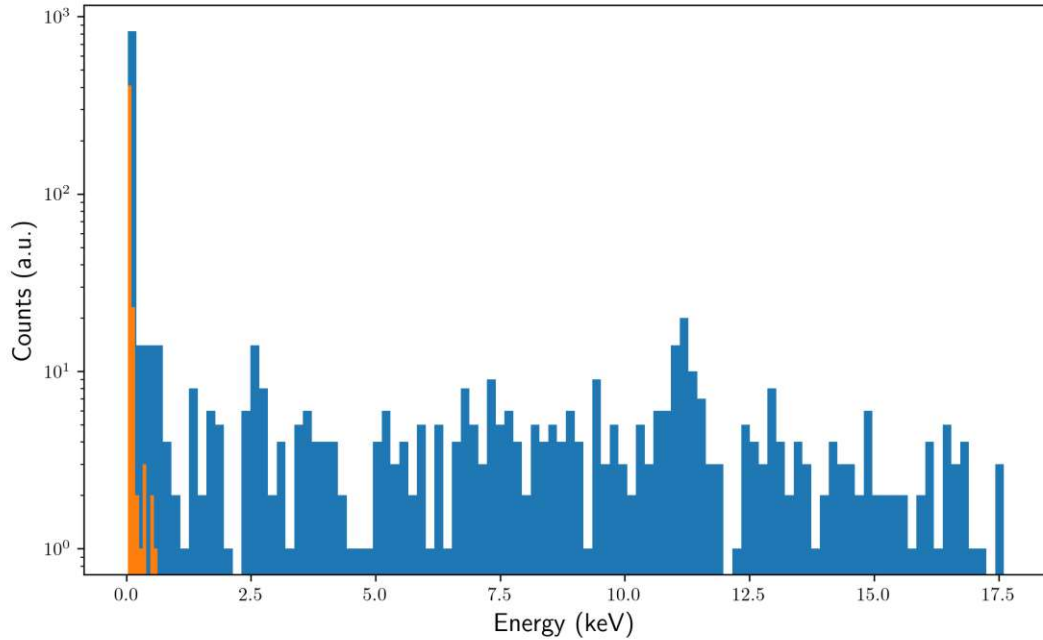
Figure 14: Blue: Energies in keV for all events surviving data selection in CRESST-III. Orange: Energies in keV for all events in the acceptance region for the CRESST-III dark matter search, as in Fig. 10.

CRESST-II Lise data is described in a similar fashion with varying parameter values. The energy range is restricted to events below 40 keV and the exposure before data selection is 52.15 kg d. E_{thr} is taken as 0.307 keV and σ_{Res} as 0.062 keV. Furthermore, ε_{Cut} and the material-specific ε_{AR} are provided in the form of a single file combining these two functions.

33

Depicted in Fig. 15 is the simulated data intended for the application of the maximum patch method. The energy range considered is limited to events below 16 keV, and the initial exposure before data selection stands at 2.87 kg d. Parameters include E_{thr} set to 0.025 keV and σ_{Res} to 0.005 keV. It's noteworthy that both ε_{Cut} and the material-specific ε_{AR} align with those employed in the description of CRESST-III data.
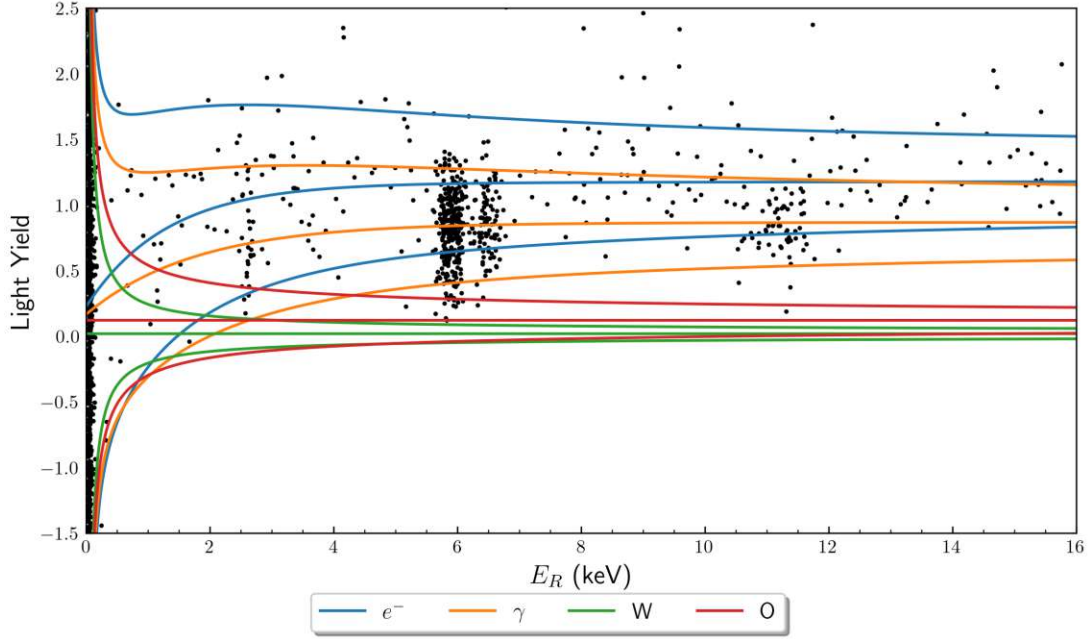
Figure 15: The light yield versus the energy of events in the dark matter data set, following the application of selection criteria. The 90% upper/lower bounds and the mean lines of the e^--band, γ-band, tungsten band and oxygen band are depicted by the blue, orange, green and red lines, respectively.

4.2 Modelling the Expected Signal

The class *SignalModel* performs the necessary calculations needed for the formulae described in section 2 using the Python packages NumPy and SciPy. For the one-dimensional case, the calculations are performed using a one-dimensional NumPy array. For the two-dimensional case, this array takes the form of a two-dimensional matrix. When utilising *ModeLimit*, firstly, the instance variables of the class are initialised, which encompass most of the aforementioned constant values like the dark matter density ρ_χ, proton mass m_p, galactic escape velocity v_{esc} etc. The step size of the energy grid can be modified by the user. When setting the step size, exercise caution to choose a value significantly smaller than the detector's

resolution, preferably five times smaller if feasible. The default value of 0.0005 keV has been found suitable for analysing the CRESST data. After the initialisation, the user sets the detector parameters. In the next step, the differential rates of given materials as in Eq. (17) and the probability density functions (PDFs) are calculated.

To accelerate the process and not compute unnecessary calculations, calculating until which energy value the integral expressed in Eq. (24) returns non-zero values is beneficial. Hence, the inequation $x_{min} \geq (z + \eta)$ dictates the upper bound of the energy grid, with x_{min}, z and η defined in Eqs. (23) and (25). Subsequently, a *for loop* is initiated, iterating over the various materials. For each material, the differential rate in (kg d keV pb)$^{-1}$ is calculated as described in Eq. (59). After the loop, the final PDF can be calculated by the weighted sum of material-specific differential rates.

Furthermore, the *SignalModel* package has methods for calculating the corresponding CDFs of the PDFs, drawing samples using the inversion method and computing the number of expected events μ by integrating over Eq. (59):

$$\mu = \int_{\text{Threshold}}^{\text{Upper limit of ROI}} dE_R \frac{d\widetilde{\Gamma}}{dE_R}(E_R) \cdot \sigma \cdot \text{exposure} \qquad (60)$$

In order to calculate the two-dimensional density function, the one-dimensional PDF can be made use of, with one modification - the PDF is not multiplied with acceptance region efficiencies[1]. Henceforth, the calculation of the two-dimensional density function can be accomplished by multiplying the one-dimensional PDF in Eq. (59) with the two-dimensional Gaussian distribution as expressed in Eq. (39). The number of expected events μ can then be calculated with

$$\mu = \int_{\text{Lower LY limit}}^{\text{Upper LY limit}} \int_{\text{Energy threshold}}^{\text{Upper limit of ROI}} \rho(E_R, LY) dE_R dLY \cdot \sigma \cdot \text{exposure}. \qquad (61)$$

Note that in contrast to the one-dimensional scenario, we set ε_{AR} to one.

---

[1]The idea behind the two-dimensional density function and the maximum patch method is making use of both the measured energy and the light yield of events. In contrast to the one-dimensional PDF, the two-dimensional density function will be applied to two-dimensional data and not one. Hence, the need for multiplying event data with acceptance region efficiencies vanishes. A pre-selection on the two-dimensional data can still be applied for excluding mainly e^-/γ events.

4.3 Utilisation of the Optimum Interval Method

In this section, the implementation of Yellin's optimum interval method will be discussed, detailing the process of tabulating and computing the necessary data required for determining upper cross section limits. Note that not only the tabulated data but also the experimental data will be utilised and processed for these limit calculations.

Starting with the calculations needed for the optimum interval method, the first step is calculating the x-values, defined in Eqs. (43) and (44), of the detected experimental data with M values for all k-largest intervals with $k \in \{0, M-1\}$. The resulting values are $x_1, x_2, \ldots, x_{M-1}$. Note that for calculating the x-values the normalised cumulative density function derived from the differential rate is used, as in (44). This normalisation allows for the tabulated data, which will be between 0 and 1, to be comparable with experimental data transformed, which will also be between 0 and 1 due to the properties of the corresponding normalised CDF.

The goal now is to determine which k and corresponding x_k is best suited to set a limit on the cross section σ. The best k is the one which provides the most information on how compatible/incompatible the experimental data and the model are. The "extremeness" value will be used as a measure of this information, which will be defined shortly. A comparison will be made regarding the degree of "extremeness" or rarity of the x_k value from the experimental data against a distribution of x_k values from the Monte Carlo generated data. Extreme in the sense of being too out of the ordinary a value as discussed in section 3.2. These extremeness values will help us set the limits on cross sections which would predict too many events in an interval where none are observed.

The next step is the generation of the tabulated data sets which will be compared to experimental data and intervals. These data sets consist of N data sets of length n drawn from a continuous uniform distribution on $[0, 1]$ for each possible $\mu(\sigma)$ of interest. The sizes of these N arrays are not constant but are Poisson-distributed with a mean value of μ, $Pois(\mu) = n$. This accounts for the fact, that the expected number of events $\mu(\sigma)$ is seldom an integer value. The calculation of the x-values for all k-largest intervals for each data set is now performed, mirroring the approach used for the experimental data. Since the data is drawn from a uniform distribution, the x-values are simply the differences between the two values. Consequently, distributions of x-values are obtained for each μ and k, see Tab. 1.

for each μ				
Data sets	$k = 0$	$k = 1$	$k = 2$	$\ldots$
1	$x_{0,1}$	$x_{1,1}$	$x_{2,1}$	$\ldots$
2	$x_{0,2}$	$x_{1,2}$	$x_{2,2}$	$\ldots$
3	$x_{0,3}$	$x_{1,3}$	$x_{2,3}$	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$x_{0,N}$	$x_{1,N}$	$x_{2,N}$	$\ldots$

Table 1: Table with x-values for the Monte Carlo generated N data sets with sizes n for all k-intervals. Such a table exists for each μ. The first indices of the x-values represent the corresponding k-value, indicating how many other elements are allowed between two elements. The second indices of the x-values represent the data set index that the x-values are a part of.
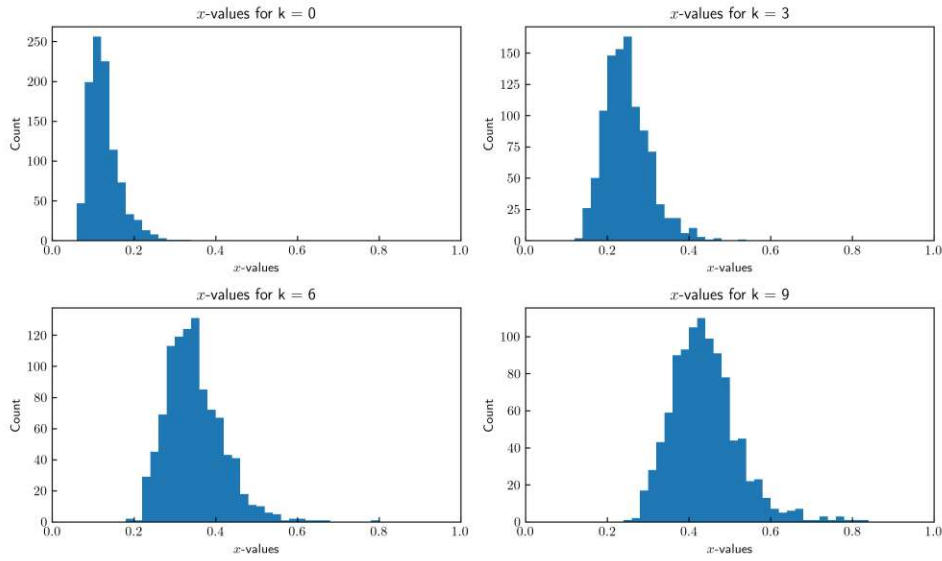
Figure 16: Histogram of the x-values for different k-values, i.e. for different columns in table 1, for an arbitrary μ.

After the generation of all data sets, the above-mentioned "extremeness", de-noted as $\Gamma_{k,N}$, can be calculated for each $x_{k,N}$ by determining the percentile of each $x_{k,N}$ in the distribution along the corresponding column in Tab. 1, which then results in Tab. 2:

for each μ					
Data sets	$k = 0$	$k = 1$	$k = 2$	$\ldots$	Γ_{Max}
1	$\Gamma_{0,1}$	$\Gamma_{1,1}$	$\Gamma_{2,1}$	$\ldots$	$\Gamma_{Max,1}$
2	$\Gamma_{0,2}$	$\Gamma_{1,2}$	$\Gamma_{2,2}$	$\ldots$	$\Gamma_{Max,2}$
3	$\Gamma_{0,3}$	$\Gamma_{1,3}$	$\Gamma_{2,3}$	$\ldots$	$\Gamma_{Max,3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$\Gamma_{0,N}$	$\Gamma_{1,N}$	$\Gamma_{2,N}$	$\ldots$	$\Gamma_{Max,N}$

Table 2: Table with Γ-values for the Monte Carlo generated N data sets with sizes of n for all k-intervals. Such a table exists for each μ. The first indices of the Γ-values represent the corresponding k-value, indicating how many other elements are allowed between two elements. The second indices of the Γ-values represent the data set index that the x-values are a part of.
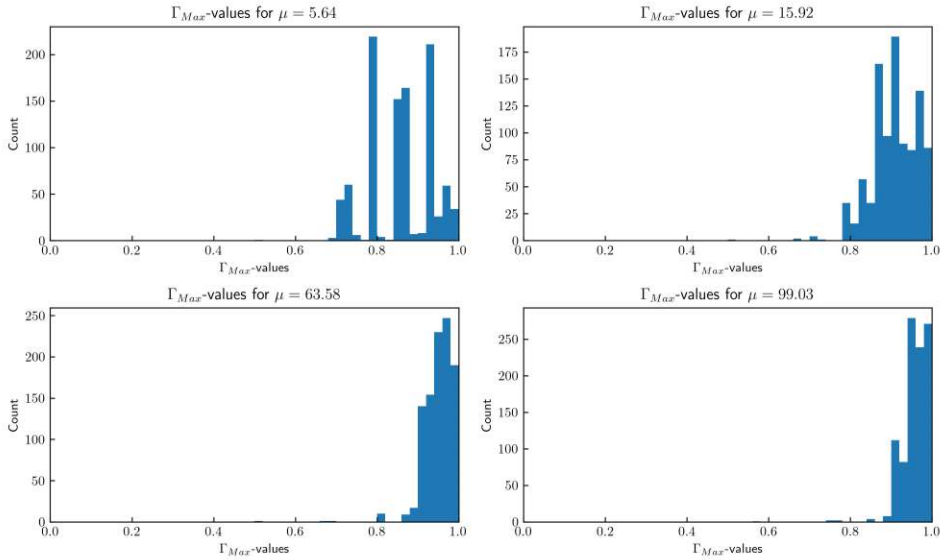
Figure 17: Histogram of the Γ_{Max}-values for different μ, i.e. histogram of the last column in Tab. 2, for various μ.

Using Tab. 2, it is now possible to determine the largest Γ-value along each row. The determination involves finding the largest Γ-value per data set, resulting in a total of N Γ_{Max}-values for each μ:

$$\Gamma_{Max,i} = \max\{\Gamma_{k,i} : k = 0, 1, \ldots, n_i\}, \ i = 0, 1, \ldots, N \tag{62}$$

38

Coming back to the experimental data, using the distributions of the x-values from the generated data, depicted in Fig. 16, the extremeness values of the experimental data can be determined for all k-intervals. The extremeness values of the experimental data are denoted as C_k, not as Γ_k. In a manner akin to the definition of Γ_{Max} for the Monte Carlo generated data, $C_{Max}(\mu)$ can be defined as

$$C_{Max}(\mu) = \max\{C_k : k = 0, 1, \ldots, M - 1\}. \tag{63}$$
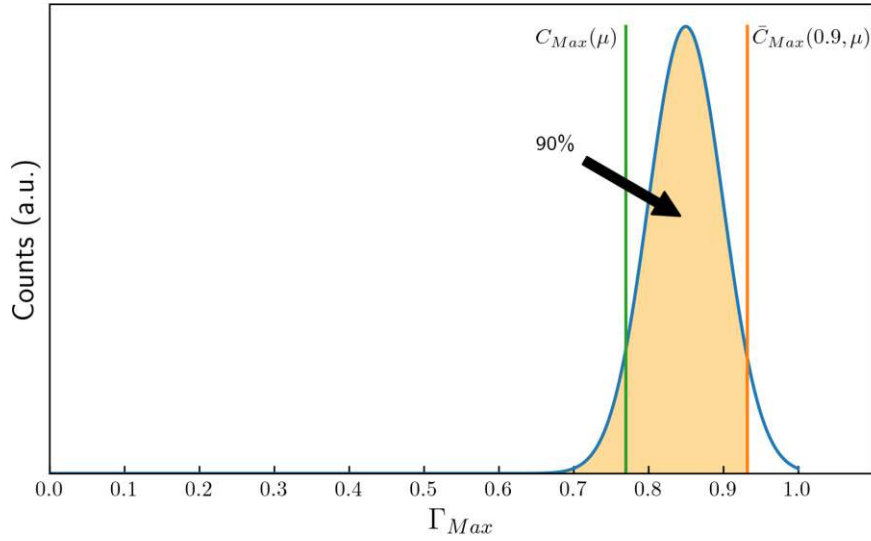
Figure 18: Illustration showing $C_{Max}(\mu)$ and $\bar{C}_{Max}(0.9, \mu(\sigma))$. There is one such graph for each dark matter mass. The blue line represents the exemplary distribution of $C_{Max}(\mu)$ values. The green line represents the C_{Max} value of an exemplary μ. The orange line represents the $\bar{C}_{Max}(0.9, \mu(\sigma))$ which is larger than 90% of the $C_{Max}(\mu)$ values.

For each μ, the percentile of $C_{Max}(\mu)$ in the corresponding distribution of Γ_{Max} can be determined, and for each confidence level, CL, and $\mu(\sigma)$, there exists a $\bar{C}_{Max}(\text{CL}, \mu(\sigma))$, which is the CL-percentile in this distribution. Assuming a confidence level of 90%, the $\bar{C}_{Max}(0.9, \mu(\sigma))$ would be larger than exactly 90% of the Γ_{Max}-values in the respective μ. Hence, the limit on $\mu(\sigma)$, and consequently the limit on σ, is given for the $\bar{\mu}$ for which the equation

$$C_{Max}(\bar{\mu}) = \bar{C}_{Max}(\text{CL}, \bar{\mu}) \tag{64}$$

holds. Because of the inherent randomness of the tabulated data, a filter is applied to these values, as in Fig. 19. The first value to exceed 0.9 thus determines $\bar{\mu}$.

39

The value, that such fraction of C of random experiments with the given μ and no unknown background will give $C_{Max} < \bar{C}_{Max}(C, \mu)$, defines the function $\bar{C}_{Max}(C, \mu)$. As a result, the cross section's upper limit at 90% confidence level is represented by the experiment's C_{Max} equalling $\bar{C}_{Max}(0.9, \mu)$, which is depicted in Fig. 19. $\bar{C}_{Max}(0.9, \mu)$, the C_{Max} value at which the 90% confidence level is attained, as a function of μ is shown in Fig. 19. The $\bar{C}_{Max}(0.9, \mu)$-values increase with increasing μ-values. By observing the pattern of Γ_{Max}-values for increasing μ-values, the rise of the $\bar{C}_{Max}(0.9, \mu)$ function is expected.
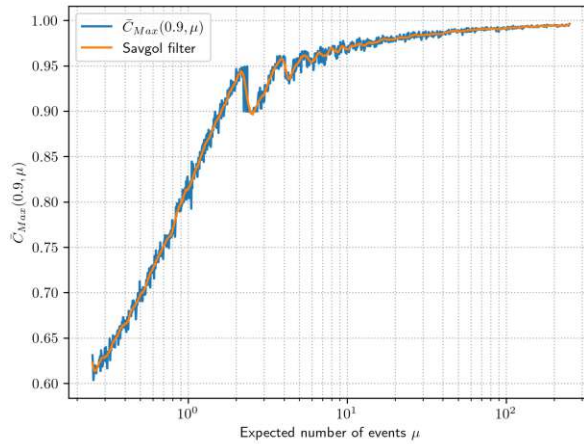
Figure 19: Plot showing $\bar{C}_{Max}(0.9, \mu)$, the C_{Max} value at which the 90% confidence level is attained, as a function of the total number of expected events μ in the experimental range. The orange line is attained by smoothing the results using the Savgol filter. The usage of the filter is for visual purposes only. The blue line is jagged due to the inherent randomness of generated data.

4.4   Approximating the Optimum Interval Method

As discussed in section 3.3, the approximation relies on two major assumptions.

1. Between the minimum and maximum μ in the pre-calculated data, the tabulated data can be approximated by the pre-tabulated data of the neighbouring μ-values.

2. For an expected number of events exceeding the maximum μ in the pre-calculated data, the tabulated data can be approximated by an extrapolation under the assumption that the tabulated data is Gaussian distributed.

40

In this work, the $\mu_{min} = 0.25$ and $\mu_{max} = 1000$ for the pre-calculated data. A total of 1000 expected numbers of events are geometrically spaced between these two values, including these two. The number of arrays for each μ is set to $N = 750$. The next step is calculating the distributions of x for each μ and k-values, and using the histograms depicted in Fig. 16 to calculate the corresponding CDFs and save these CDFs locally. A similar approach is then applied to histograms of Γ_{Max}-values depicted in Fig. 17, but instead of calculating the CDFs and then saving them, the whole histograms are saved by their unique values in pre-defined bins and unique counts.

Two scenarios come to mind when utilising this approximation for the tabulated data:

1. In the first case, the user makes use of the same μ-values that are also used for the pre-calculated and pre-tabulated data. In that case, the only errors in play would arise due to the finite binning choices. The CDF of x-distributions for example is defined by 1000 data points between 0 and 1. The more data points there are between 0 and 1, the better the approximation would be. However, increasing the number of data points would also mean more storage space and run-time. In that case, the approximation loses its value because it is exactly developed for getting estimates in a short amount of time without demanding a lot of computational resources.

2. In the second case, the user makes use of an arbitrary expected number of events that are not a part of the μ used for the pre-calculated and pre-tabulated data. In this case, not only do the errors of the previous case come into play, but new error sources arise due to approximating the tabulated data for these certain μ-values using the tabulated data of neighbouring μ-values. The weighted averages of the CDFs of x-distributions and histograms of Γ_{Max}-values are calculated with weights depending linearly on the distance of these neighbours in μ-space to each other.

Under the conditions of the first scenario, the newly tabulated and approximated x- and Γ_{Max}-distributions for μ values can be seen in Fig. 20. As anticipated, the centres of the x-distributions tend to shift to smaller numbers for increasing μ-values. The variance in distribution also sinks. Increasing the k-value results in larger values, because more events are allowed to be between the values subtracted from each other. This effect is also observed in Fig. 16. The x-distributions show significant similarities in their behaviour. The Γ_{Max}-distributions, on the other hand, do not seem to overlap to the same extent as the x-distributions. This is true, especially for smaller μ-values. Note, that this is the case where the newly tabulated data is tabulated for the same μ-values as

41

the pre-tabulated data was tabulated for. Thus, even for two identical values of μ, two sets of Monte Carlo experiments yield slightly different Γ_{Max}-values.
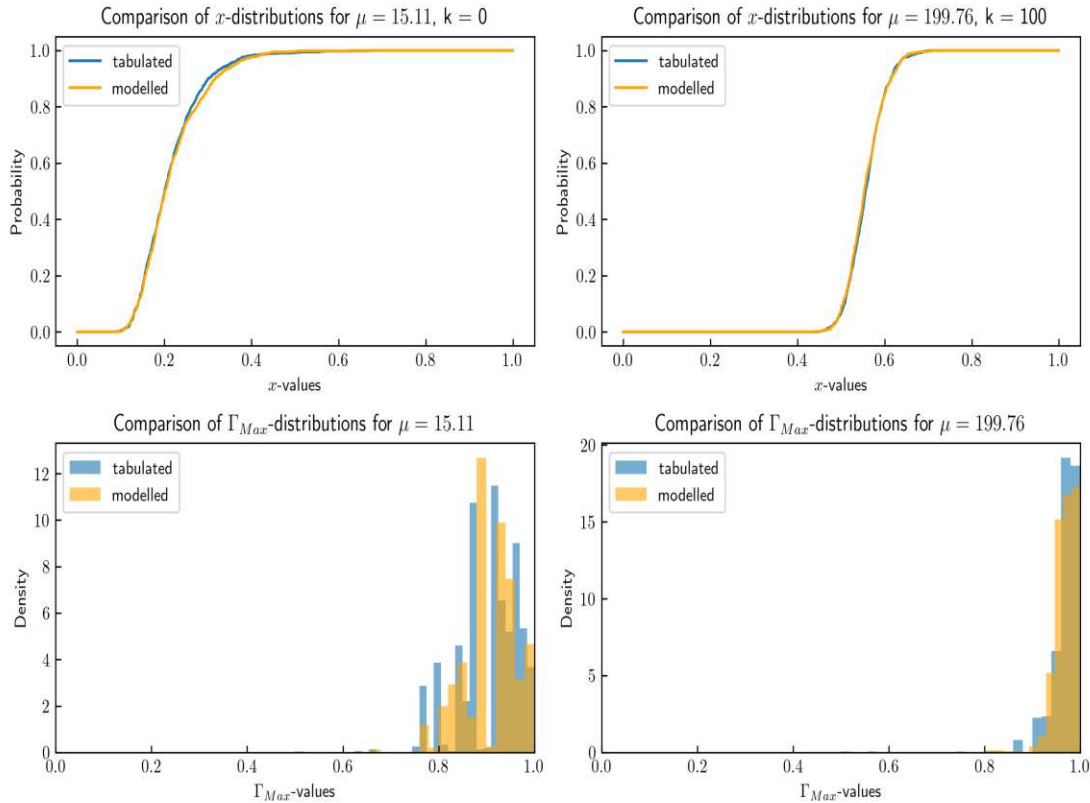
Figure 20: Top left plot depicts the comparison of x-distributions for a given μ and k, in this case $\mu = 15.11$ and $k = 0$. The bottom left depicts the comparison of Γ_{Max}-values for the same μ. The plots on the right side depict the same comparisons but for different μ and k, namely $\mu = 199.76$ and $k = 100$. In this figure, the μ-values of the newly tabulated and pre-tabulated data coincide.

Under the conditions of the second scenario, where the μ-values of the tabulated and pre-tabulated data do not share any values, the error would be expected to rise due to the nature of the approximation. However, Fig. 21 shows no apparent increase in the error of the estimations. The CDF graphs of the x-distributions still show significant overlap and the histograms of Γ_{Max} show, at least for high μ-values, a considerable amount of overlap. For low μ-values the overlap does not seem to be noteworthy, but that is not the case even when the μ-values are shared as in Fig. 20, therefore, to expect such overlap now would not be correct because Monte Carlo experiments by nature tend to result in different Γ_{Max} distributions, even if all the starting parameters and conditions are same.
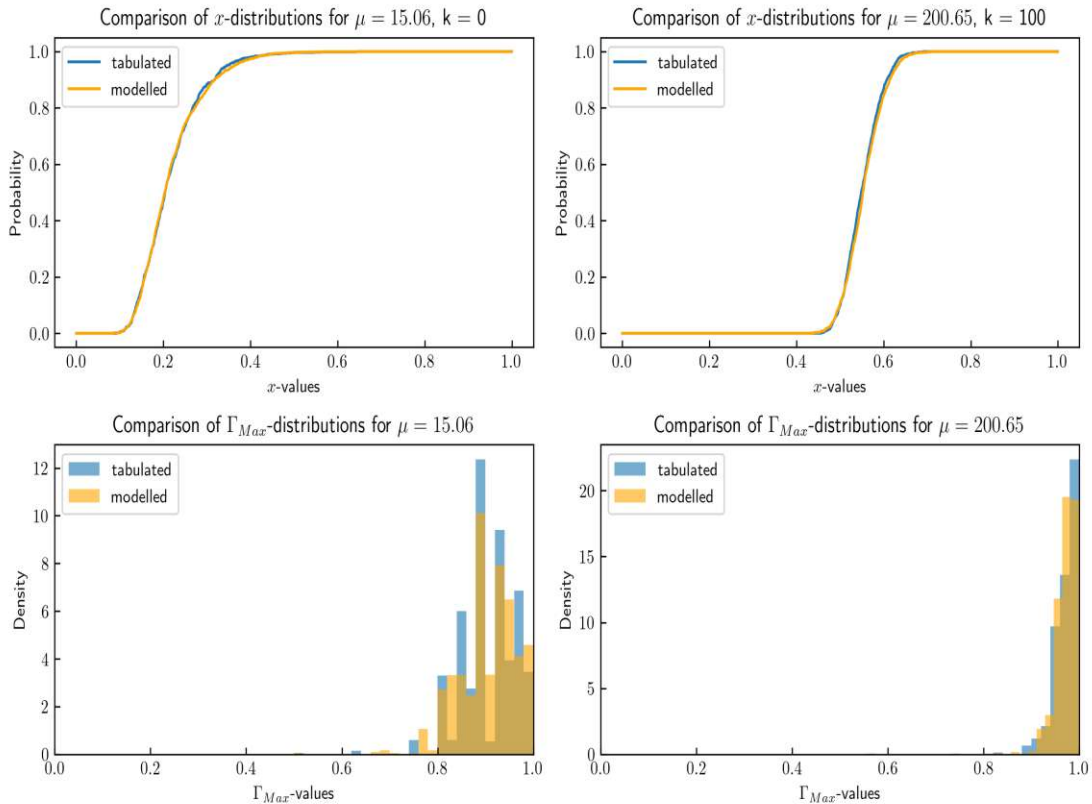
Figure 21: Top left plot depicts the comparison of x-distributions for a given μ and k, in this case $\mu = 15.06$ and $k = 0$. The bottom left depicts the comparison of Γ_{Max}-values for the same μ. The plots on the right side depict the same comparisons but for different μ and k, namely $\mu = 200.65$ and $k = 100$. In this case, the μ-values of the newly tabulated and pre-tabulated data do not coincide.

If the maximum number of expected events exceeds the maximum μ, for which data was pre-tabulated, the second assumption starts playing a role. From the plots depicted in Fig. 20 and 21 it can be inferred that this assumption is only acceptable under certain conditions, e. g. for x-distributions when the μ-values are large and the k-values are small in comparison, and for the Γ_{Max} distributions in general for high μ-values. For x-distributions the error tends to grow with increasing k-values, as can be seen in Fig. 22. If the optimum interval is not found for small k-values, the estimation error will grow.

The data used for the approximation can be modified by the user. The maximum μ value for the pre-tabulated data is already set to a high value, but if the user reckons that this μ_{max} will not be sufficient, the user may choose to create a new pre-tabulated data. By setting the maximum μ of the pre-tabulated data

43

high, the second scenario as a whole can be avoided to ensure better results.
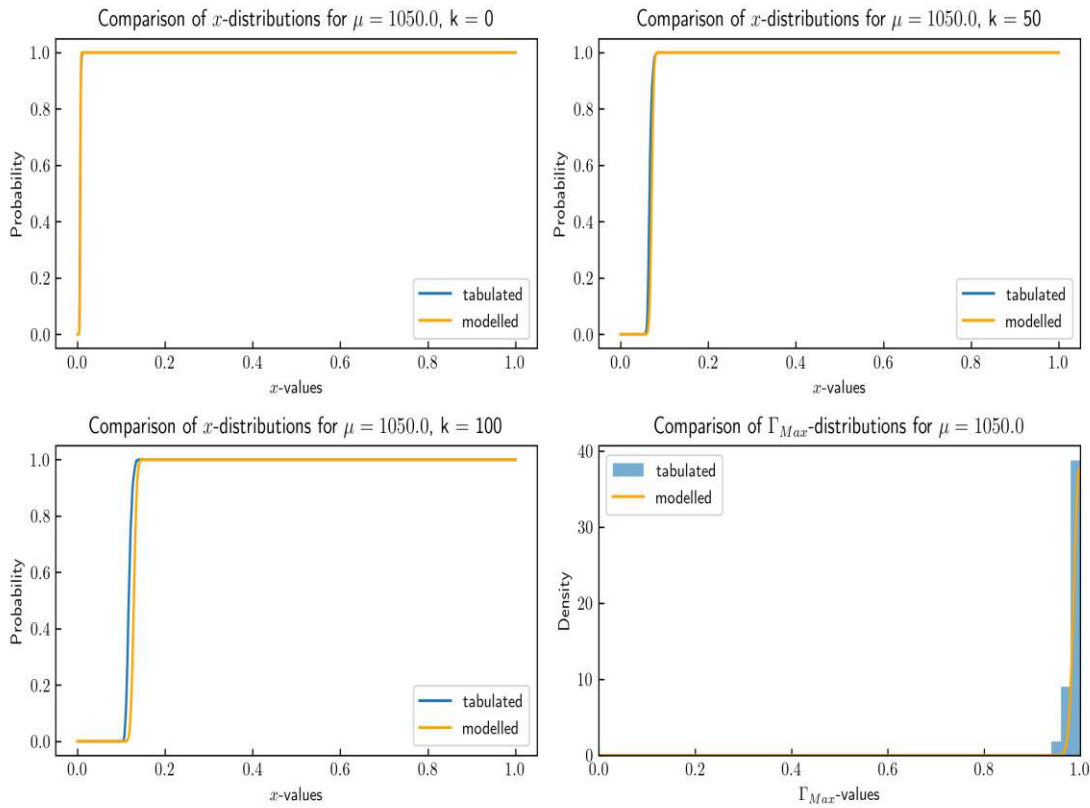
Figure 22: Top left plot depicts the comparison of x-distributions for an out of bounds μ-value of 1050 and $k = 0$. The top right and bottom left plots depict the same comparison for different k-values, namely 50 and 100. The bottom left depicts the comparison of Γ_{Max}-values for the same μ. In this case, the μ-values of the newly tabulated and pre-tabulated data do not coincide and the maximum number of expected events in the newly tabulated data0 exceeds that of the pre-tabulated data.

The extrapolation of μ-values is done by fitting Gaussian functions to the distributions and saving the mean and standard deviation values into a dataframe. These are then used to predict the mean and standard deviation values of the distributions for the μ-values out of bounds by utilising a similar logic with distance calculation to the neighbouring μ-value as is estimated for values between two neighbouring μ-values.

The script for creating the pre-tabulated data needed for the approximation will be included in the script, so that certain modifications, such as changing the μ-values, can be applied to the pre-tabulated data.

44

4.5 Data Tabulation for the Maximum Patch Method

The foundational working principles of the maximum patch method are explained in section 3.4. This section will concentrate on the algorithm for identifying the largest rectangles with no data points inside them and on the distributions of the rectangle sizes, which is necessary for calculating $C_0'(A, N)$ in Eq. (58). A more in-depth algorithm can be found in [39], which also involves finding rectangles with a set number of data points in it. Alg. 1 is a simplified and optimised algorithm for maximum patch purposes only and can be utilised to calculate the distributions for maximum patch sizes:

Algorithm 1 Maximum Patch Algorithm

  for N in $[1, N_{max}]$ **do**
    Create empty list A_{Max}
    for N_{MC} in $[1, N_{MC_{Max}}]$ **do**
      $L_r \leftarrow$ N random pairs of numbers in a x_1-x_2 space with $0 \leq \{x_1, x_2\} \leq 1$
      Append $(0, 0)$ and $(1, 1)$ to L_r
      Sort L_r by the first coordinate x_1
      $A_r \leftarrow$ Rectangle areas $(x_{1,i+1} - x_{1,i}) \cdot (1 - 0)$ calculated by using the pairs from L_r
      for i in $[0, N + 1]$ **do**
        for j in $[i + 2, N + 2]$ **do**
          $x_{ji,diff} \leftarrow x_{1,j} - x_{1,i}$
          $S_r \leftarrow$ Sublist of L_r between the indices $i + 1$ and $j - 1$, including these two
          Append $(0, 0)$ and $(1, 1)$ to S_r
          Sort S_r by the second coordinate x_2
          $A_s \leftarrow$ Rectangle areas $x_{ji,diff} \cdot (x_{2,k+1} - x_{2,k})$ calculated by using the pairs from S_r
          Append A_s to A_r
        end for
      end for
      Append largest value in A_r to A_{Max}
    end for
    Through the distribution of areas, calculate the cumulative density functions representing the areas being less or equal to certain values needed for Eq. (58)
  end for

Note that $N_{\max}$ theoretically needs to be infinitely large according to the sum in Eq. (58). However, given that the individual terms within the sum diminish as

N increases, truncating the sum at a sufficiently large N proves adequate. In the context of this study, N_{max} is set to 516.

In Fig. 23 the rectangles with the largest areas not containing any points can be observed for two different N values, $N = 25$ and $N = 150$. As expected, the largest rectangles tend to shrink with increasing N values because the area becomes more crowded and finding large areas devoid of any points becomes a challenge.
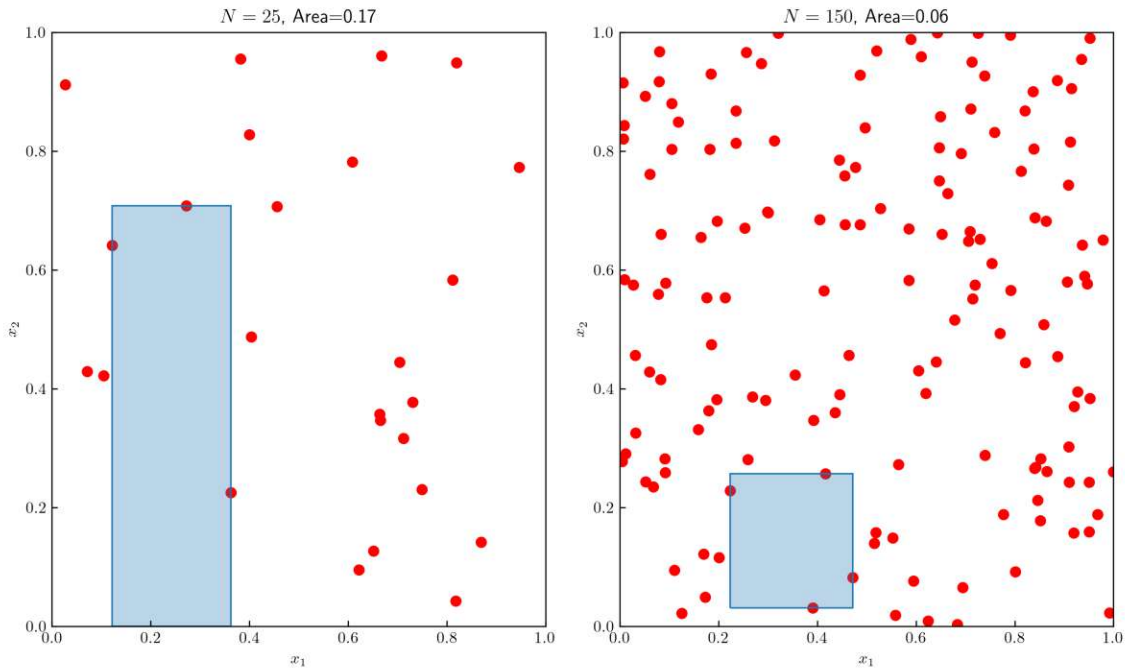
Figure 23: Two plots of randomly generated points and the corresponding maximum patches for different N values. The plot on the left depicts 25 points representing the randomly generated data in Alg. 1, and the plot on the right depicts 150 such points. The blue rectangles represent the maximum patches for these Monte Carlo experiments. The area of the maximum patch is 0.29 for $N = 25$ and 0.06 for $N = 150$.

Fig. 24 shows the last step in Alg. 1, where the cumulative density function is calculated using the distribution of areas. For this plot, a total of 10,000 Monte Carlo experiments were performed. The amount of Monte Carlo experiments here is crucial. If the number is too low, the CDF might be jagged which might lead to an important issue. It is known that with each random point added to x_1-x_2 space, the average space for one point and thus the average size of the maximum patch must decrease. For a low number of events in the x_1-x_2 space, adding 1 point

still results in a significant enough change of the expected maximum patch size so that even a low number of Monte Carlo experiments would be able to differentiate between two neighbouring distributions in the N-space. However, with increasing N, the addition of 1 point to x_1-x_2 space has a much less significant effect on the size of the maximum patch, hence a low number of Monte Carlo experiments might not be able to differentiate between two neighbouring distributions in the N-space properly and the probabilities might get mixed up. For this reason, performing as many Monte Carlo experiments as possible is very important, even though it is very time-consuming.
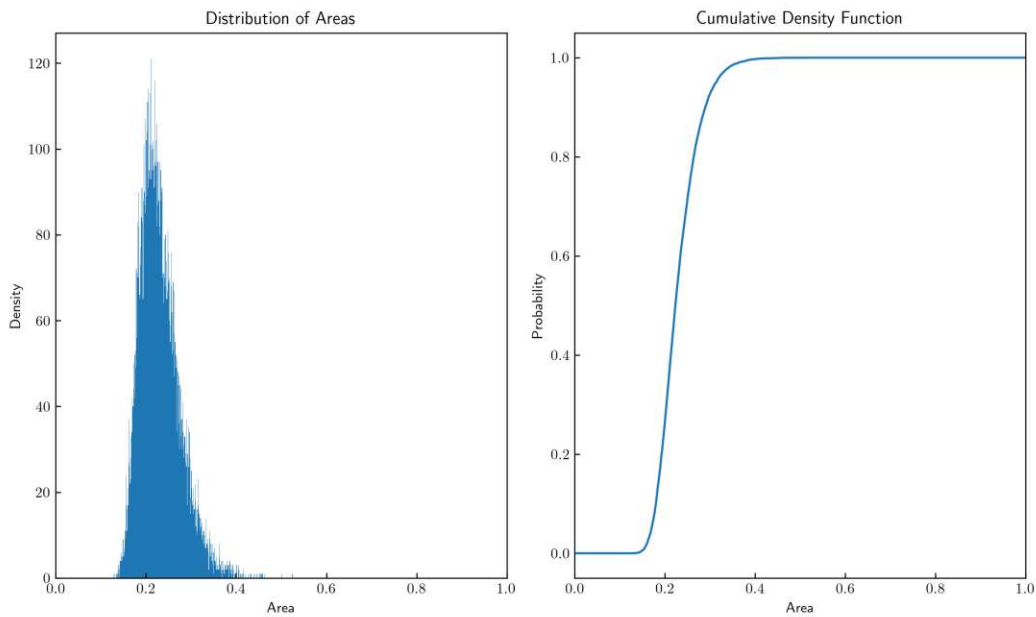
Figure 24: Left: The distribution of maximum patch areas calculated for $N = 25$. The histogram shifts towards 0 for increasing N values. Right: Corresponding cumulative distribution function of maximum patch areas. 10000 Monte Carlo experiments were performed for these results.

Using the transformation rules expressed in Eqs. (52) and (53), one can calculate the inverse transformation to get from x_1-x_2 plane back to E-LY plane. This allows the visualisation of the rectangular maximum patch in the x_1-x_2 plane in the E-LY plane. As expected, the maximum patch is no longer a rectangle, as shown in Fig. 25. By examining the distribution of the two-dimensional density function in the E-LY plane (39), one could have anticipated the observed shape of the maximum patch.

47
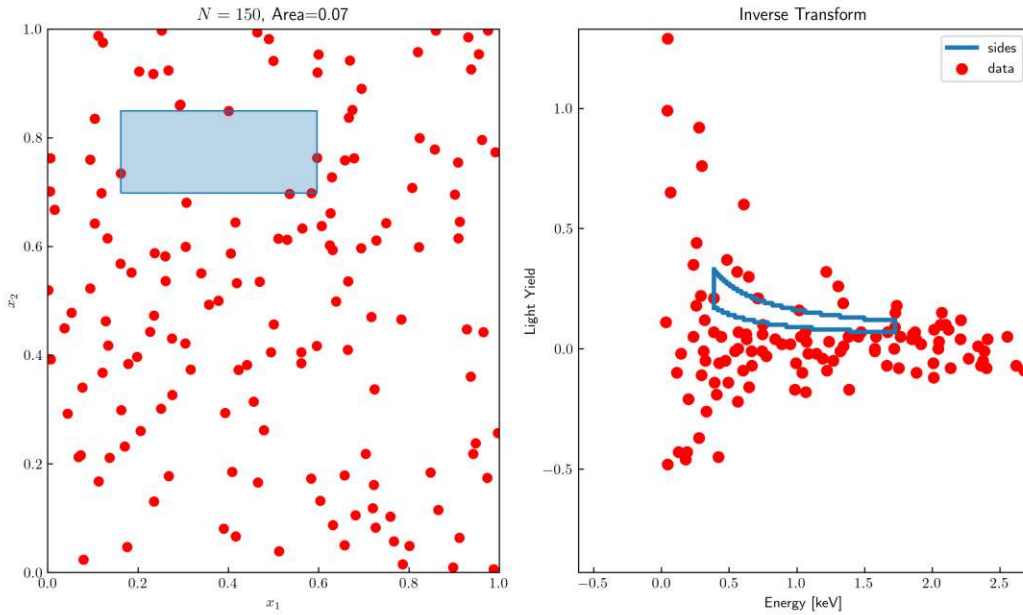
Figure 25: Left: The plot depicts 150 points representing the randomly generated data in Alg. 1. The blue rectangle represents the maximum patch for this Monte Carlo experiment. Right: The inverse transformation of these points and the maximum patch back to E-LY plane. Note that the graph is zoomed in around the transformed patch, hence not all 150 data points are visible in this plot.

It is also worth mentioning that $C_0(x, \mu)$ in Eq. (58) can be pre-calculated for preset N and μ-values. Since A is dependent on μ, by binning the A-values and using the distributions of maximum patch areas. Alg. 1, used for calculating the largest empty rectangle, is a slightly modified version of the algorithm described in [52] and is of at least the time complexity $\mathcal{O}(n^2)$. Thus, calculating $C_0(x, \mu)$ in the presence of many observed events increases the run-time significantly. For this reason, the implementation of the pre-calculated $C_0(x, \mu)$-values can be beneficial and time-saving. In the search for the largest empty rectangle, more intricate algorithms are available for utilisation. An alternative algorithm, as detailed in [52], exhibits a time complexity of $\mathcal{O}(n \log(n))$, offering the potential to further reduce the overall run-time.

4.6   Usage of Multiprocessing

Using Python's multiprocessing module is beneficial when dealing with processes that can be parallelised. In this study, as aforementioned, the tabulation of data needed for Yellin's optimum interval and maximum patch methods can be time-consuming. This is especially true for the optimum interval method if the expected number of events μ is very high or for the maximum patch method if the amount of data before any kind of pre-selection is very large. Luckily, these processes can be parallelised in a pretty straightforward manner.

For the optimum interval method, a total of four key calculations are parallelised:

- Calculation of x-values shown in Tab. 1.

- Calculation of x-distributions depicted in Fig. 16.

- Calculation of Γ-values shown in Tab. 2.

- Calculation of Γ_{Max} distributions depicted in Fig. 17.

In order to share the workload proportionately among multiple cores, a special sorting algorithm of μ-values is utilised. Suppose that the μ-values range from μ_{min} to μ_{max} and there is a total of four cores. First, four empty lists are created. The first list gets the largest and the smallest μ-values, μ_{max} and μ_{min}, and these values are then deleted from the list of μ-values. The second list gets the new largest and smallest μ-values, $\mu_{max,new}$ and $\mu_{min,new}$ and so on. After the fourth list, the loop sorts the new largest and smallest μ-values into the first list again. This procedure continues until all the μ-values are exhausted and sorted. This way, all cores get approximately the same workload. Parallelising the tabulation without this sorting would result in the first list being completed fairly quickly and the last list still computing maybe its first μ-value, because the larger μ-values require more computation. Without the sorting beforehand, the multiprocessing would save only a negligible amount of time.

The multiprocessing applied to the tabulation of data needed for the maximum patch method is in contrast very straightforward. The different number of random data points N is shared between cores. The cores then run N_{MC} Monte Carlo experiments for that particular N. The results are then saved into .txt files.

4.7 Required Inputs

The inputs needed for the proper execution of the code vary depending on one key factor; whether or not the user utilises the WIMP signal model described in section 2.1. The following inputs are required independent of the model used:

- List of dark matter masses, for which the limit should be calculated,

- Estimates of minimum and maximum values for the expected number of events μ,

- Confidence level of the limit,

- Unbinned measured data by the experiment,

- Whether to tabulate new random data for the optimum interval method and the number of lists in such data,

If the user chooses to use the dark matter model described in this work, then the following additional inputs are required:

- List of target materials containing information about the atomic mass, the mass ratio in relation to the molecule, the corresponding chemical symbol,

- Acceptance region efficiencies of the materials as a function of energy as explained in 1.2.2, if there are any at all,

- Detector-related information, i.e. exposure, resolution, threshold, upper energy limit of the region of interest, cut efficiency,

- The upper limit and the step size for the energy grid. These are parameters that are needed for the definition of the energy axis, which will be used for evaluating energy-dependent values,

- The sample size for the random variables that a user can draw from the summed probability distribution modelled with *SignalModel*. It is needed only if the rvs method will be used.

However, if the user chooses to use another signal model, but wishes to make use of the methods implemented, then the required additional inputs are:

- List of cross section σ values and corresponding expected number of events μ, for each dark matter mass,

- List of cumulative density functions (CDFs) with the corresponding energy grid values, for each dark matter mass,

- If the data is two-dimensional, then the list of two-dimensional density functions (2D-DFs) with the corresponding energy grid and light yield values, for each dark matter mass,

- Whether μ's dependence on σ is linear or not.

Note that the algorithm can compute $\bar{\sigma}$ only if $\mu(\sigma)$ is a strictly monotonic function.

5 Results

In this chapter, the results of the implementation of the signal model and Yellin's analysis methods will be presented and discussed.

5.1 One-Dimensional Density Functions

The signal model presented in this work depends on many variables, as explored in section 2.1, however, this section will focus on the dependence of probability density functions (PDFs) and cumulative density functions (CDFs) on dark matter masses.
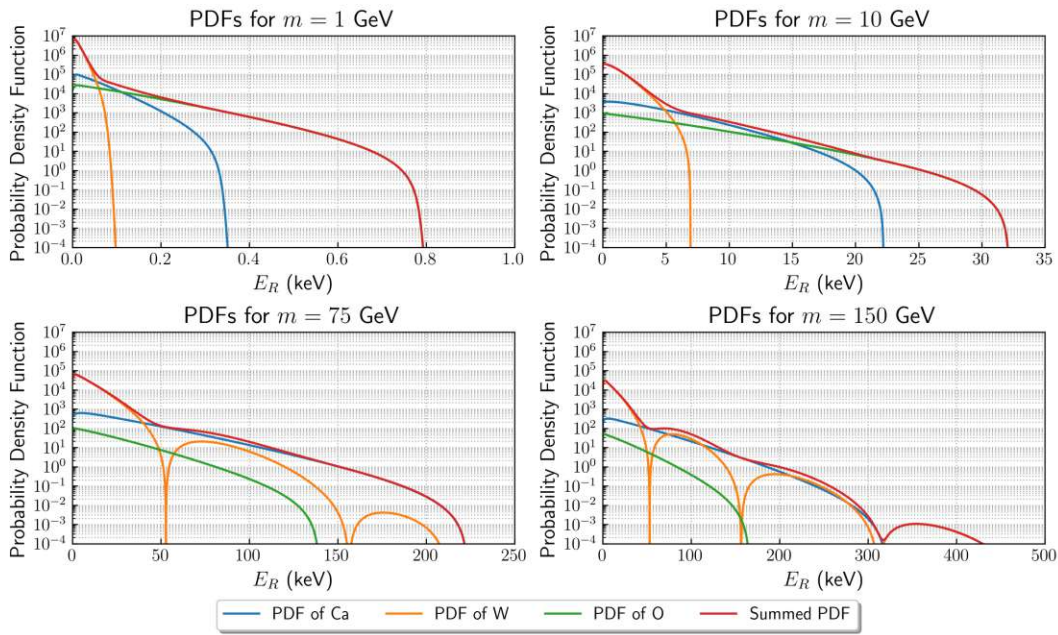
Figure 26: Graphs of non-normalised PDFs multiplied with their respective relative mass density in the molecule $CaWO_4$ for dark matter masses m_χ equalling 1 GeV, 10 GeV, 75 GeV and 150 GeV. The blue coloured graph represents the PDF of calcium, the orange coloured graph represents the PDF of tungsten, the green coloured graph represents the PDF of oxygen and the red coloured graph represents the weighted sum of all these three PDFs. The computations of the PDFs incorporate distinct parameters unique to CRESST, including resolution, cut efficiency, exposure, energy threshold, the chemical composition of the crystal, and acceptance region efficiencies of individual elements.

As expected from the dependence of Eq. (17) on dark matter mass m_χ and on the material, different PDFs for different masses and different materials are

52

obtained. As can be seen from Fig. 26, different materials dominate the summed PDF for different dark matter masses. For small masses, the summed PDF is dominated by tungsten for lower energies and mainly dominated by oxygen for higher energies, and the PDF of calcium has hardly any impact on the summed PDF. With increasing masses, however, one sees the increasing influence of calcium more clearly for medium and high energies. For large masses, such as 150 GeV, one sees the effect of the form factor (9) and thus the influence of the PDF of tungsten on the summed PDF. Furthermore, one can see that the maximum achievable recoil energies increase with increasing mass.
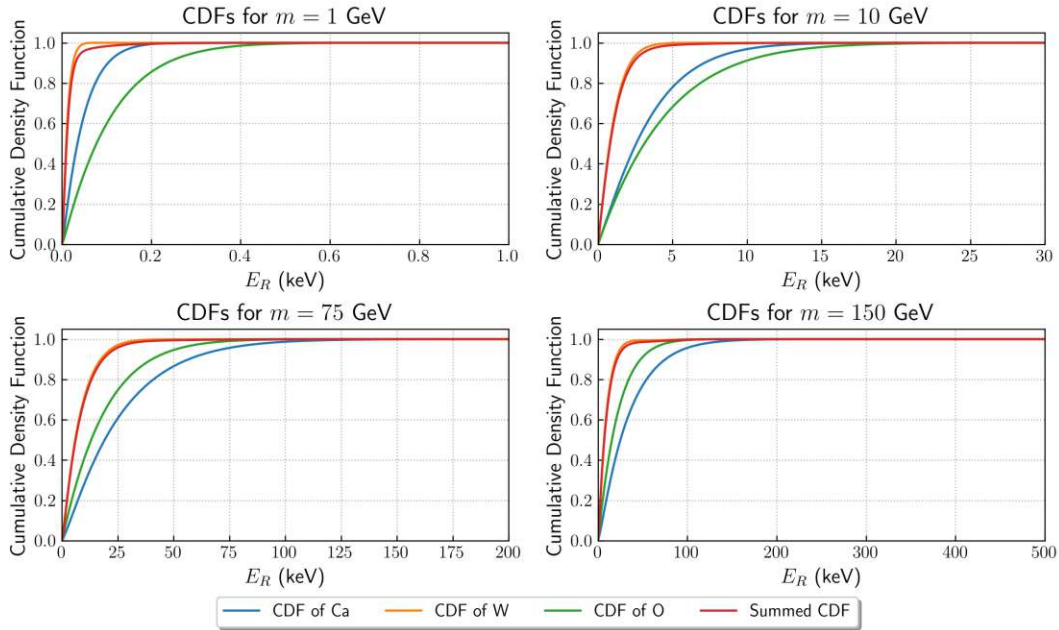
Figure 27: Graphs of normalised CDFs for dark matter masses m_χ equalling 1 GeV, 10 GeV, 75 GeV and 150 GeV. The blue coloured graph represents the CDF of calcium, the orange coloured graph represents the CDF of tungsten, the green coloured graph represents the CDF of oxygen and the red coloured graph represents the sum of all these three CDFs. The computations of the CDFs incorporate distinct parameters unique to CRESST, including resolution, cut efficiency, exposure, energy threshold, the chemical composition of the crystal, and acceptance region efficiencies of individual elements.

In the figure of the CDFs (Fig. 27), one can observe how the summed CDF becomes more and more dominated by the CDF/PDF of tungsten as the mass increases, as one would expect, as the area under the PDF of tungsten in the graph of PDFs (Fig. 26) becomes increasingly larger compared to that of the other elements.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU Bibliothek
Your knowledge hub
WIEN

5.2 Two-Dimensional Density Functions

As mentioned in section 2.2, the expansion of the one-dimensional PDF into a two-dimensional density function (2D-DF) in the energy-light yield plane can be achieved by inserting the one-dimensional PDFs, save for the multiplication with the acceptance region efficiencies, into Eq. (39). Therefore, since the mass dependency of the PDFs is already discussed in the previous section 5.1, this section will focus on the target-nuclei dependency of the 2D-DFs.
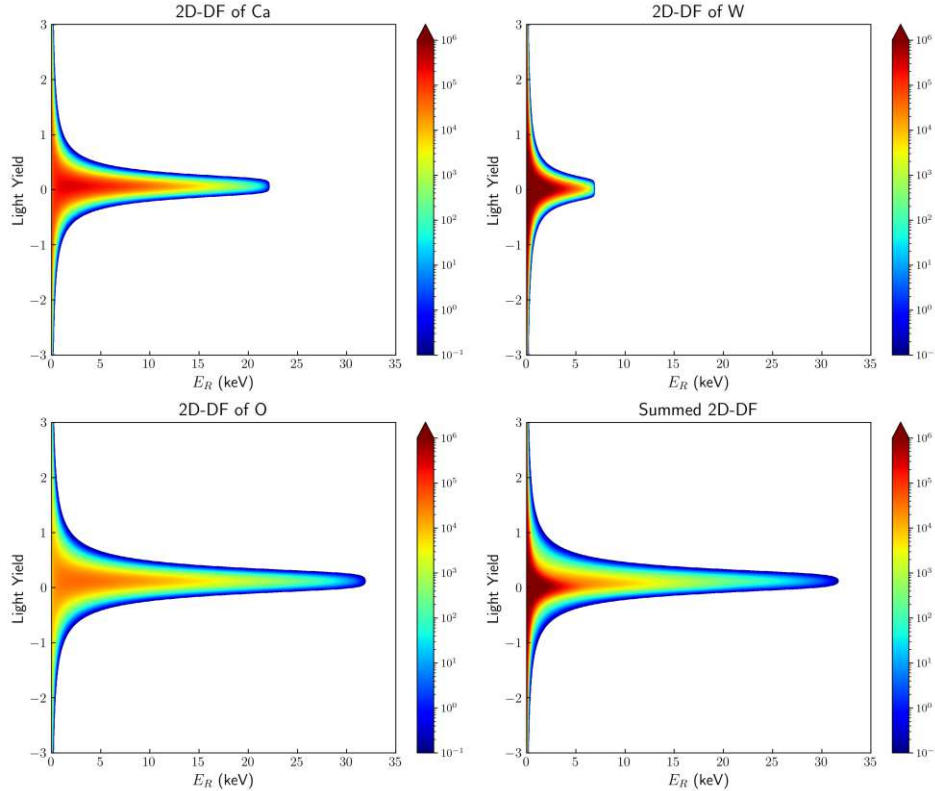
Figure 28: Graphs of non-normalised 2D-DFs multiplied with their respective relative mass density in the molecule $CaWO_4$ for a dark matter mass m_χ of 10 GeV. Top left, top right and bottom left plots show the partial 2D-DFs of calcium, tungsten and oxygen respectively. The bottom right plot depicts the weighted sum of these three partial 2D-DFs, resulting in the total expected signal distribution in the E-LY space. The computations of the 2D-DFs incorporate distinct parameters unique to CRESST, including resolution, cut efficiency, exposure, energy threshold and the chemical composition of the crystal.

Fig. 28 consists of four plots, three of them representing the partial 2D-DFs for the three nuclei of $CaWO_4$ and the fourth one representing the weighted sum of

these three 2D-DFs. As anticipated from the results depicted in Fig. 26, different nuclei stretch up until different energy values along the horizontal axes and the spread in light yield along the vertical axes are different, which can be explained by the nuclei-specific variables found in expression (39). Following the pattern shown in Fig. 26, the lower energy region is heavily influenced by the presence of tungsten, while the influence of other nuclei becomes more apparent at higher energies, particularly as the probability of tungsten recoil events get effectively negligible.
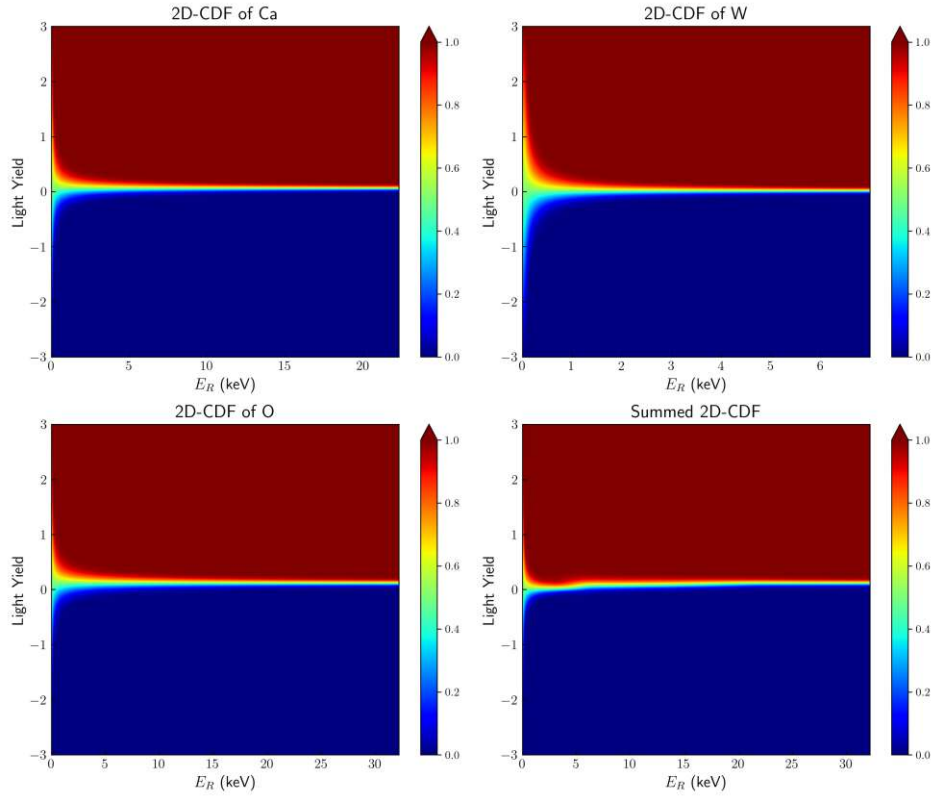
Figure 29: Graphs of 2D-CDFs for a dark matter mass m_χ of 10 GeV and $CaWO_4$. The top left, top right and bottom left plots show the 2D-CDFs of calcium, tungsten and oxygen respectively. The bottom right plot depicts the 2D-CDF constructed from the weighted sum of these three partial 2D-DFs. Note that these 2D-CDFs are constructed by summing the values of the 2D-DFs along their light yield axis. The energy axes stretch up to maximum recoil energies that can be calculated by utilising the Eq. (24). The computations of the 2D-CDFs incorporate distinct parameters unique to CRESST, including resolution, cut efficiency, exposure, energy threshold and the chemical composition of the crystal.

55

The presented visualisations in Fig. 29 illustrate the two-dimensional cumulative density functions (2D-CDFs). These functions are generated by cumulatively summing the values of the two-dimensional probability density functions along the light yield axis. The varying lengths of the 2D-CDFs along the energy axis can be explained by Eq. (24). This equation plays a crucial role in determining the energy value beyond which the probability density function is equal to zero.

In contrast to the representation in Fig. 28, the differences in energy-dependent factors specific to different nuclei, which influence the Gaussian spread along the light yield axis, become more evident in the cumulative distribution functions. Additionally, the summed 2D-CDF exhibits noticeable asymmetry in the lower energy region, particularly up to 5 keV, a feature not as clearly visible in Fig. 28. This asymmetry predominantly arises from the impact of tungsten recoil events that dominate the lower energy regions, as highlighted in the top-right plot of Fig. 28.

5.3 Exclusion Charts

Four different Yellin's methods were discussed in section 3. This chapter will present the results these methods yield when applied to data provided by CRESST-III Det. A and CRESST-II Lise detector modules [26, 51]. The results of these methods implemented in this study will then be compared to the limits calculated by the original script written by S. Yellin. It is important to emphasise that the data made available by these experiments include only one-dimensional recoil energy data, and, the light yield data is not included in these data sets. Hence, in this study, the maximum patch method will only be applied to simulated data.

An exclusion limit represents the border between two areas. Above the exclusion limit, physicists can say that if there was a WIMP with such and such mass and cross section, an experiment should have been able to detect this WIMP with a certain confidence level. Under the exclusion limit is the area where the experiment with its current specifications is simply not able to detect or rule out the existence of a WIMP with such and such mass and cross section. Since the detector should have been able to detect WIMP presence with particle mass and cross section landing in the area above the exclusion limit and since no dark matter particles were detected in this area, these mass-cross section pairs can be excluded by this experiment.
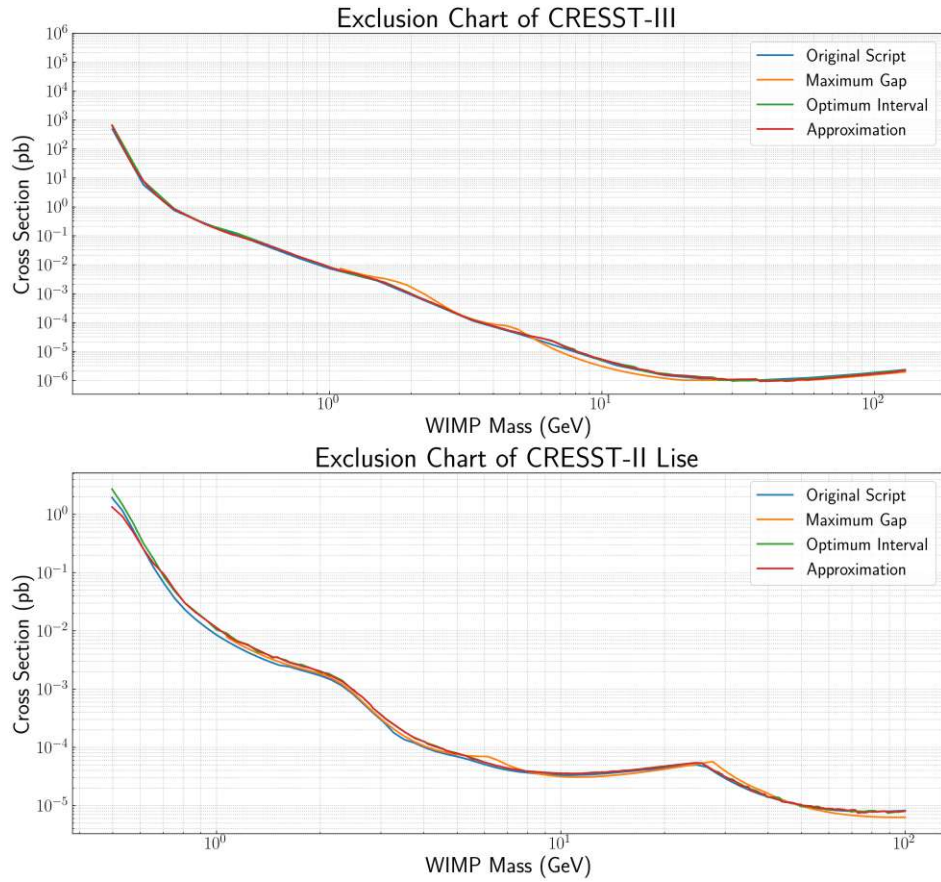
Figure 30: Exclusion charts of CRESST-III and CRESST-II Lise experiments showing the upper limits on cross sections of WIMP-nucleon scattering given in pb, for corresponding WIMP masses given in GeV. The blue line represents the cross sections determined by the original implementation of Yellin's optimum interval method written in Fortran. The orange, green and red exclusion lines represent the cross sections determined by *ModeLimit* using the maximum gap method, optimum interval method and the approximation for the optimum interval method respectively.

The exclusion charts of CRESST-III and CRESST-II Lise experiments are shown in Fig. 30. It is observed that the results of the approximation and the optimum interval method implemented in *ModeLimit* seem to overlap almost perfectly with the results calculated by the original implementation of Yellin's optimum interval method, if a bit more conservative. The slight differences may have been caused by the parameter differences in calculations of the expected signal discussed in 2.1. The maximum gap method seems to deviate in its results from the other methods. Section 3 explains how the maximum gap method works and how it

only focuses on gaps with no other data points. This is a very strong limitation because the gap with zero data points in it may not be the best option to set the upper limit. Therefore, such deviations are not unexpected.

As described in section 3.2, in order to calculate the limits shown in Fig. 30, first the corresponding $\bar{\mu}$-values are being calculated. The information on the $\bar{\mu}$-values needed for calculation may be useful for the recommended workflow, which is described in section A. Using the approximation of the optimum interval method, one can calculate the estimate of the needed maximum value of $\bar{\mu}$ in a quick and convenient way. Then, making use of this particular information, one can tabulate new data with μ-values reaching beyond the estimated maximum value of $\bar{\mu}$ for more accurate results.
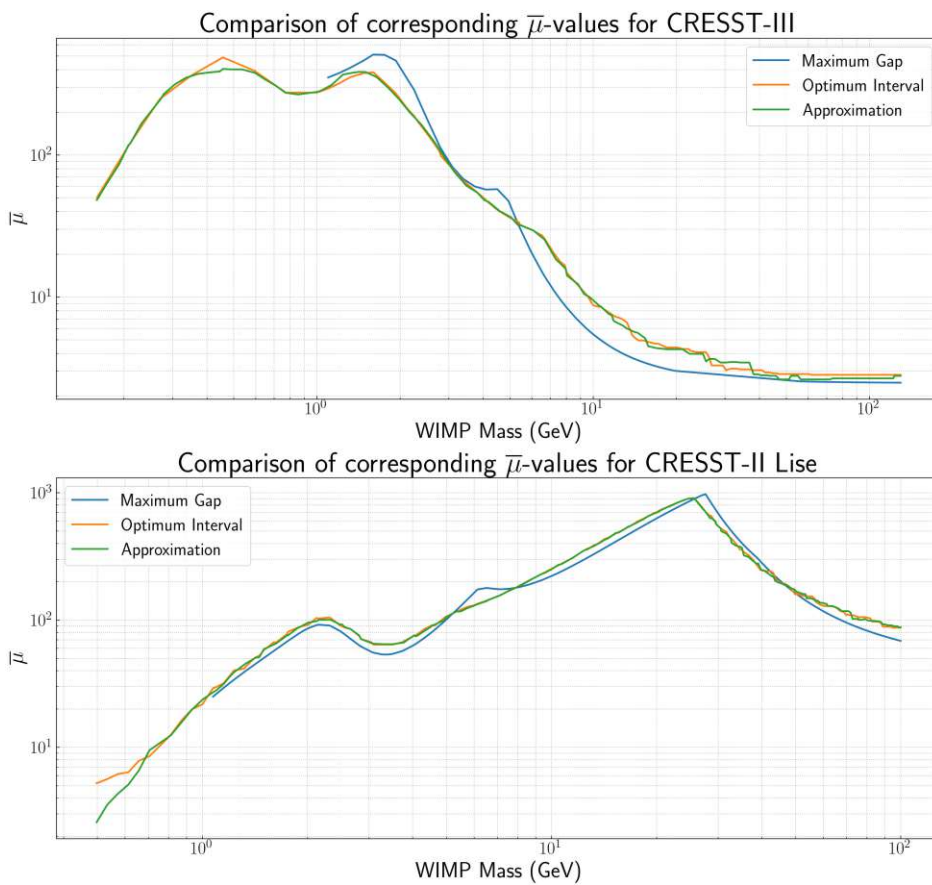
Figure 31: Corresponding $\bar{\mu}$-values to the cross sections visualised in Fig. 30. The blue, orange and green lines represent the $\bar{\mu}$-values determined by *ModeLimit* using the maximum gap method, optimum interval method and the approximation for the optimum interval method respectively.

The graphical representation in Fig. 31 showcases the $\bar{\mu}$-values plotted against

dark matter masses. These $\bar{\mu}$-values play a pivotal role in determining the limits depicted in Figure 30. Similar to the observations in Fig. 30, the results of the approximation and the optimum interval method with newly tabulated data agree with each other. However, the results obtained through the maximum gap method deviate from the aforementioned approaches, mirroring the trends observed in the exclusion chart plot.

Specifically, the calculated $\bar{\mu}$-values for the experimental data from CRESST-III hover around 500, while those derived from the CRESST-II Lise data reach values up to 1000. It is noteworthy that tabulating data for achieving such high expected numbers of events, necessary for calculating these $\bar{\mu}$ values, proves to be a resource-intensive and time-consuming process. This consideration prompts a closer examination of the advantages and disadvantages associated with the approximation of the optimum interval method, as detailed in section 5.4. Furthermore, the optimal usage of the approximation within an ideal workflow is discussed in section A.

Moving forward, a comprehensive examination of the outcomes yielded by the maximum patch method when applied to simulation data becomes the focus of attention. The results obtained through the maximum patch method will be compared to the results of the optimum interval method applied to the full data and the subset of the data within the acceptance region, as defined in Fig. 10.

As can be seen in Fig. 32, the exclusion limits determined by the maximum patch method demonstrate similar behaviour as the exclusion limits determined by the optimum interval method. However, a notable distinction lies in the strength of the limits, with the maximum patch method outperforming due to its incorporation of not only the energy of the event but also the light yield. This consideration provides a more refined assessment of the exclusion limits. As anticipated, the optimum interval method, when applied to the full dataset, yields the weakest limits.

59
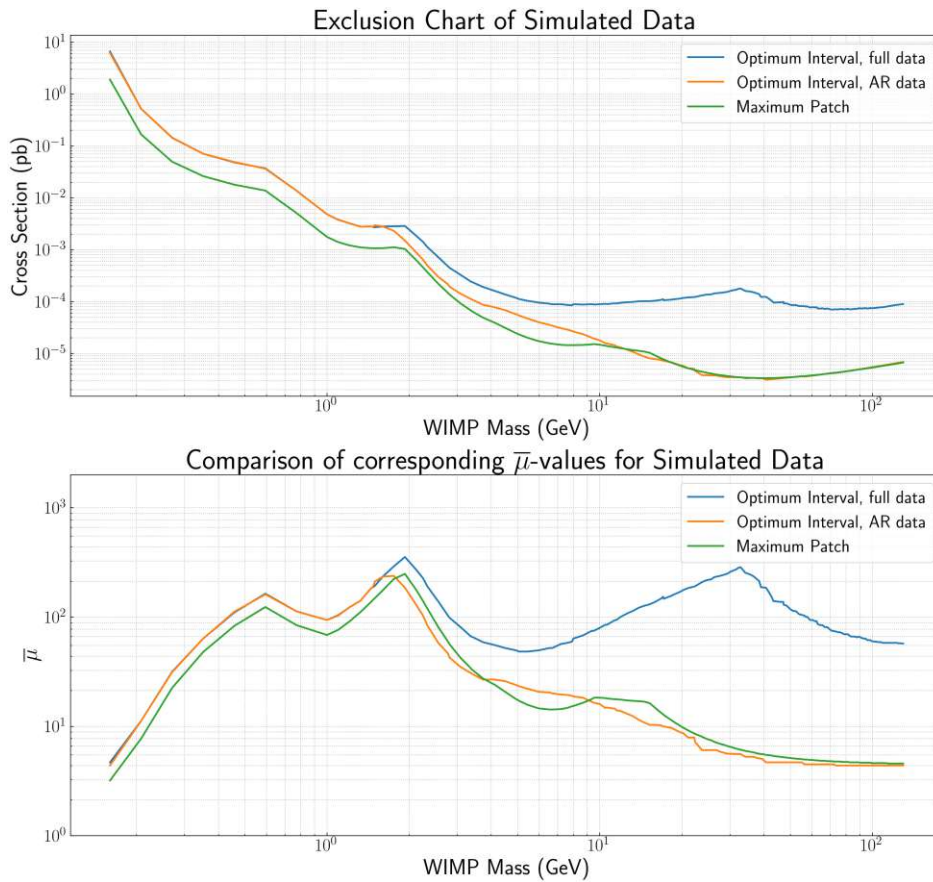
Figure 32: Top: Exclusion chart of the simulation data showing the upper limits on cross sections of the WIMPs given in pb for corresponding WIMP masses given in GeV. Bottom: Corresponding $\bar{\mu}$-values to the cross sections plotted in the figure above. The blue, orange, and green lines show values from different methods: The optimum interval method applied to full data, the optimum interval method applied to data in the acceptance region, and the maximum patch method, respectively.

In Fig. 33, it can be seen how the limits shown in Fig. 32 relate to each other. A better comprehension of behavioural disparities in various mass regions can be achieved by calculating the quotients of different limits set by different methods. For lower masses, the optimum interval method applied to the full data and data in the acceptance region provide similar results, hence the quotients are close to unity. In this mass region, the maximum gap method yields limits that are between 30% and 40% of the limits set by the optimum interval method.

In the medium mass region, the limits' relations show significant changes. The optimum interval method no longer yields the same limits when applied to full

data and data in the acceptance region. The results of the optimum interval method applied to data in the acceptance region tend to move towards the limits determined by the maximum patch method.

For higher masses, these methods result in similar limits, hence the quotients are close to unity. The gap between the results of the optimum interval method applied to full data and the maximum patch method/optimum interval method applied to data in the acceptance region is larger in the high mass region than in the low and medium mass regions.
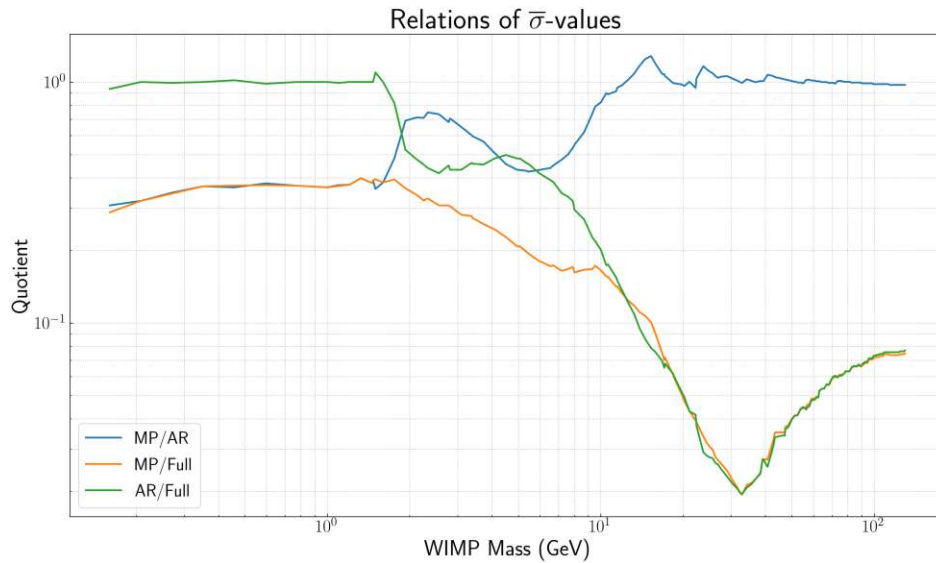
Figure 33: Top: Exclusion chart of the simulation data showing the upper limits on cross sections of the WIMPs given in pb for corresponding WIMP masses given in GeV. Bottom: Corresponding $\bar{\mu}$-values to the cross sections plotted in the figure above.

5.4 Advantages and Disadvantages of the Approximation

While the approximation shows significant overlap with the results of actually tabulated data in Fig. 30, it is important to note its limitations. As discussed in section 4.4, using the μ-values used in data tabulation for the approximation yields the best results because it models the distribution of x- and Γ-values the least and delivers a downsampled version of actually pre-tabulated data. Using μ-values that are already covered by the μ-range of the pre-tabulated data but are not exactly the same μ-values used for the pre-tabulated results in deviations. Using μ-values that are not covered by the μ-range of the pre-tabulated data adds another layer of deviation from true limits, due to the behaviour shown in Fig. 22.

61

So, it is of importance to not use only the approximation for end results, due to accuracy concerns.

	Tabulation	Saved data	Approximation
RAM [GB]	20.31	10.16	0.363
Storage [GB]	1.93	0.62	0.4
Run-time [s]	3217.21	1738.25	250.49

Table 3: Table comparing the results of RAM usage, storage space usage and run-time measurements. The approximation is calculated using a total of 1000 geometrically spaced μ-values between 0.25 and 2000 and 1000 lists are generated for each μ. For the calculation by tabulating new data, a total of 600 geometrically spaced μ-values are used with 1000 lists consisting of random numbers with Poisson distributed sizes for each μ. The limits are determined for 168 dark matter masses spanning from 0.16 GeV up to 129.746 GeV.

The implementation of this approximation is driven by its notable advantages. In a quantitative comparison between optimum interval calculations by tabulating new data, by using pre-tabulated saved data and the approximation, the approximation excels not only in run-time but also in RAM usage and storage space needed. Table 3 shows that the approximation is about 13 times faster than tabulating new data and about 7 times faster than using saved pre-tabulated data. The storage space needed for the approximation is also smaller than both, especially considering that the approximation in this case covers a μ-range between 0.25 and 2000, whereas the tabulated data only covers the μ-range between 0.25 and 1000. Since, as explained in section 4.3, the sizes of the lists are Poisson distributed, tabulating data for higher μ-values would result in more RAM usage and storage space needed for saving data. The RAM comparison is very significant. Whereas the tabulation of new data costs 20 GB of RAM, using the pre-tabulated saved data costs only half of that and making use of the approximation is only a mere fraction of about 2%. As one can see, even though the accuracy of the approximation is of course not as good as that of tabulated data, these advantages still make it useful.

6 Conclusion

The Python package *ModeLimit* equips researchers engaged in the domain of dark matter experiments with a versatile and flexible tool. It allows researchers to model both the one-dimensional and two-dimensional PDFs and CDFs of the expected spin-independent dark matter signal as described in section 2. Then, using these CDFs in combination with the measurement data or simulated data, *ModeLimit* enables the determination of upper limits on the cross sections of the dark matter particles in the presence of an unknown background by leveraging analysis methods proposed by S. Yellin, namely the maximum gap method, optimum interval method, and maximum patch method. In addition to the expected dark matter signal implemented in *ModeLimit*, researchers may choose to tweak the default class variables, such as the dark matter density, or completely ignore the implemented expected signal to employ their models of expected signals for the determination of the upper limits using Yellin's methods.

In contrast to the maximum gap method, the optimum interval method and the maximum patch method cannot be solved analytically. For this reason, *ModeLimit* is equipped with an approximation of the optimum interval method as described in sections 3.3 and 4.4. This approximation offers a resource-efficient alternative without compromising significant accuracy, see Fig. 30. Detailed in the recommended workflow in section A is an exemplary utilisation of the approximation. Furthermore, by parallelising the data tabulation processes for the optimum interval method and maximum patch method, the run-time is reduced substantially. This efficiency gain is particularly high for tables with high μ-values as the size of the tabulated data grows with increasing μ-values.

One of the notable contributions of *ModeLimit* is its integration of Yellin's maximum patch method. Unlike the maximum gap and optimum interval methods, which are designed to analyse only one-dimensional data like recoil energy, *ModeLimit* broadens the scope by incorporating the maximum patch method and hence enabling the analysis of two-dimensional data, such as recoil energy and light yield. This approach eliminates the need for the acceptance region chosen by researchers based on their knowledge of the data, as illustrated in Figure 10. By algorithmically determining the most optimal area for analysis, the maximum patch method ensures a more unbiased outcome, thereby enhancing the overall objectivity and reliability of the results.

Crucially, the results generated by the *ModeLimit* package, as extensively detailed in Section 5.3, exhibit a high degree of alignment with the findings published by the CRESST Collaboration, calculated by using the earlier Fortran-based original script. This robust alignment underscores the reliability and accuracy of *ModeLimit* in reproducing outcomes consistent with established calculations.

Moreover, *ModeLimit* distinguishes itself through its user-friendly and easily

integrable design into Python-based research environments. This characteristic streamlines the implementation process and equips researchers with a straightforward and accessible tool for computing upper limits in rare event searches, particularly in the context of experiments focused on dark matter. Thus, the package's adaptability and simplicity contribute to widening the accessibility of the aforementioned analysis methods and fostering collaboration in the field of rare event searches.

In summary, the open-source Python package *ModeLimit* emerges as an important and potentially useful implementation of Yellin's methods for setting upper limits when analysing experimental or simulated data of dark matter searches with a presumed unknown background. Its successful alignment with established results, coupled with its incorporation of innovative features, signals a promising step forward in the world of dark matter research.

Looking forward, it's clear that *ModeLimit* holds promise well beyond the boundaries of the CRESST experiment. A particularly exciting avenue lies in employing *ModeLimit* across a range of experiments, each approaching the quest for elusive dark matter in its own unique way. For instance, experiments such as LZ, PandaX, and XENONnT, which employ liquid xenon in the search of elusive dark matter particles, can potentially benefit from the versatility of Yellin's methods implemented in *ModeLimit*. Thus, *ModeLimit* can prove invaluable for setting upper limits on the cross section of dark matter by utilising data from diverse experiments in the field.

List of Tables

List of Figures

70

References

[1] S. Yellin. "Finding an upper limit in the presence of an unknown background". In: *Physical Review D* 66.3 (2002). URL: https://doi.org/10.1103%2Fphysrevd.66.032005 (cit. on pp. 1, 25, 27).

[2] S. Yellin. "Extending the optimum interval method". In: *arXiv* (2007). URL: https://arxiv.org/abs/0709.2701 (cit. on pp. 1, 25, 30).

[3] S. Yellin. "Software for computing an upper limit given unknown background". URL: https://www.slac.stanford.edu/~yellin/ULsoftware.html (visited on 11/26/2023) (cit. on p. 1).

[4] C. Tunnell and J. Aalbers. "PyOptimumInterval". URL: https://github.com/tunnell/optimum_interval (visited on 11/26/2023) (cit. on p. 2).

[5] D. Abrams et al. "Exclusion limits on the WIMP-nucleon cross section from the Cryogenic Dark Matter Search". In: *Physical Review D* 66.12 (2002). URL: https://doi.org/10.1103%2Fphysrevd.66.122003 (cit. on p. 3).

[6] E. Aprile et al. "First Dark Matter Search with Nuclear Recoils from the XENONnT Experiment". In: *Physical Review Letters* 131.4 (2023). URL: https://doi.org/10.1103%2Fphysrevlett.131.041003 (cit. on pp. 3, 6).

[7] D.S. Akerib et al. "The LUX-ZEPLIN (LZ) experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 953 (2020). URL: https://doi.org/10.1016%2Fj.nima.2019.163047 (cit. on pp. 3, 6).

[8] Yue Meng et al. "Dark Matter Search Results from the PandaX-4T Commissioning Run". In: *Physical Review Letters* 127.26 (2021). URL: https://doi.org/10.1103%2Fphysrevlett.127.261802 (cit. on pp. 3, 6).

[9] N. Aghanim et al. "Planck2018 results: VI. Cosmological parameters". In: *Astronomy amp; Astrophysics* 641 (2020). URL: http://dx.doi.org/10.1051/0004-6361/201833910 (cit. on p. 3).

[10] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helvetica Physica Acta* 6 (1933). URL: https://ui.adsabs.harvard.edu/abs/1933AcHPh...6..110Z (cit. on p. 3).

[11] Vera C. Rubin and Jr. Ford W. Kent. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *Astrophysical Journal* 159 (1970). URL: https://ui.adsabs.harvard.edu/abs/1970ApJ...159..379R (cit. on p. 4).

[12] Vera C. Rubin. "The Rotation of Spiral Galaxies". In: *Science* 220.4604 (1983). URL: https://www.science.org/doi/abs/10.1126/science.220.4604.1339 (cit. on p. 4).

[13]   Wikipedia. "Galactic Rotation Curve". URL: `https://upload.wikimedia.org/wikipedia/commons/b/b9/GalacticRotation2.svg` (visited on 11/26/2023) (cit. on p. 4).

[14]   Douglas Clowe, Anthony Gonzalez, and Maxim Markevitch. "Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter". In: *The Astrophysical Journal* 604.2 (2004). URL: `https://doi.org/10.1086%2F381970` (cit. on p. 4).

[15]   M. Markevitch et al. "Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56". In: *The Astrophysical Journal* 606.2 (2004). URL: `https://doi.org/10.1086%2F383178` (cit. on p. 4).

[16]   Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". In: *The Astrophysical Journal* 648.2 (2006). URL: `https://doi.org/10.1086%2F508162` (cit. on p. 4).

[17]   Chandra X-Ray Observatory. "More Images of 1E 0657-56". URL: `https://chandra.harvard.edu/photo/2006/1e0657/more.html` (visited on 11/26/2023) (cit. on p. 5).

[18]   D. N. Spergel et al. "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters". In: *The Astrophysical Journal Supplement Series* 148.1 (2003). URL: `https://doi.org/10.1086%2F377226` (cit. on p. 5).

[19]   Planck Collaboration et al. "Planck 2018 results". In: *Astronomy &amp Astrophysics* 641 (2020). URL: `https://doi.org/10.1051%2F0004-6361%2F201833910` (cit. on p. 5).

[20]   ESA. "Cosmic Microwave Background seen by Planck". URL: `https://sci.esa.int/web/planck/-/51553-cosmic-microwave-background-seen-by-planck` (visited on 11/26/2023) (cit. on p. 6).

[21]   MAGIC collaboration. "Limits to dark matter annihilation cross-section from a combined analysis of MAGIC and Fermi-LAT observations of dwarf satellite galaxies". In: *Journal of Cosmology and Astroparticle Physics* 2016.02 (2016). URL: `https://doi.org/10.1088%2F1475-7516%2F2016%2F02%2F039` (cit. on p. 6).

[22]   Benjamin Zitzer and VERITAS Collaboration. "The VERITAS Dark Matter Program". (2017). URL: `https://doi.org/10.48550/arXiv.1708.07447` (cit. on p. 6).

[23] Micael Andrade et al. "Dark Matter searches in Dwarf Galaxies with the Southern Wide-field Gamma-ray Observatory". In: *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*. (2023). URL: `https://doi.org/10.22323%2F1.444.1413` (cit. on p. 6).

[24] Francesco Gabriele Saturni et al. "Dark matter searches in dwarf spheroidal galaxies with the Cherenkov Telescope Array". In: *arXiv* (2023). URL: `https://doi.org/10.48550/arXiv.2309.09607` (cit. on p. 6).

[25] A. Albert et al. "Search for Decaying Dark Matter in the Virgo Cluster of Galaxies with HAWC". In: *arXiv* (2023). URL: `https://doi.org/10.48550/arXiv.2309.03973` (cit. on p. 6).

[26] CRESST Collaboration et al. "Description of CRESST-III Data". In: *arXiv* (2019). URL: `https://arxiv.org/abs/1905.07335` (cit. on pp. 6, 12, 13, 32, 33, 56).

[27] R. Bernabei et al. "The dark matter: DAMA/LIBRA and its perspectives". In: *arXiv* (2021). URL: `https://doi.org/10.48550/arXiv.2110.04734` (cit. on p. 6).

[28] XMASS Collaboration et al. "Direct dark matter searches with the full data set of XMASS-I". 2023. URL: `https://doi.org/10.1103/PhysRevD.108.083022` (cit. on p. 6).

[29] D. S. Akerib et al. "Results from a Search for Dark Matter in the Complete LUX Exposure". In: *Physical Review Letters* 118.2 (2017). URL: `https://doi.org/10.1103%2Fphysrevlett.118.021303` (cit. on p. 6).

[30] CERN. "Dark matter - Invisible dark matter makes up most of the universe – but we can only detect it from its gravitational effects". URL: `https://home.cern/science/physics/dark-matter` (visited on 11/26/2023) (cit. on p. 6).

[31] Oliver Buchmueller, Caterina Doglioni, and Lian-Tao Wang. "Search for dark matter at colliders". In: *Nature Physics* 13.3 (2017). URL: `https://doi.org/10.1038%2Fnphys4054` (cit. on p. 6).

[32] Patrick J. Fox et al. "LEP shines light on dark matter". In: *Physical Review D* 84.1 (2011). URL: `https://doi.org/10.1103%2Fphysrevd.84.014028` (cit. on p. 6).

[33] Mads T. Frandsen et al. "LHC and Tevatron bounds on the dark matter direct detection cross-section for vector mediators". In: *Journal of High Energy Physics* 2012.7 (2012). URL: `https://doi.org/10.1007%2Fjhep07%282012%29123` (cit. on p. 6).

[34] Yang Bai, Patrick J. Fox, and Roni Harnik. "The Tevatron at the frontier of dark matter direct detection". In: *Journal of High Energy Physics* 2010.12 (2010). URL: https://doi.org/10.1007%2Fjhep12%282010%29048 (cit. on p. 6).

[35] CRESST. "Cryogenic Rare Event Search with Superconducting Thermometers". URL: https://cresst-experiment.org/ (visited on 11/26/2023) (cit. on pp. 7, 11).

[36] A. H. Abdelhameed et al. "Geant4-based electromagnetic background model for the CRESST dark matter experiment". In: *The European Physical Journal C* 79.10 (2019). URL: https://doi.org/10.1140%2Fepjc%2Fs10052-019-7385-0 (cit. on pp. 8, 10).

[37] Florian Reindl. "Exploring Light Dark Matter With CRESST-II Low-Threshold Detectors". PhD thesis. Technische Universität München, 2016 (cit. on p. 10).

[38] Mariangela Lisanti. "Lectures on Dark Matter Physics". In: *New Frontiers in Fields and Strings*. WORLD SCIENTIFIC, (2016). URL: http://dx.doi.org/10.1142/9789813149441_0007 (cit. on p. 14).

[39] Jens Michael Schmaler. "The CRESST Dark Matter Search - New Analysis Methods and Recent Results". PhD thesis. TU München, (2010). URL: https://mediatum.ub.tum.de/node?id=998304 (cit. on pp. 14, 21, 22, 29, 45).

[40] F. Donato, N. Fornengo, and S. Scopel. "Effects of galactic dark halo rotation on WIMP direct detection". In: *Astroparticle Physics* 9.3 (1998). URL: https://doi.org/10.1016%2Fs0927-6505%2898%2900025-5 (cit. on pp. 14, 16, 21).

[41] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle dark matter: evidence, candidates and constraints". In: *Physics Reports* 405.5-6 (2005). URL: https://doi.org/10.1016%2Fj.physrep.2004.08.031 (cit. on p. 15).

[42] Richard H. Helm. "Inelastic and Elastic Scattering of 187-Mev Electrons from Selected Even-Even Nuclei". In: *Phys. Rev.* 104.5 (1956). URL: https://link.aps.org/doi/10.1103/PhysRev.104.1466 (cit. on p. 16).

[43] J. Engel. "Nuclear form factors for the scattering of weakly interacting massive particles". In: *Physics Letters B* 264.1 (1991). URL: https://www.sciencedirect.com/science/article/pii/037026939190712Y (cit. on pp. 16, 17).

[44] J.D. Lewin and P.F. Smith. "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil". In: *Astroparticle Physics* 6.1 (1996). URL: `https://www.sciencedirect.com/science/article/pii/S0927650596000473` (cit. on pp. 17, 20).

[45] Prajwal Raj Kafle et al. "On the Shoulders of Giants: Properties of the Stellar Halo and the Milky Way Mass Distribution". In: *The Astrophysical Journal* 794.1 (2014). URL: `https://doi.org/10.1088%2F0004-637x%2F794%2F1%2F59` (cit. on p. 21).

[46] R. Strauss et al. "Energy-dependent light quenching in $CaWO_4$ crystals at mK temperatures". In: *The European Physical Journal C* 74.7 (2014). URL: `https://doi.org/10.1140%2Fepjc%2Fs10052-014-2957-5` (cit. on pp. 22, 23).

[47] R. F. Lang et al. "Scintillator Non-Proportionality and Gamma Quenching in $CaWO_4$". In: *arXiv* (2009). URL: `https://doi.org/10.48550/arXiv.0910.4414` (cit. on pp. 22, 23).

[48] NumPy. "NumPy". URL: `https://numpy.org/` (visited on 11/26/2023) (cit. on p. 32).

[49] SciPy. "SciPy". URL: `https://scipy.org/` (visited on 11/26/2023) (cit. on p. 32).

[50] Matplotlib. "Matplotlib". URL: `https://matplotlib.org/` (visited on 11/26/2023) (cit. on p. 32).

[51] G. Angloher et al. "Description of CRESST-II data". In: *arXiv* (2017). URL: `https://doi.org/10.48550/arXiv.1701.08157` (cit. on pp. 32, 56).

[52] A Naamad, D.T Lee, and W.-L Hsu. "On the maximum empty rectangle problem". In: *Discrete Applied Mathematics* 8.3 (1984). URL: `https://www.sciencedirect.com/science/article/pii/0166218X84901240` (cit. on p. 48).

Appendix

A Recommended Workflow with the *ModeLimit*

In this section, an optimal workflow for the *ModeLimit* Python package will be explored. The focus lies in setting upper limits for cross sections related to different dark matter masses. A clear and efficient workflow is key here. This approach not only makes the most of the *ModeLimit* package but also helps get valuable insights from the data.

1. Use the approximation for the optimum interval method to see the range of needed μ-values for setting proper upper limits.

2. Tabulate new data with very dense geometrically spaced μ-values with a high number of lists consisting of random number for each μ using the information on the μ-range acquired in step 1. This step will cost time but will also yield higher accuracy of results. Save the tabulated data. If the μ-values needed are high, using multiprocessing may save a considerable amount of time.

3. If you want to recalculate the limits for different masses or slight changes in the signal model, you can either use the saved data or use the approximation again. The former is recommended if the changes will not result in drastic changes of the μ-range. If it does, resetting the μ-range with the approximation is recommended.

4. If two-dimensional data is available, using the maximum patch method is an option.

Code examples for setting all the necessary parameters will be provided in the package to achieve a user-friendly and easy-to-use experience.