SAPIENZA
UNIVERSITÀ DI ROMA

Search for low-mass dark matter with direct detection experiments

Scuola di dottorato Vito Volterra

Dottorato di Ricerca in Fisica – XXXIV Ciclo

Candidate

Stefano Piacentini
ID number 1597381

Thesis Advisor

Prof. Andrea Messina

January 2022

Thesis defended on 31 May 2022
in front of a Board of Examiners composed by:

Prof. Irene Di Palma (chairman)
Prof. Riccardo Bellan
Prof. Maura Pavan

Search for low-mass dark matter with direct detection experiments
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LATEX and the Sapthesis class.

Version: 31 May 2022

Author's email: stefano.piacentini@uniroma1.it

Dedicated to
my family

Abstract

The present thesis, in the context of the direct detection of dark matter, focuses on the search for low mass dark matter candidates. We set the most stringent bound on the dark matter spin independent cross section in the GeV/c^2 mass range using the DarkSide-50 experiment. We extend the bound of the experiment in the sub-GeV region thanks to the exploitation of the Migdal effect. Finally, we investigate the possibility of measuring such an effect with the CYGNO experimental approach.

At present, dark matter is only known via gravitational effects and its nature, e.g. its interactions with ordinary matter or its mass, has not yet been discovered. The search for dark matter weakly interacting massive particles with noble liquids has probed masses down and below a GeV/c^2. Detecting the scattering of dark matter particles in the sub-GeV "low mass" range is a challenging task, since, in this mass region, the typical energy transfer is below the experimental threshold. However, the nuclear recoils induced by dark matter particles interacting in the detector can be followed by the excitation and ionization of the recoiling atom, the Migdal effect.

We improve the preliminary results of the 2018 DarkSide-50 analysis thanks to a refined detector response model, data selection, and background model. We develop an innovative analysis tool in the Bayesian approach that is able to reproduce the detector response in a semi analytical way. As a result of the analysis, the DarkSide-50 observed sensitivity turns out to be the world's best sensitivity in the $m_{DM} = [0.7, 4.7]$ GeV/c^2 mass region, being able to exclude at 90% credible interval a dark matter spin independent (SI) cross-section $\sigma_{DM}^{SI} = 2.1 \times 10^{-43}$ cm^2 at $m_{DM} = 4.5$ GeV/c^2.

We compute the rate of the so-called Migdal effect for hydrogen, helium and for argon atoms: thanks to the inclusion of this effect, the sensitivity of DarkSide-50 can be extended to lower masses, down to 60 MeV/c^2, where the experiment has currently the strongest sensitivity.

Up to this day, the Migdal effect has been measured only in α and β decays, and its verification in nuclear scattering would be a strong confirmation of the results obtained by the direct detection experiments. For this purpose, we investigate the possibility of measuring the Migdal effect induced by a neutron source with a prototype of the CYGNO experiment. With its light target atoms and great 3D position resolution, CYGNO will be a complementary experiment with respect to DarkSide, being able to access the direction of the nuclear scatterings and potentially distinguish the dark matter induced recoils from the irreducible neutrino coherent scattering background.

Contents

Introduction

Since the 1930s, we have collected an incredible amount of astronomical and cosmological evidences suggesting that there is something we are missing in our description of the Universe. The available models for gravitation are not consistent with the observed data on every scale. The two main explanations for these inconsistencies are that the current gravitational models do not describe precisely our universe, or that there is a massive particle, or a set of particles, called "dark matter", that we are still missing and that we have not yet detected.

The observations are quite difficult to reconcile with modified gravity theories, and this is the reason why the scientific community has shown a great and growing interest in the detection of possible dark matter particles in the last decades. Indeed, establishing the nature and the properties of dark matter is one of the most open and compelling problems of contemporary physics and, up to this day, no incontrovertible evidence of dark matter particles has been found yet.

Several and very different experimental techniques have been developed to be able to detect dark matter on Earth. These experimental approaches can be divided mainly in three categories: accelerators can look for events in which the interaction between ordinary matter particles results in the production of dark matter particles; on the contrary, the so-called "indirect detection" techniques look for events in which interacting dark matter particles could produce ordinary matter particles; finally, the so-called "direct detection" experiments study the possible recoils that the dark matter particles should induce in the constituents of the ordinary matter, in particular mainly in the atomic nuclei. In fact, as suggested by the astronomical data and as described by the so-called "Standard Halo Model", a dark matter halo permeates the Milky Way, and a dark matter wind coming from the direction of the Cygnus constellation hits the Solar System during its motion around the galactic center.

My research activity focused on the dark matter direct detection. In this field, many experimental approaches have been developed, each of them giving their best performances for different possible values of the dark matter mass. However, the experimental approaches producing the best results in the most theoretically motivated dark matter mass region, namely the electroweak scale at $\mathcal{O}(1 \text{ GeV}/c^2 - 1 \text{ TeV}/c^2)$, are based on Time Projection Chambers (TPC) using noble liquids or gas. These experiments are characterized by a very strong sensitivity, being able to exclude incredibly small dark matter interaction cross-sections.

As we will see in more detail in this thesis, the highest sensitivity to the nuclear recoil interactions of this kind of detectors is reached for dark matter masses of the same order of magnitude of the mass of the target nuclei. For this reason, among all

the experimental results in the direct detection field, experiments like XENON1T, based on liquid Xe, have the best sensitivity in the $\mathcal{O}(100\,\text{GeV}/c^2)$ region, while experiments based on liquid Ar, such as DarkSide-50, have the strongest sensitivity in the $\mathcal{O}(5\,\text{GeV}/c^2)$ region. The GeV and sub-GeV regions, usually referred to as the "low mass" region, even if still theoretically motivated, are in principle inaccessible by these experiments, and are typically explored by semiconductor devices and bolometers. Their sensitivity to lower dark matter masses is indeed limited by the lower energy threshold in the detection of nuclear recoils, which is typically expected to be $\mathcal{O}(\text{keV})$, corresponding to $\mathcal{O}(1\,\text{GeV}/c^2)$ dark matter interacting with noble liquids such as argon or xenon.

Within this context, the goal of this thesis is to contribute to the search for dark matter candidates, with particular focus on the low mass region. In particular, we contributed to the conclusive analysis of the DarkSide-50 experiment, improving the 2018 DarkSide-50 previous analysis [1] by means of a better description of the detector response model, the data selection, and the background model. We developed an innovative way to include the detector response model in a semi analytic way inside the likelihood and we carried out the analysis in the Bayesian approach. DarkSide-50 is a dual phase TPC based on liquid argon which collected data underground at the Laboratori Nazionali del Gran Sasso (LNGS) from 2015 until 2019, for a total exposure of $(9968 \pm 150)\,\text{kg d}$. The readout system consists of two arrays of 19 PhotoMultiplier Tubes (PMTs), placed at the top and the bottom caps of a cylindrical TPC, that are able to collect, as a consequence of a particle releasing energy inside the detector, both the scintillation light (S1 signal) and the ionization (S2 signal), amplified in the gaseous phase. Thanks to the deeper comprehension of the detector response and background models, we were able to measure, with our innovative approach, the sensitivity of DarkSide-50 to dark matter spin independent interactions: the experiment turned out to set, up to this date, the most stringent bound in the $[0.7, 4.7]\,\text{GeV}/c^2$ dark matter mass region. In addition, thanks to the parallel phenomenological studies we carried out, we were able to extend this bound to lower dark matter masses, down to $60\,\text{MeV}/c^2$, where, again, DarkSide-50 revealed to currently set the world's strongest bound.

Indeed, in the phenomenological analysis of Ref. [2], we demonstrated that it is possible to extend the sensitivity of liquid argon experiments to lower masses taking into account a small effect, called the "Migdal effect", in addition to the usual nuclear recoil analysis. To be more specific, the Migdal effect is the emission of an electron as a direct consequence of the perturbation of the electron cloud induced by the recoil of the atomic nucleus. We computed the Migdal effect rate for hydrogen, helium and, using the argon ionization probabilities computed in Ref. [3], argon atoms: we found that, particularly for argon, it is a negligible effect when the nuclear recoil induced by the dark matter has enough energy to be detected, but, when this is not the case, it turns out to be the dominant contribution, since the emitted electron is detectable even when the nuclear recoil is not.

Due to its importance in this research field, a growing interest in the measurement of the Migdal effect is arising. Indeed, up to this day, it has been measured only in α and β decays, and the observation of the Migdal effect in nuclear scattering would be a strong confirmation of the possibility to exploit it by the direct detection experiments. For this purpose, we believe in the possibility of measuring the Migdal

effect induced by a neutron source with a CYGNO prototype. The CYGNO experimental collaboration has the goal of constructing a gaseous TPC detector with optical readout for the high precision 3D tracking of $\mathcal{O}(1\ \mathrm{GeV}/c^2)$ and sub-GeV dark matter induced nuclear recoils. To enhance the sensitivity in this mass region, the experiment, which is concluding its R&D phase, will be therefore based on a gas mixture containing light atoms like helium and fluorine. The readout system, consisting of a triple Gas Electron Multiplier (GEM) stage optically coupled to an sCMOS camera, will allow, thanks to its great spatial resolution, measuring the directionality of the recoils, being potentially able to confirm the astrophysical source of a potential signal or to discriminate the signal from the irreducible coherent neutrino scattering background. Due to its spatial resolution power, this experimental approach could be in principle able to resolve the possible Migdal effect signatures when exposed to a suitable neutron source.

Finally, the analyses performed in this thesis are carried out in the Bayesian approach, generally rarely used by the experimental collaborations in the dark matter direct detection field. In particular, the innovative approach developed for the DarkSide-50 analysis was implemented in a C++ software, exploiting the BAT (Bayesian Analysis Toolkit) libraries and the NVIDIA CUDA libraries for GPU acceleration. A simplified version of this approach has also been used to preliminarily assess the sensitivity for a possible future CYGNO detector, and for a phenomenological study [4] of the data published by the DAMA experimental collaboration.

The thesis is structured as follows.

In Chapter 1 the most significant dark matter evidences and possible particle candidates are introduced. In addition, the standard model describing the galactic dark matter halo in the Milky Way is also described.

In Chapter 2 the main features of the modern direct detection searches are illustrated. A theoretical description of the typical dark matter-nucleus interaction process is given, and the state-of-the-art of this research field is reported.

Chapter 3 focuses on the Migdal effect: an analytical calculation for hydrogen and helium atoms is presented. Up to our knowledge, this represents the first explicit calculation of the rate of the Migdal process for a helium target. Then, following our phenomenological study of Ref. [2], the role of this effect in direct detection searches is explored, showing its importance when an argon target is used.

In Chapter 4 the statistical formalism of the Bayesian approach, used in the subsequent analyses, is shown. In this chapter, the analysis of the DAMA residuals we carried out in Ref. [4] is illustrated.

In Chapter 5 the DarkSide-50 experimental setup is described, and the detector response model, the calibration measurements, and the background model are presented.

Chapter 6 is dedicated to the conclusive analysis of the DarkSide-50 experiment: the innovative implementation of the detector response model is shown, along with the full detailed description of the experimental likelihood. The results of the analysis and the bound on the dark matter spin independent cross section are presented here in detail.

In Chapter 7 the CYGNO experimental approach is described, with a particular focus on the features that make CYGNO a suitable experiment to measure the Migdal effect in nuclear scattering. In particular, the performances of the current

prototype, the LEMON detector, are presented, and the details of the CYGNO plan for the construction of a large TPC detector will be given.

Finally, Chapter 8 focuses on the future perspectives for this work. In particular, after a brief description of the available sensitivity projections for the future upgrades of the CYGNO and DarkSide projects, in the last section, a particular emphasis is put on the possibility of measuring the Migdal effect with a CYGNO prototype.

Chapter 1

The dark matter hypothesis

As suggested by the recent astronomical and cosmological observations [5], most of the matter in our Universe, approximately 80%, is made of an unknown source called dark matter (DM). Its existence is postulated solely by means of its gravitational effects which affect the galaxy rotation curves, galaxy clusters, gravitational lensing, Cosmic Microwave Background (CMB), and large-scale structure. Great effort has been made in recent years to determine the nature of DM, both from the theoretical and experimental point of view. Several models have been proposed to identify the most suitable particle candidate, and a lot of experimental approaches have been pursued. This thesis focuses on the search for dark matter candidates with direct detection experiments.

In this introductory chapter, a brief summary of the main DM evidences and candidates is given in Sec. 1.1 and Sec. 1.2, respectively. In Sec. 1.3 a description of the standard assumptions on the DM distribution in the Milky Way, that will be one of the inputs needed for the next chapters, is reported.

1.1 Evidences for dark matter

1.1.1 Galaxy rotation curves

One of the most known evidences for the existence of DM comes from the study of the rotation curves of the spiral galaxies [6,7]. This class of galaxies, which the Milky Way belongs to, is characterized by a flat disk mainly made of stars and gas clouds arranged in spiral arms and rotating in circular orbits, and by a central agglomerate of stars called "bulge". The rotation curve is defined as the relation between the circular speed of the stars in the galaxy, that can be measured using the Doppler shift of the Hydrogen 21 cm line, and the distance of these stars from the galactic center. Using Newton's law of gravity and the Gauss theorem, it can be easily shown that the circular speed v of a star around the galactic center is given by

$$v = \sqrt{\frac{GM(R)}{R}}, \tag{1.1}$$
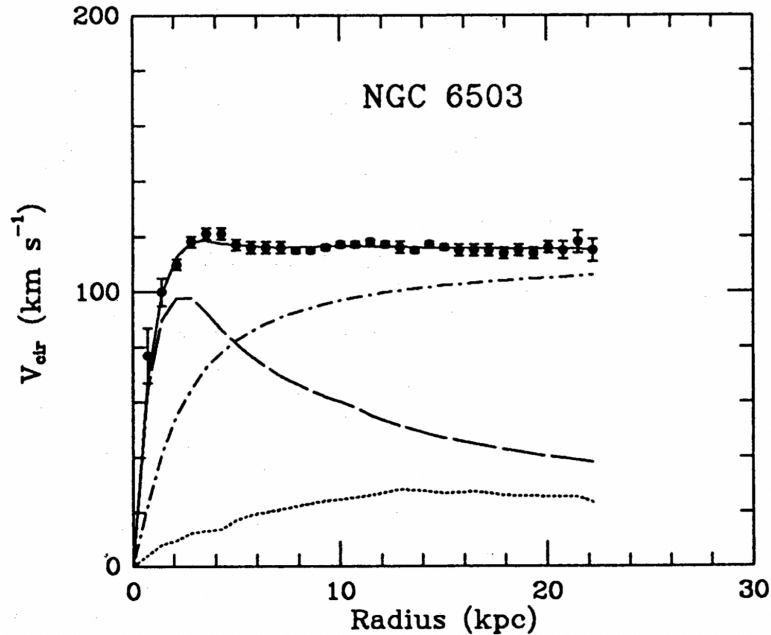
Figure 1.1. Rotation curve of the galaxy NGC6503 from Ref. [8]. The lines correspond to the fitted visible components (dashed), gas component (dotted), DM halo component (dash-dotted), and to the sum of the three (solid).

where R is the distance from the galactic center, G is the gravitational constant and

$$M(R) = \int_{|\boldsymbol{r}|<R} \rho(\boldsymbol{r}) \, d\boldsymbol{r} \qquad (1.2)$$

is the mass contained within a distance R from the galactic center, with $\rho(\boldsymbol{r})$ being the visible mass density profile as a function of the position $\boldsymbol{r}$ with respect to the galactic center. At large R, namely in the region where the mass density ρ is expected to be zero, $v \propto 1/\sqrt{R}$, and this is what is observed for example for the planets in the Solar System. However, on the galactic scale, the rotation curve is observed to be quite flat at large R, as shown in Fig. 1.1. This observation can be explained by an invisible DM halo, with a density profile at large distances $\rho(\boldsymbol{r}) \propto r^{-2}$, which permeates the spiral galaxies.

In general, however, the rotation curve at large distances is not always flat [9], and the DM halo density profile dependence on r can be different from r^{-2} at small or very large radii [10, 11].

Finally, an additional evidence for DM related to the velocities of stars in galaxies comes from the so-called "bar instability" [12] in disk galaxies. Disk galaxies consist of a flattened circular disk of stars. However, it has been shown by numerical simulations that self-gravitating disks are not stable and tend to form bar-like objects. However, adding a massive, more extended, spherical halo keeps the disks stable.

1.1.2   Galaxy Clusters, Gravitational Lensing and Bullet Cluster

Galaxy clusters are the biggest gravitationally bound structures in the Universe, containing $\mathcal{O}\left(10^{2\div3}\right)$ galaxies and hot gas. Since they are so largely extended, they can be considered as representative of the Universe in general, and, by studying some of their properties, it is possible to get evidences of DM.

One of the first hints for DM has been in fact the study of the total mass of the clusters [13, 14]. To get a first rough estimation, using the virial theorem, it is in fact possible to express the mass of the cluster as

$$M = R_G \frac{\langle v^2 \rangle}{G}, \tag{1.3}$$

where $\langle v^2 \rangle$ is the mean square velocity of the cluster, measured via Doppler effect, and the gravitational radius R_G is defined as

$$R_G = 2 \left(\sum_i m_i \right)^2 \left(\sum_i \sum_{j \neq i} \frac{m_i m_j}{r_{ij}} \right)^{-1}, \tag{1.4}$$

measured by reconstructing the positions of the galaxies of the cluster. For a typical cluster, the mass to luminosity ratio is

$$\frac{M}{L} \sim 10^2 \frac{M_\odot}{L_\odot}, \tag{1.5}$$

with $M_\odot$ and $L_\odot$ indicating the mass and the luminosity of the Sun, respectively. The large $\mathcal{O}\left(10^2\right)$ factor can be explained by the presence of invisible matter widespread within the cluster, and it is equivalent to a total mass density $\rho_m \sim 0.3\rho_c$ [15], where ρ_c is the critical density defined as

$$\rho_c = \frac{3H_0^2}{8\pi G}, \tag{1.6}$$

with H_0 being the Hubble constant today.

Another observable that can be measured studying galaxy clusters is the baryon fraction f_b, defined as

$$f_b = \frac{M_b}{M}, \tag{1.7}$$

where M is the total mass of the cluster, and M_b is the mass of the normal baryonic matter. Assuming that the intergalactic hot X-ray gas is spherically symmetric and in hydrostatic equilibrium, one can estimate the baryon fraction by measuring the properties of the X-ray spectrum of the gas obtaining $f_b \sim 0.144 \pm 0.005$ [16], which is, therefore, a small fraction of the total mass of the cluster.

Further evidences of DM come from the so-called "gravitational lensing". According to General Relativity, in fact, massive objects modify the local geometry of space-time. Therefore, a photon traveling in the curved space around a massive object - mainly a galaxy or a galaxy cluster - is deflected and, as a consequence, it may happen that the image of a star or a galaxy which is behind the object appears distorted to the observer, or even duplicated in multiple copies. This effect, whose

intensity depends on the gravitational potential and increases as the mass of the massive object does, is known as gravitational lensing. According to the scale of the mass involved in the process, different regimes are defined. When the light is deflected by large angles it is referred to as strong lensing [17]: multiple images, or even an Einstein ring, if the light source and the massive object are aligned with the observer, can be formed. Weak lensing [18], instead, occurs when the light is deflected by small angles, and the image of the light source appears to be distorted. Finally, microlensing [19] occurs when the deflection is too small to be resolved by the instruments of the observer: the result is that the light source is temporarily brightened. The crucial aspect is that, being a pure gravitational effect, both visible and dark matter give their contribution to it. Indeed, by measuring the intensity of the gravitational lensing, the mass of the massive object can be estimated; such observations when compared to the mass of the visible matter composing the object indicate a lot of missing mass. In other words, the mass of the stars and the gas composing the galaxies and the galaxy clusters is not enough to explain such big observed gravitational lensing effects, and this is considered to be a strong evidence of DM.

Finally, a famous physical system related to galaxy clusters with strong evidences of dark matter is the so-called Bullet Cluster [20] - see Fig. 1.2. The picture shows two clusters right after their collision: this is a very peculiar observation, since the different behavior of the various components of the clusters - stars, intergalactic hot gas and hypothetically DM - can be studied. Since the probability of a clash between the stars is very low, the stellar component is basically unaffected by the clash, as observed. On the other hand, the intergalactic gas interacts, mainly electromagnetically, and we can detect the resulting X-ray emission (pink halo in the picture): these interactions have the effect of slowing down the gas with respect to the motion of the other components of the cluster pair. However, the gravitational lensing effect does not follow the baryonic matter, whose main component is gas itself, but instead appears unchanged by the clash. This strongly supports the idea of the existence of a non-baryonic DM halo in the two clusters (blue halo in the picture), which, unlike the gas, interacts very weakly, except for gravity.

1.1.3   Cosmic Microwave Background

The most commonly endorsed model of Cosmology today is the Big Bang model. The simplest parameterization of the Big Bang model, which is consistent with the majority of the astronomical observations, is the six-parameters ΛCDM (Lambda Cold Dark Matter) model [22]. According to ΛCDM, the Universe is made of three main components: a cosmological constant Λ related to dark energy, the ordinary matter and a cold dark matter (CDM) component. The reason why it is called "cold" dark matter is because, in this model, DM is assumed to be non-relativistic at the time of decoupling. This assumption is motivated by large scale structure properties, as we will briefly see in Sec. 1.1.4.

One of the fundamental observational bases of the ΛCDM model, which is considered to be the Standard Model of contemporary Cosmology, is the Cosmic Microwave Background. However, before going into the details of its importance as an evidence for dark matter, let's make a brief summary description of the evolution of
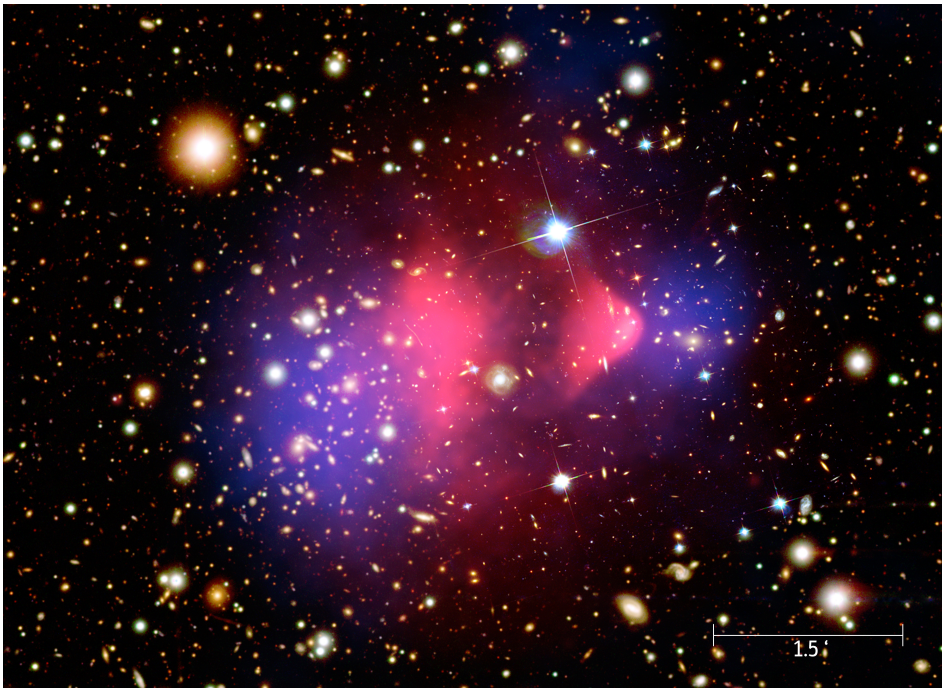
Figure 1.2. Picture of the cluster merger 1E0657-558 ($z = 0.296$) named Bullet Cluster. The pink halo represents the X-ray emission, generated by the intergalactic gas. The blue halo is the mass distribution evaluated from gravitational lensing. Figure from Ref. [21].

the Universe. The Friedmann equation describes how the expansion of the Universe is related to its energy content and its geometry and can be written as

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}, \tag{1.8}$$

where $H(t)$ is the Hubble constant at a certain time t, $a(t)$ is the scale factor and parameterizes the expansion of the Universe, ρ is the total energy density, k/a^2 is the spatial curvature of the Universe and Λ is the cosmological constant. We have followed the convention of setting the speed of light $c = 1$. The scale factor is often normalized to unity today, namely $a_0 = 1$, and the Hubble constant today is often written as $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$, with h being a dimensionless constant parameterizing the uncertainty. Indeed, different kind of observations find different values of Hubble constant today - this is known as "Hubble tension" - and using h is useful to propagate the uncertainty on the Hubble constant today to the other parameters of the model. Just to give the numerical order of magnitude for h, for example Planck measurements of the CMB give $h = 0.673 \pm 0.006$ while measurements from type 1a Cepheid supernovae show $h = 0.740 \pm 0.014$ [23]. The total energy density ρ_c for which the Universe is flat, namely $k = 0$, is called critical density, and it is defined as in Eq. 1.6. It is therefore useful to define the so-called density parameter Ω such that

$$\Omega = \frac{\rho}{\rho_c}. \tag{1.9}$$

Therefore in a flat Universe $\Omega = 1$, and this is consistent with the current astronomical observations [5]. Since however ρ takes contributions from the various components of the Universe, one can define each single density parameter as

$$\Omega_i = \frac{\rho_i}{\rho_c}, \qquad (1.10)$$

where the index i indicates the i-th component. These components are: the energy density of all the relativistic particles, called radiation density ρ_r; the energy density of the matter ρ_m which in turn can be separated in a baryonic matter energy density ρ_b and a CDM energy density ρ_{CDM}; the energy density of the cosmological constant ρ_Λ.

According to the current accepted explanation of the evolution of the Universe, for significant periods of time the total energy density has been dominated by one component [24]. When, in the initial phases, the Universe was hot and dense and the temperature was greater than $k_B T > 3$ eV, the Universe was dominated by radiation, which, as the Universe expands, dilutes as $\rho_r \propto a^{-4}$. The scale factor in this period, as described by the Friedmann equation, evolved as $a \sim \sqrt{t}$. At a time of $\sim 10^4$ y after the Big Bang, corresponding to a temperature $k_B T \sim 3$ eV, ρ_m started to be greater than ρ_r, and there was a phase in which the Universe was dominated by matter. Since $\rho_m \propto a^{-3}$, during this phase, the scale factor evolved as $\rho_m \propto t^{2/3}$. Finally, the matter started to dilute, and the cosmological constant became dominant, and during this period the scale factor began to grow exponentially with time. If we define the redshift z by means of the relation

$$a = \frac{1}{1+z}, \qquad (1.11)$$

we can rewrite the Friedmann equation as

$$H^2 = H_0^2 \left(\Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0} \right), \qquad (1.12)$$

where the Ω_k term accounts for the curvature term in Eq. 1.8.

At this point, we have all the elements to understand the existence and origin of the Cosmic Microwave Background. As already said in the first phases the Universe was hot and dense. When matter domination began, the Universe was a plasma of electrons and baryons in thermal equilibrium. As the temperature decreased, the initial neutral atoms started to form, and the Universe became transparent to photons, which are coupled to charged particles. At redshift $z \sim 1100$, radiation and matter were essentially decoupled, and the photons started to travel freely in space. Those photons are still streaming in the Universe today and are called Cosmic Microwave Background.

The CMB appears to be a black body radiation with a temperature of $T_0 = (2.7255 \pm 0.0006)$ K [25]. However, it is not perfectly uniform in all directions, but, instead, is characterized by small fluctuation of order $\Delta T/T \sim 10^{-5}$, as depicted in Fig. 1.3. These anisotropies are an incredible source of information on the formation of structures in our Universe and on the amount of DM. From the point of view of the intensity, the $\Delta T/T \sim 10^{-5}$ amplitude itself provides hints for DM. Indeed, the anisotropies are caused by density perturbations at the time of photon
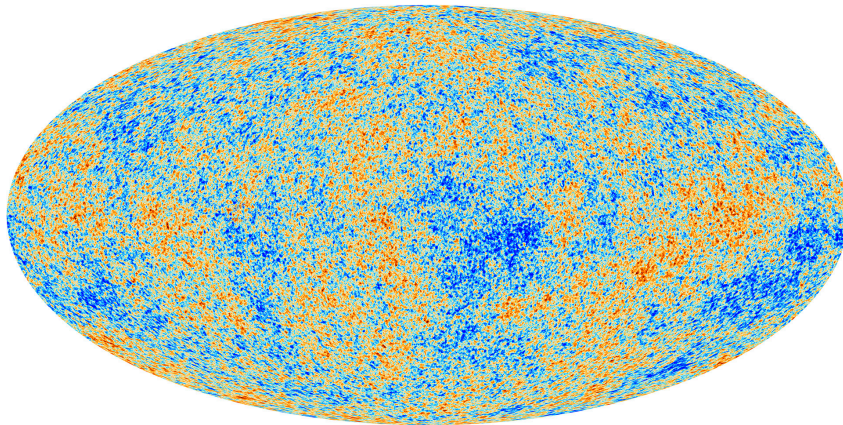
Figure 1.3. Cosmic Microwave Background anisotropies from the Planck satellite. All the distortions and spurious effects caused by the motion of the Earth with respect to the CMB galactic frame have been subtracted and corrected here. Figure by ESA https://www.esa.int/ESA_Multimedia/Images/2013/03/Planck_CMB

decoupling, and, studying the evolution of these perturbations, it can be probed that the requirement for the structures - galaxies, galaxy cluster etc. - to form in a Universe without non-baryonic DM is $\Delta T/T \sim 10^{-4}$ in the CMB.

In addition to that, other evidences for DM come from the angular properties of the CMB. In fact, this temperature fluctuations can be analyzed expanding the CMB in spherical harmonics:

$$\frac{\Delta T(\theta,\phi)}{\overline{T}} \equiv \frac{T(\theta,\phi) - \overline{T}}{\overline{T}} = \sum_l \sum_{m=-l}^{m=l} a_{lm} Y_l^m(\theta,\phi), \tag{1.13}$$

where $T(\theta,\phi)$ is the temperature in a certain direction of the sky, $\overline{T}$ is the mean temperature, $Y_l^m(\theta,\phi)$ are spherical harmonics and a_{lm} are the expansion coefficients. At this point the so-called power spectrum can be computed: it is a function of l and it is defined as

$$C_l = \left\langle |a_{lm}|^2 \right\rangle_m, \tag{1.14}$$

with the brackets indicating a simple average. The result is depicted in Fig. 1.4. The relative height of the peaks in the temperature power spectrum depends on the abundance of the different components of the Universe at the photon decoupling, and this gives therefore information on the total amount of DM; on the other hand the baryon matter density influences the ratio between the odds and the even peaks amplitudes. The fit of the CMB temperature power spectrum to the ΛCDM parameters performed by the Planck collaboration [5] gives

$$\begin{aligned} \Omega_b h^2 &= 0.02237 \pm 0.00015, \\ \Omega_{CDM} h^2 &= 0.1200 \pm 0.0012. \end{aligned} \tag{1.15}$$

Therefore not only Ω_{CDM} is greatly different from zero but, as expected from the other astrophysical observations described before, the DM is also the main ingredient of all the matter in the Universe.
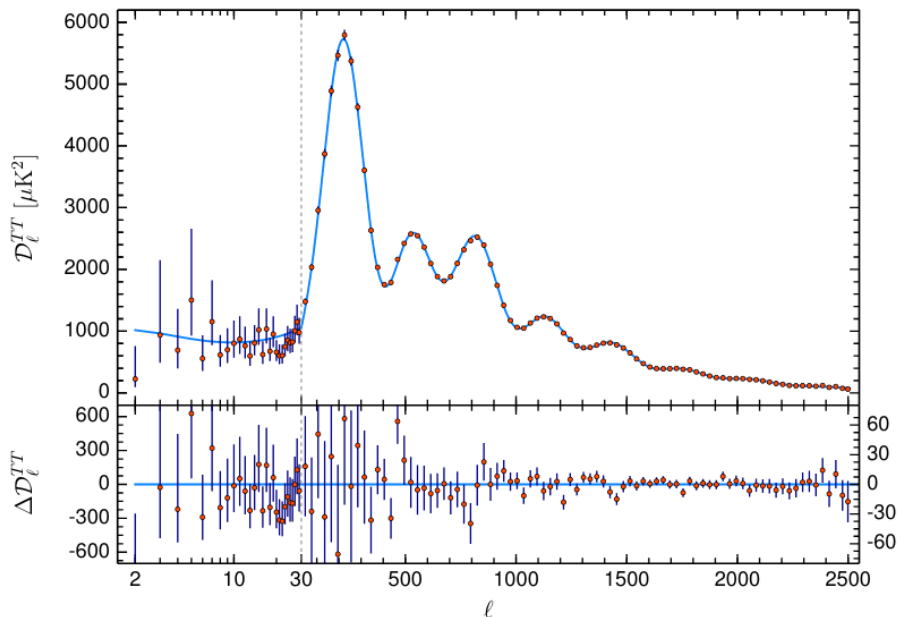
Figure 1.4. Planck 2018 temperature power spectrum (TT) [5]. The blue solid line is a fit to the ΛCDM theoretical prediction. The vertical gray dashed line at $l = 30$ denotes a change in the horizontal axis scale.

1.1.4 Large scale structure

The astronomical observations and mapping of the plentiful amount of objects in the Universe have allowed us to probe and analyze the various structures up to a very large scale. These structures seem to organize in a hierarchical model, which is in agreement with the hypothesis of a cold DM. Although the study of these structures are not as informative in quantitative and qualitative terms as the CMB anisotropies, combining different datasets can provide some constraints on the cosmological parameters. The latest results from DES (Dark Energy Survey) [26], for instance, combining observations from galaxy-galaxy lensing, galaxy clustering and weak lensing, provide $\Omega_m = 0.34 \pm 0.03$, which, combined with other datasets such as, for instance, Planck, BAO (Baryonic Acoustic Oscillations) or BBN (Big Bang Nucleosynthesis), yields $\Omega_m = 0.306^{+0.004}_{-0.005}$.

1.2 Dark matter candidates

In the previous section, some of the main evidences of DM have been discussed. A broad list of theoretical models has been proposed as a solution for those empirical observations, but, however, if DM is actually made of particles it is possible to summarize the common properties that it has to satisfy. In particular, all these models must provide a particle which:

- is not baryonic;

- is stable on a cosmological timescale: we currently have evidences of the existence of DM today, so it is not decayed yet;

- is neutral, since it does not undergo electromagnetic interactions;

- has no color charge, because it does not feel QCD strong interactions;

- if coupled to ordinary matter except for gravity, has at most weak interactions with the Standard Model particles.

In addition to that, as mentioned in Sec. 1.1.4, most of the dark matter has to be "cold", i.e. non-relativistic to explain large scale structure. There exists a set of models proposing "hot", i.e. relativistic, dark matter, even if, however, its contribution must be small, at most $\sim 1\%$, with respect to the whole amount of DM. Since none of the known Standard Model (SM) particles has all the previous properties at the same time, many extensions of the SM have been proposed. In the rest of the section more details will be given on some of the main possible proposed DM candidates: greater emphasis will be given to WIMPs (Weakly Interacting Massive Particles), and mentions will be devoted to lighter DM candidates, i.e. axions and sterile neutrinos.

1.2.1 WIMPs

Weakly Interacting Massive Particles is probably the most studied class of particle DM candidates both from theoretical and experimental point of view. The reason is that, thanks to the so-called "WIMP miracle", they arise in a natural way in many theories and naturally have the correct cosmological properties. Regarding their general features, they are stable, neutral particles with a mass that lives in the $\mathcal{O}(1\,\text{GeV} - 10\,\text{TeV})$ range. Usually named χ, they interact with the Standard Model particles at the weak scale and were not relativistic at the photon decoupling. One of the most famous candidate as a WIMP is the neutralino of Supersymmetry: nonetheless, even if its mass is expected to be close to weak scale ($m_{weak} \sim 100 GeV$) the accelerator searches have only given null results so far [27]. Many other WIMP candidates have however been proposed [28], but since this thesis will consider the general features of these candidates, giving more details goes beyond the scope of this work.

The WIMP miracle

The dark matter we observe today may be produced in the early Universe, as a thermal relic of the Big Bang [29]. At those stages, as already mentioned, the Universe was a hot and dense plasma of particles assumed to be in thermal equilibrium. The WIMP DM is generally assumed to undergo 2-to-2 interactions with the SM particles, namely the following interaction can happen

$$\chi + \overline{\chi} \leftrightarrow \phi\,\phi^*, \tag{1.16}$$

where $\phi\,\phi^*$ indicates a couple of SM particles, i.e. leptons, quarks or electroweak bosons. This process, at high temperatures, proceeds in both directions. Nonetheless, since the number density of DM particles in equilibrium depends exponentially

on the temperature as $e^{-m_\chi/k_B T}$, as the temperature of the Universe decreases below the WIMP mass m_χ the number of DM particles becomes exponentially suppressed. Parallel to this, the Universe expands, and, from a certain moment on, it becomes so large, and the DM becomes so dilute, that the DM annihilation process turns irrelevant. The transition between the exponential fall regime and the expansion dominated regime is called "freeze out".

From a quantitative point of view, the time evolution of the DM number density n_χ is described by the Boltzmann equation

$$\frac{dn_\chi}{dt} = -3Hn_\chi - \langle \sigma_A v \rangle \left(n_\chi^2 - n_{\chi,eq}^2 \right), \tag{1.17}$$

where H is the Hubble parameter, $\langle \sigma_A v \rangle$ is the thermal average of the annihilation cross-section, and $n_{\chi,eq}$ is the DM number density in equilibrium. If we focus on the right-hand side of Eq. 1.17, the first term implements the dilution from the expansion of the Universe, the n^2 term accounts for the annihilation process, while the n_{eq}^2 term accounts from the reverse production process. In literature, instead of the number density n_χ, it is often referred to DM yield Y_χ, defined as

$$Y_\chi \equiv \frac{n_\chi}{s}, \tag{1.18}$$

where s is the entropy density. Assuming thermal equilibrium, a gas state equation for the matter plasma, and using the Second Law of Thermodynamics, the entropy density can be expressed as

$$s = \frac{2\pi^2}{45} g_* T^3, \tag{1.19}$$

where T is the temperature and g_* counts the number of degrees of freedom of the different species in the gas. Since in thermal equilibrium the total entropy is conserved, the entropy density is a measure of the comoving volume of the Universe, $s \propto a^{-3}$. Therefore, the DM yield Y_χ is also known as comoving number density of the DM. Fig. 1.5 shows the solution of Eq. 1.17 in terms of the comoving number density as function of the temperature (and time), assuming a WIMP mass $m_\chi = 100\,\text{GeV}$. As explained, there's an exponential fall which, after the freeze out transition, turns into a constant behavior.

It appears clear at this moment that the freeze out event is crucial. From a quantitative point of view, the freeze out is the moment in which the production and the dilution terms in Eq. 1.17 are comparable, namely

$$n_\chi^{FO} \langle \sigma_A v \rangle \simeq H, \tag{1.20}$$

where with the index FO we indicate that a certain quantity is evaluated at freeze out.

At this point we can write the DM density parameter today as

$$\Omega_\chi h^2 \equiv \frac{\rho_\chi^0}{\rho_c} h^2 = \left(\frac{s_0 h^2}{\rho_c} \right) m_\chi Y_\chi^0, \tag{1.21}$$

where, as usual, we denote with index 0 the quantities evaluated today, and where, in the last equality, we used the definition of Eq. 1.18. However, using the freeze
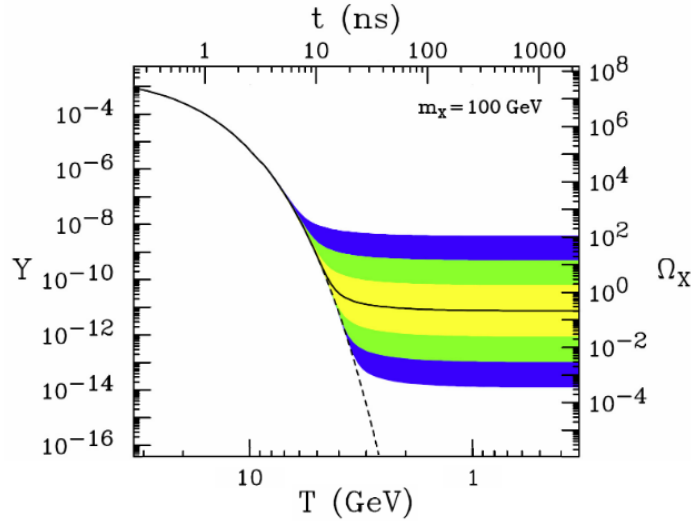
Figure 1.5. Comoving number density Y (and correspondent WIMP density parameter Ω_χ) as a function of the temperature (and time) for a WIMP mass $m_\chi = 100$ GeV. The solid black line is the solution of the Boltzmann equation for a cross-section that yields the observed DM density parameter. The colored regions correspond to a value of the cross-section that differs by a factor of 10, 100 or 1000 from this value. The dashed black line represents the evolution of an equivalent particle that remains instead in thermal equilibrium. Figure from Ref. [28].

out assumption, $Y_\chi^0 = Y_\chi^{FO}$ holds, and plugging Eq. 1.20 into Eq. 1.21 we obtain

$$
\begin{aligned}
\Omega_\chi h^2 &= \left(\frac{s_0 h^2}{\rho_c}\right) m_\chi Y_\chi^0 = \left(\frac{s_0 h^2}{\rho_c}\right) m_\chi Y_\chi^{FO} = \\
&= \left(\frac{s_0 h^2}{\rho_c}\right) m_\chi \frac{n_\chi^{FO}}{s^{FO}} \simeq \left(\frac{s_0 h^2}{\rho_c}\right) m_\chi \frac{H^{FO}}{s^{FO} \langle \sigma_A v \rangle} \propto \frac{1}{\langle \sigma_A v \rangle},
\end{aligned}
\tag{1.22}
$$

where in the third equality we used the definition of Eq. 1.18 at freeze out. By putting the numbers into the last equation we obtain

$$
\Omega_\chi h^2 \simeq \frac{3 \cdot 10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle \sigma_A v \rangle}.
\tag{1.23}
$$

Neglecting phase space and order one factors, and assuming a weak coupling for the annihilation interaction, we can roughly estimate the WIMP cross-section as

$$
\langle \sigma_A v \rangle \sim \frac{\alpha_W^2}{m_\chi^2}.
\tag{1.24}
$$

By using this estimation in Eq. 1.23, resolving for m_χ, and assuming the desired Ω_χ, we obtain $m_\chi = \mathcal{O}(1 \text{ TeV})$. This is what is called the "WIMP miracle": weak-scale particles freeze out with the desired DM density factor [28].

Even if the WIMP paradigm is so appealing, no signal has been discovered yet in the expected $\mathcal{O}(1 \text{ GeV} - 10 \text{ TeV})$ mass region - except for the DAMA/LIBRA

experiment [30–32]. The failure of observing incontrovertible evidences of a WIMP-like DM signal stimulated the effort of exploring different mass ranges [33], with particular attention to the lower mass region. Although in fact, by repeating the steps we showed in this section, low mass dark matter seems not to reproduce the right relic density, there exist theories in which, thanks to appropriate mechanisms, the dependence of the cross-section, and therefore Ω_χ, on the mass of the DM mass is weakened [34, 35]. In this context, both the CYGNO and DarkSide projects, which I am collaborating with, are interested in investigating the low mass region, and, in particular, one of the main goals of this thesis is the conclusive low mass analysis of the DarkSide-50 concluded experiment.

1.2.2 Axions

Axions are particles theorized by R. D. Peccei and H. R. Quinn [36] as an elegant solution to the so-called Strong CP problem of QCD (Quantum ChromoDynamics). QCD is the theory describing the strong interactions between quarks and gluons. It is an SU(3) gauge theory and has some important exact symmetries such as Lorentz invariance and CPT invariance. Besides these, it is characterized also by two approximated symmetries (let's just consider only the up u and down d quarks for simplicity):

- since $m_u \sim 2.5 \text{MeV}/c^2$ and $m_d \sim 4.9 \text{MeV}/c^2$ [37], their mass difference is very small if compared to the intrinsic QCD energy scale which is $\Lambda_{QCD} \sim 200 MeV$. Therefore, QCD is a theory in which $m_u \sim m_d$: if we impose this property, the theory acquires a new global U(2) symmetry in the flavor space. This symmetry can be decomposed as $\text{SU}_V(2) \times \text{U}(1)$, where the SU(2) corresponds to the up-down isospin symmetry, while U(1) corresponds to the up-down baryon number conservation.

- At the same time, it is also true that both the up and down masses are much smaller than Λ_{QCD}, meaning that QCD is a theory in which $m_u \sim m_d \sim 0$. If we assume this property, the theory now is symmetric under a $\text{U}_V(2) \times \text{U}_A(2)$, where the subscript V and A indicates that this global symmetry can be decomposed in a vector and axial part.

The interesting thing is that the axial part of this global approximate symmetry is not observed in nature, not even in an approximate way. The solution is that this $\text{U}_A(2) = \text{SU}_A(2) \times \text{U}_A(1)$ symmetry must be spontaneously broken: in fact, the vacuum of QCD breaks the symmetry by creating the so-called chiral condensate [38]. Therefore, by the Goldstone theorem, we should observe four bosonic nearly massless particles (the symmetry is also explicitly broken by the quark masses), namely the related Goldstone bosons. However, only the Goldstone bosons related to the $\text{SU}_A(2)$ part of the symmetry are observed - i.e. the three pions ($m_\pi \sim 130 \, MeV/c^2$, which is considered to be sufficiently smaller than Λ_{QCD}). The failure of observing a good candidate to be the Goldstone boson of the $\text{U}_A(1)$ part of the symmetry is called "$\text{U}_A(1)$ problem". The $\text{U}_A(1)$ problem can be solved by noting that $\text{U}_A(1)$ is an anomalous symmetry, i.e. its associated current is conserved in the classical approach, i.e. by the Noether theorem, but it is not if we include one-loop quantum

corrections. These corrections cannot be eliminated due to the topological non-trivial properties of the QCD vacuum [39], and they violate the CP symmetry. This CP violation, however, has not been observed so far. This problem is known as "Strong CP problem".

The solution proposed by Peccei and Quinn is to add an additional $U_{PQ}(1)$ chiral symmetry to the Standard Model. If spontaneously broken, the void expectation value of this symmetry can dynamically remove the CP violating contributions. However, the price to pay for this cancellation is that, due to the anomaly, the Goldstone boson associated $U_{PQ}(1)$ acquires a mass. This new neutral massive bosonic pseudo-scalar particle is called axion.

Axions are promising DM candidates, and, with masses in the $\mathcal{O}\left(10^{-5} - 10^{-3}\,\text{eV}\right)$ range, can satisfy the appropriate stability requirements [40]. In the most common axion theories, they are coupled to electrons and photons, so they can be produced by astrophysical sources, mainly stars, and be in principle detected on Earth. Indeed, recently there has been great interest in the observed low energy electron recoil excess in the XENON1T experiment [41], and one of the possible explanations considered by the XENON collaboration has been a possible signal induced by the axions produced by the Sun.

1.2.3 Sterile Neutrinos

The Standard Model of particle physics provides three species of left-handed neutrinos, ν_e, ν_μ and ν_τ, which can interact with leptons via weak interactions. They have very small but non-zero masses, not greater than $\mathcal{O}(0.2\,\text{eV})$ [5, 42]. They are therefore expected to be relativistic at the decoupling. For this reason, the standard neutrinos are a hot DM candidate, and, as such, they could only be a small fraction of the total amount of DM in the Universe.

An interesting class of theories tries however to extend the set of neutrino particles adding another particle, the so-called "sterile neutrino", which is a singlet under the Standard Model gauge group - i.e. it has no weak, strong or electromagnetic interactions with the SM particles. If its mass is greater than $\sim 1\,\text{keV}$, as a particle it has all the good properties requested to be DM: it is electrically and color neutral, it becomes non-relativistic quite early (being a good candidate for cold DM), and, if the mixing angles with other neutrinos are very small, it is very weakly coupled with other particles and can be stable over cosmological scales [43]. However, since, even if with low probabilities, they can oscillate to standard neutrinos, they are not stable in an absolute way. Some of their decays can produce mono-energetic keV photon emission, which could be observable in galaxies or galaxy clusters due to the expected high DM density [44]. Indeed, in 2014, a XMM-Newton X-Ray telescope data analysis of 73 galaxy clusters observations showed an unidentified X-ray line at $\sim 3.6\,\text{keV}$ [45], showed in Fig. 1.6. This excess can be interpreted as a signal from the decay of a $\sim 7\,\text{keV}$ sterile neutrinos, but the debate on the goodness of this result, as well as other null results, is still ongoing.
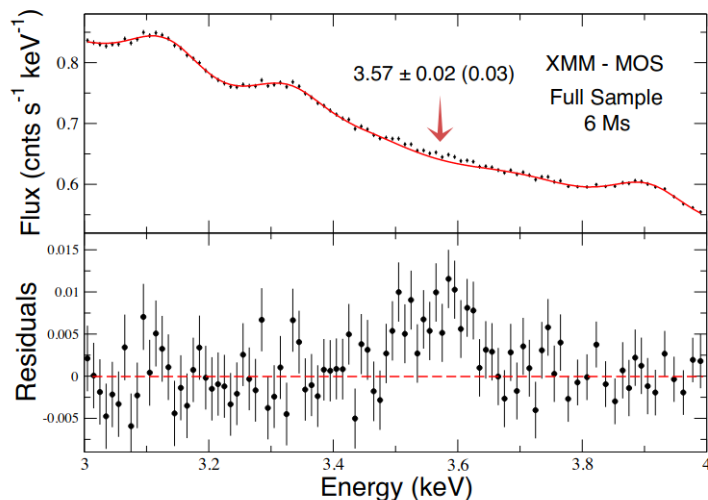
Figure 1.6. XMM-Newton $3 - 4 \, \mathrm{keV}$ spectrum obtained analyzing data from 73 galaxy clusters. The spectrum shows an unexplained peak at $\sim 3.6 \, \mathrm{keV}$. Figure from Ref. [45].

1.3   Dark matter local density and velocity distributions

The goal of this thesis is to contribute to the search for dark matter candidates with direct detection experiments. In fact, even if we have strong evidences of its existence, to confirm it we have to detect it. Since the experiments this work will deal with, namely DarkSide-50 and CYGNO, are located on Earth, in order to evaluate the possible DM signature we need to know how it is spatially and kinetically distributed. In particular, direct detection experiments aim to probe DM density and speed in the Solar system, which is located a $R_\odot = 8.2 \, \mathrm{kpc}$ from the galactic center of our galaxy, the Milky Way (MW).

The standard model describing the properties of the DM halo in the MW is called Standard Halo Model (SHM) [46, 47]. Let's consider the phase space distribution function of the DM particles $f(\boldsymbol{x}, \boldsymbol{v}, t)$, defined such that $f(\boldsymbol{x}, \boldsymbol{v}, t) d^3\boldsymbol{x} d^3\boldsymbol{v}$ gives the number of particles in an infinitesimal phase space volume $d^3\boldsymbol{x} d^3\boldsymbol{v}$ located at position $\boldsymbol{x}$ with speed $\boldsymbol{v}$ at time t. Assuming collisionless DM in steady state, the phase space distribution function can be found solving the collisionless Boltzmann equation

$$\frac{df}{dt} = 0. \tag{1.25}$$

By writing the total derivative in a more explicit way we obtain

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial \boldsymbol{x}} \cdot \boldsymbol{v} - \frac{\partial f}{\partial \boldsymbol{v}} \cdot \frac{\partial \Phi}{\partial \boldsymbol{x}} = 0, \tag{1.26}$$

where Φ denotes the gravitational potential. Assuming the DM halo is a self-gravitating system, where the potential is generated by its density distribution, Φ is related to the density by the Poisson's equation

$$\nabla^2 \Phi = 4\pi G \rho. \tag{1.27}$$

The SHM assumes an isotropic isothermal spherical DM halo with density profile $\rho(\boldsymbol{r}) \propto r^{-2}$. The solution of the collisionless Boltzmann equation in this case is the so-called Maxwell velocity distribution

$$f(\boldsymbol{v}) = \frac{1}{N}\left(\frac{3}{2\pi\sigma^2}\right)^{3/2} e^{-\frac{3|\boldsymbol{v}|^2}{2\sigma^2}}, \tag{1.28}$$

where N is just a normalization factor, and σ is the root-mean-square velocity dispersion. As a consequence of the isothermal hypothesis, the halo has a flat rotation curve, and the velocity dispersion σ is related to the circular velocity v_c of rotation of objects around the Galactic center by $\sigma = \sqrt{3/2}v_c$. In addition, the density distribution assumed by SHM extends formally up to infinity, and hence the velocity distribution does it too. However, the MW halo is finite, and, therefore, there exists a velocity $v_{esc}(r) = \sqrt{2|\Phi(r)|}$ at which a DM particle is not gravitationally bound to the MW. The usual way to implement this effect into the model is just truncating the Maxwell distribution at the local escape speed, namely, for our purposes, $v_{esc}(R_\odot)$.

Therefore, to summarize, the parameters we are interested in for direct detection searches on Earth are: the local DM density $\rho(R_\odot)$, the local circular speed $v_c(R_\odot)$ and the local escape speed $v_{esc}(R_\odot)$. The local density $\rho(R_\odot)$ is measured by using a large set of observations such as kinematics of nearby stars, MW rotation curve, velocity dispersion of halo stars, and so on. The standard parameter value used for direct detection searches is $\rho(R_\odot) = 0.3\,\mathrm{GeV\,cm^{-3}}\,c^{-2}$, even if there are large systematic uncertainties affecting the experimental determination [48, 49] and the observations varies in the $0.3 - 0.5\,\mathrm{GeV\,cm^{-3}}\,c^{-2}$ range. The local circular speed is measured by using various methods, too: from the study of the proper motion of Sgr A*, the massive black hole at the center of the MW, recent measurements give $v_c(R_\odot) = (248 \pm 7)\,\mathrm{km\,s^{-1}}$ [50], while other methods employing tracer stars give $v_c(R_\odot) = (229.0 \pm 0.2)\,\mathrm{km\,s^{-1}}$ [51]. Finally, regarding $v_{esc}(R_\odot)$, methods based on the study of high velocity stars from the RAVE survey give $v_{esc}(R_\odot) = 553^{+54}_{-41}\mathrm{km\,s^{-1}}$ [52], while other analyses based on Gaia Data Release 2 give $v_{esc}(R_\odot) = (580 \pm 63)\,\mathrm{km\,s^{-1}}$ [53].

Chapter 2

Dark matter direct detection

The goal of this thesis is to contribute to the search for dark matter candidates with direct detection experiments. But what does it mean "direct detection"? The attribute "direct" means observing the effect of the interaction of the dark matter particles on ordinary matter in a controlled environment. We know that DM interacts gravitationally with the SM particles, but, in the hypothesis in which these interactions are not only gravitational, the DM could in principle undergo scattering processes with ordinary matter, transferring its energy-momentum. The scattering process might happen, for instance, off an atomic nucleus or an electron. For the purposes of this work, we will assume the theoretical framework in which the dark matter is considered to be heavy, with a mass in the $\mathcal{O}(\mathrm{GeV} - \mathrm{TeV})$ range, and weakly interacting. In these conditions, assuming the Standard Halo Model, the dominant effect is the elastic scattering with the ordinary matter nuclei.

The chapter is structured as follows: in Sec. 2.1 we illustrate the general properties of the interactions between the DM particles and the ordinary matter nuclei; in Sec. 2.2 we describe the main experimental features of the modern direct detection detectors; finally, in Sec. 2.3 we report the state-of-the-art of this research field.

2.1 Dark matter nuclear scattering

2.1.1 Kinematics of the dark matter nuclear scattering

According to the Standard Halo Model, in the Milky Way and at our radius from the galactic center, the mean velocity of a dark matter particle is $\sim 220 \,\mathrm{km\,s^{-1}}$, corresponding to $\beta = v/c \sim 10^{-3}$. This means that we are in the non-relativistic regime, and we can treat the hypothetical scattering processes by means of non-relativistic kinematics and quantum mechanics. At this speed, the kinetic energy K for a DM mass of $m_{DM} = 10 \,\mathrm{GeV}/c^2$ is:

$$K = \frac{1}{2} m_{DM} \beta^2 c^2 = 5 \,\mathrm{keV}. \tag{2.1}$$

Since the typical binding energy in heavier nuclei is $\mathcal{O}(5 - 10 \,\mathrm{MeV})$, the possible effects related to the nuclear structure (and even more to the nucleon structure) can be safely neglected. In addition, we will consider an elastic DM-nucleus scattering, which is a good approximation in most of the applications, and it is useful to study
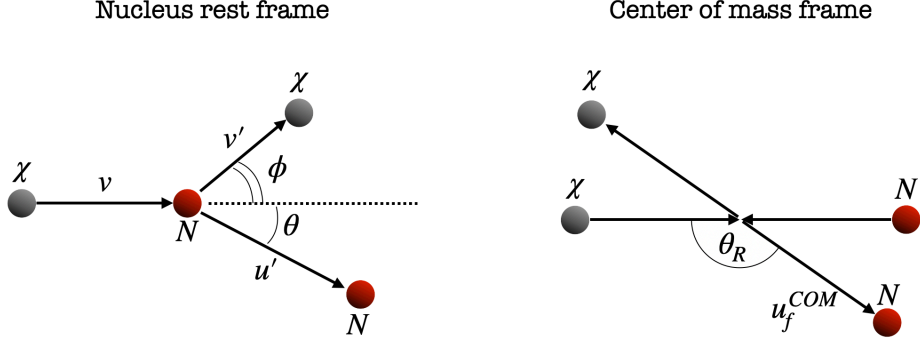
Figure 2.1. Elastic scattering of a DM particle (χ) and a nucleus (N) in the nucleus rest frame (**left**) and in the center of mass frame (**right**).

the kinematics of the process. However, possible photons or electrons can be emitted after the interaction, and in Chap. 3 we will study in detail this kind of effects.

In order to better understand the properties of the DM-nucleus scattering, let's give a brief description of the kinematics. Let's assume to have a certain detector looking for DM nuclear scattering: the typical mass of the target nucleus is $m_N \sim \mathcal{O}(10\,\mathrm{GeV}/c^2)$. For a temperature $T \sim \mathcal{O}(100\,\mathrm{K})$, the velocity of the nuclei of the target is

$$K = \frac{1}{2}m_N v_N^2 \sim k_B T \rightarrow v_N \sim \sqrt{\frac{2k_B T}{m_N}} \sim 10^3\,\mathrm{m\,s^{-1}}, \tag{2.2}$$

corresponding to $\beta \sim 10^{-6}$. We can therefore safely assume that, in the laboratory frame, the nuclei are at rest. For our non-relativistic and elastic treatment, we assume energy and momentum conservation. Let $\boldsymbol{v}$ and $\boldsymbol{u}$ denote the speed of the DM and the nucleus before the collision, respectively. Let's use the $\boldsymbol{v}'$ and $\boldsymbol{u}'$ notation to indicate the velocities after the collision, and let ϕ and θ indicate the scattering angles of $\boldsymbol{v}'$ and $\boldsymbol{u}'$, respectively, with respect to the DM initial velocity. For simplicity, let's choose the horizontal x-axis as aligned to the initial DM velocity, while the vertical y-axis is aligned to the corresponding transverse direction. With this notation, see Fig. 2.1, the momentum conservation laws can be written as

$$m_{DM}v = m_{DM}v'\cos\phi + m_N u'\cos\theta, \tag{2.3}$$

$$0 = m_{DM}v'\sin\phi - m_N u'\sin\theta. \tag{2.4}$$

By isolating the ϕ terms we obtain

$$v'\cos\phi = v - u'\frac{m_N}{m_{DM}}\cos\theta, \tag{2.5}$$

$$v'\sin\phi = u'\frac{m_N}{m_{DM}}\sin\theta. \tag{2.6}$$

$$\tag{2.7}$$

At this point we can square and sum the two equations, finding

$$m_{DM}(v')^2 = m_{DM}v^2 + \frac{m_N^2}{m_{DM}}(u')^2 - 2m_N v u'\cos^2\theta \tag{2.8}$$

Since for the energy conservation law

$$\frac{1}{2}m_{DM}v^2 = \frac{1}{2}m_{DM}(v')^2 + \frac{1}{2}m_{DM}(u')^2, \tag{2.9}$$

we can solve for u', obtaining

$$u' = 2\frac{\mu}{m_N}v\cos\theta, \tag{2.10}$$

where μ is the reduced mass defined as

$$\mu \equiv \frac{m_{DM}m_N}{m_{DM} + m_N}. \tag{2.11}$$

At this point the nuclear recoil (NR) energy is given by

$$E_R \equiv \frac{1}{2}m_N(u')^2 = 2\frac{\mu^2}{m_N}v^2\cos^2\theta. \tag{2.12}$$

It is often useful to express the previous equation in terms of the scattering angle θ_R in the center of mass (COM) frame, defined as the angle between the initial and final momenta of the target nucleus. To do that, we can relate θ and θ_R by writing the Galileo boost in the x direction, namely

$$u_f^{COM}\cos(\pi - \theta_R) + V_{COM} = u'\cos\theta, \tag{2.13}$$

where u_f^{COM} is the magnitude of the final velocity of the nucleus in the COM frame, V_{COM} is the velocity of the center of mass, and the π shift is due to the different definition of θ_R and θ (θ is referred to the initial velocity of the DM particle, while θ_R is referred to the initial velocity of the nucleus in the COM frame, which has the opposite direction, see Fig. 2.1). Noting that $u_f^{COM} = V_{COM} = (\mu/m_N)v$, we find

$$-\frac{\mu}{m_N}v\cos\theta_R + \frac{\mu}{m_N}v = 2\frac{\mu}{m_N}v\cos^2\theta, \tag{2.14}$$

and, simplifying,

$$1 - \cos\theta_R = 2\cos^2\theta. \tag{2.15}$$

Therefore, in terms of the scattering angle in the COM frame, the recoil energy is given by

$$E_R = \frac{\mu^2}{m_N}v^2\left(1 - \cos\theta_R\right). \tag{2.16}$$

This means that the maximum energy that the incoming DM particle can provide to a nucleus at rest via elastic scattering is

$$E_R^{max} = 2\frac{\mu^2}{m_N}v^2. \tag{2.17}$$

It is often useful to rewrite the above equations in terms of the dimensionless parameter r, defined as

$$r \equiv 4\frac{m_{DM}m_N}{(m_{DM} + m_N)^2}. \tag{2.18}$$
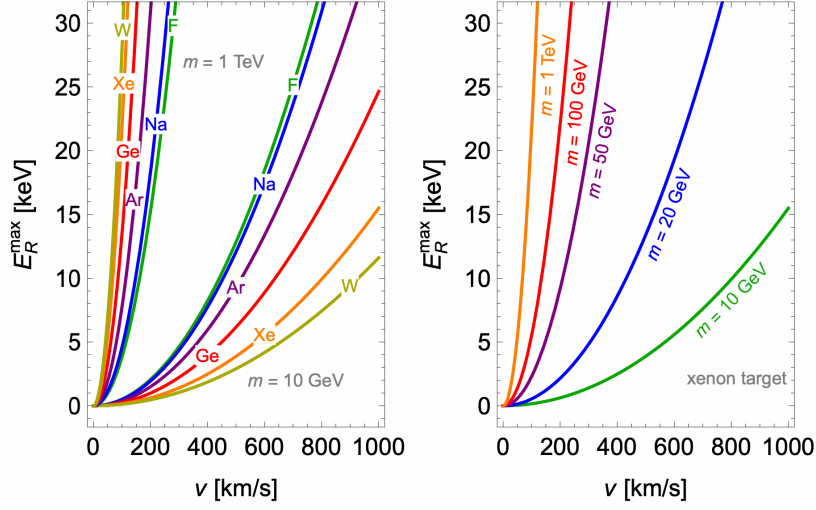
Figure 2.2. E_r^{max} as a function of the DM velocity v. Thanks to Eq. 2.20, the same plot can be read as v_{min}, on the x-axis, as a function of E_R, on the y-axis. On the **left**, the relation is shown for different targets and for $m_{DM} = 10\,\text{GeV}/c^2$ and $m_{DM} = 1\,\text{TeV}/c^2$. On the **right**, the relation is shown assuming a Xenon target for different values of m_{DM}. Figure from Ref. [54].

Using this parameterization Eq. 2.17 now becomes

$$E_R^{max} = rE_i, \tag{2.19}$$

with $E_i = (1/2)m_{DM}v^2$ indicating the initial kinetic energy of the incoming DM particle. The r parameter in the above equation can be interpreted as the energy transfer efficiency due to the kinematics of the process.

On the other hand, if we consider E_R as a fixed quantity, we can compute the minimum dark matter velocity that can produce a recoil with that energy as

$$v_{min} = \sqrt{\frac{m_N E_R}{2\mu^2}}. \tag{2.20}$$

These relations, depicted in Fig. 2.2, have an interesting implication. Lighter DM must have larger v_{min} to generate a certain recoil: this means that, since the SHM constraints the velocity distribution of DM, if an experiment has a certain energy threshold to detect a nuclear recoil, it would be more difficult to detect lighter particles with respect to heavier particles. In other words, at a fixed energy threshold, there exists a limit in the DM mass under which the experiment loses sensitivity. In the modern direct detection noble liquids experiments, for example, the typical energy threshold is of $\mathcal{O}(1\,\text{keV})$, meaning that this kind of experiments loses sensitivity to nuclear recoils if $m_{DM} \lesssim 1\,\text{GeV}/c^2$.

Finally, assuming the usual DM velocity $v \sim 220\,\text{km s}^{-1}$, the typical momentum of the incoming DM particle is, for $m_{DM} \sim \mathcal{O}(10\,\text{GeV}/c^2)$, $p \sim \mathcal{O}(10\,\text{MeV})$. The de Broglie wavelength corresponding to this momentum is given by

$$\lambda = \frac{h}{p} \sim 120\,\text{fm}. \tag{2.21}$$

On the other hand, the typical nuclear size can be estimated by its radius as $R_N \sim 1.25 A^{1/3}$ fm, where A is the atomic mass number of the target. This means that for instance Argon nuclei, with $A = 40$, have a radius of ~ 4 fm. Since the DM typical de Broglie wavelength exceeds by more than an order of magnitude the size of even large nuclei, the hypothetical elastic nuclear scattering will be "coherent". In other words, for this process, rather than a bound state of nucleons, the nucleus can be considered as a single object.

2.1.2   Scattering rate and cross-section

In the following, we will show how to obtain the differential event rate for a DM-nucleus interaction. Let's assume to have a certain detector looking for nuclear recoils in a heavy atom target and taking data for a certain amount of time t. The number of events induced by a DM halo with velocity v and number density n can be written as

$$N = tnvN_T\sigma, \tag{2.22}$$

where t is the exposure time and N_T is the number of target particles. In particular, we are interested to the energy spectrum of DM recoils, and to do this we can take the derivative with respect to the (measurable) recoil energy of the target particles, namely

$$\frac{dN}{dE_R} = tnvN_T\frac{d\sigma}{dE_R}. \tag{2.23}$$

However, as we know, the dark matter particles do not have a fixed velocity v, which, instead, is distributed according to its local velocity distribution $f(\boldsymbol{v})$. According to the SHM there is an upper limit in the maximum velocity determined by the Milky Way escape velocity v_{esc} but, in addition, since we are interested in the rate as a function of E_R, if we fix the nuclear recoil energy there exist a minimum possible DM velocity v_{min}, as described by Eq. 2.20. This mean that Eq. 2.23 should be written as

$$\frac{dN}{dE_R} = tnN_T \int_{v_{min}(E_R)}^{v_{esc}} vf(\boldsymbol{v})\frac{d\sigma}{dE_R}d\boldsymbol{v}. \tag{2.24}$$

Let's define the exposure $E = tM_T$ as the product of the exposure time t and the total mass of the target M_T. Noting that $n = \rho_0/m_{DM}$ and $N_T = M_T/m_N$, where ρ_0 is the local DM mass density, we obtain

$$\frac{dN}{dE_R} = E\frac{\rho_0}{m_{DM}m_N} \int_{v_{min}(E_R)}^{v_{esc}} vf(\boldsymbol{v})\frac{d\sigma}{dE_R}d\boldsymbol{v}. \tag{2.25}$$

The event rate R is defined as the number of events per unit target mass per unit time, so the differential event rate is given by

$$\frac{dR}{dE_R} = \frac{\rho_0}{m_{DM}m_N} \int_{v_{min}(E_R)}^{v_{esc}} vf(\boldsymbol{v})\frac{d\sigma}{dE_R}d\boldsymbol{v}. \tag{2.26}$$

The crucial quantity to be determined is therefore the differential cross-section. Since we do not know the nature of the DM-nucleus interaction, it is usually written as the sum of a spin-independent (SI) contribution and a spin-dependent (SD)

contribution,

$$\frac{d\sigma}{dE_R} = \left[\left(\frac{d\sigma}{dE_R}\right)_{SI} + \left(\frac{d\sigma}{dE_R}\right)_{SI}\right]. \tag{2.27}$$

Spin-independent contributions would come from scalar or vector DM coupling to the quarks forming the nuclei, while spin-dependent contributions would come from an axial-vector coupling to the quarks in the nuclei. Without going into the details of the cross-section calculations starting from the Lagrangian, a quite model independent parameterization is

$$\frac{d\sigma}{dE_R} = \frac{m_N}{2\mu^2 v^2}\left[\sigma_0^{SI}F_{SI}^2(q^2) + \sigma_0^{SD}F_{SD}^2(q^2)\right], \tag{2.28}$$

where $F_{SI,SD}$ indicates the nuclear form factors, which encode the dependence on the momentum transfer q^2 and the nuclear structure. In the non-relativistic case, we are in the low q^2 regime, and this form factors can be approximated to their zero-momentum values, namely $F_{SI,SD} \approx F_{SI,SD}(q^2 = 0) = 1$. Regarding the cross-sections, the most used parameterization in the literature are [54]

$$\sigma_0^{SI} = \frac{\mu^2}{\pi}\left[Zf_p + (A - Z)f_n\right]^2, \tag{2.29}$$

where Z is the atomic number and f_p (f_n) is the proton (neutron) coupling factor, and

$$\sigma_0^{SD} = \frac{\mu^2}{\pi}\frac{J + 1}{J}\left[a_p\langle S_p\rangle + a_n\langle S_n\rangle\right]^2, \tag{2.30}$$

where J is the nuclear angular momentum, a_p (a_n) is the proton (neutron) effective coupling constant and $\langle S_p\rangle$ ($\langle S_n\rangle$) is the proton (neutron) spin expectation value.

In the SI case, the standard assumption is the so-called isosinglet condition, i.e. $f_p = f_n$ [54]. This implies that the scattering cross-section is simply proportional to A^2. Indeed, at low momentum transfer, Eq. 2.29 can be written as

$$\sigma_0^{SI} = \sigma_p\frac{\mu^2}{\mu_p^2}A^2, \tag{2.31}$$

with σ_p being the DM-proton cross-section and μ_p being the DM-proton reduced mass.

In order to understand which is the dependence of the event rate on E_R, let's rewrite Eq. 2.26, assuming $v_{esc} \to \infty$ for simplicity, as

$$\frac{dR}{dE_R} \propto \int_{v_{min}(E_R)}^{\infty} v f(\boldsymbol{v})\frac{1}{v^2}d\boldsymbol{v}. \tag{2.32}$$

Using Eq. 1.28 but taking into account the speed $\boldsymbol{v}_E$ of the Earth, where the detector is located, with respect to the DM halo, we obtain

$$\frac{dR}{dE_R} \propto \int_{v_{min}(E_R)}^{\infty} \frac{\exp\left\{-\frac{|\boldsymbol{v}+\boldsymbol{v}_E|^2}{v_0^2}\right\}}{v}d\boldsymbol{v}, \tag{2.33}$$

where $v_0^2 \equiv 2\sigma^2/3$. Since we are interested in the general behavior of the event rate, we will introduce an additional simplification, namely we will neglect v_E. Using Eq. 2.20 e by means of a change of variable, the event rate can be written as

$$\frac{dR}{dE_R} \propto \int_{E_R}^{\infty} \exp\left\{ -\frac{E}{rE_0} \right\} dE, \tag{2.34}$$

where, by definition,

$$\begin{aligned} E_0 &\equiv \frac{1}{2} m_{DM} v_0^2 \\ r &\equiv 4 \frac{m_{DM} m_N}{(m_{DM} + m_N)^2}. \end{aligned} \tag{2.35}$$

For our simplistic calculation, the event rate, as a function of the nuclear recoil energy, can therefore be expressed as [55]

$$\frac{dR}{dE_R} = \frac{R_0}{rE_0} \exp\left\{ -\frac{E_R}{rE_0} \right\}, \tag{2.36}$$

where R_0 is the total event rate, and it is given by

$$R_0 = \int_0^{\infty} \frac{dR}{dE_R} dE_R. \tag{2.37}$$

As we have seen in this chapter, DM particles can deposit a small amount of energy, amounting to just $\mathcal{O}(10\,\text{keV})$, into the ordinary matter nuclei. In addition, as highlighted by Eq. 2.36, our very simplistic approach leads to a featureless and exponentially falling rate. Even with more realistic calculations, for instance taking into account a finite v_{esc}, keeping the $\boldsymbol{v}_E$ dependence, or considering additional higher order E_R terms in the scattering cross-section, the situation is still the same, see Fig. 2.3. There are no expected peaks, resonances, knees in the spectrum, making it difficult to distinguish from other background sources. Moreover, the total rate is expected to be very small, $\mathcal{O}(1\,\text{evt kg}^{-1}\,\text{y}^{-1})$[1], making the DM direct detection a real challenge. It is therefore not surprising that the existing DM direct detection experiments are characterized by a large exposure and very low background rates.

However, even if the focus of these searches has mainly been the event rate integrated over a long period of time, there is a growing interest in its time structure [56]. In fact, it is now clear that one of the inputs in the calculation of the DM rate is the average DM initial velocity before the DM-nucleus collision in the Earth reference frame. The motion of the Sun through the DM halo is expected to be in the direction of the Cygnus constellation, on the plane of the galaxy. In addition, the orbit of the Earth around the sun is tilted, with respect to the galactic rotation plane, by $\sim 60°$ above the line of motion. This means that the Earth's speed component parallel to the galactic rotation plane changes with time, with a period of 1 year. In other words, in half of the year our planet is moving "upwind" with respect to the DM halo, and in the other half is moving "downwind". This is, of course, a small effect, since the circular velocity v_{orbit} of the Earth in its motion

---

[1]Assuming an ^{40}Ar target, $\rho_0 = 0.3\,\text{GeVcm}^{-3}\,c^{-2}$, $m_{DM} = 10\,GeV/c^2$, and $\sigma_{DM}^{SI} \sim 10^{-38}\,\text{cm}^2$.
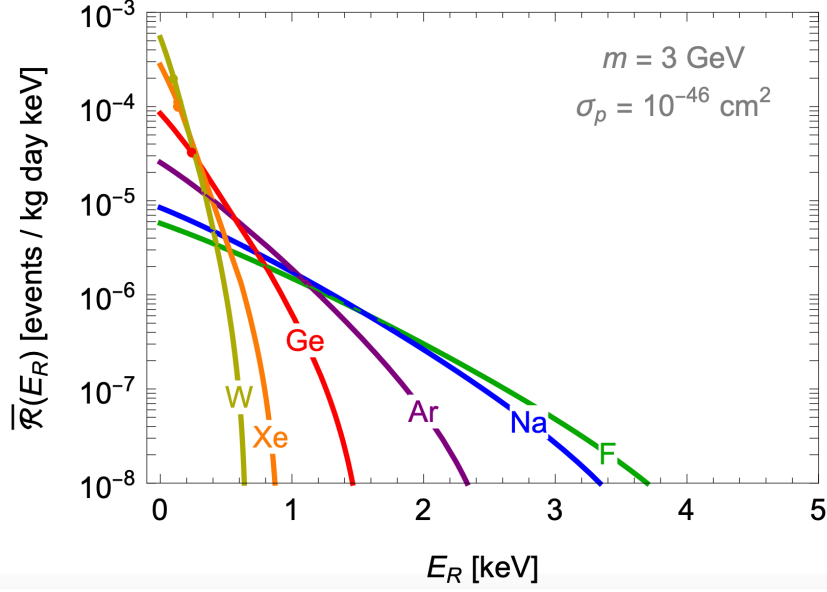
Figure 2.3. Annual average of the SI DM-nucleus scattering rate for a DM mass $m_{DM} = 3 \text{ GeV}/c^2$, assuming a DM-proton scattering cross-section $\sigma_p = 10^{-46} \text{ cm}^2$. Different lines correspond to different target nuclei. Figure from Ref. [54].

around the Sun is quite smaller than the circular velocity v_{sun} of the Sun around the galactic center, with $v_{orbit}/v_{sun} \sim 0.07$. However, it induces a small variation in total event nuclear recoil DM rate that can, in principle, be detected by looking for a possible time modulation of the rate. From a more quantitative point of view, the time dependence of the rate can be approximately computed as [57]

$$\frac{dR(t)}{dE_R} \simeq \frac{dR}{dE_R}\left[1 + \Delta(E_R)\cos\left(\frac{2\pi(t-t_0)}{T}\right)\right], \qquad (2.38)$$

where T is the expected 1 year period of the modulation, t_0 is the phase, which is such that the maximum is expected on June 2nd, and $\Delta(E_R)$ is the modulation amplitude, which, of course, depends on the recoil energy of the event. The above equation is the result of a first order Taylor expansion in the rate, since $\Delta(E_R)$ is expected to be of the order of $1 - 10\%$ [57].

Finally, due to the fact that the direction to the Cygnus constellation depends on the position of the detector on the Earth surface, small directional effects on the nuclear recoil rate can be observed. As a consequence, since the Earth is also rotating on its axis, we could see variations in the number of nuclear recoil events along a particular direction in the laboratory frame with a period of 1 day. Assuming the SHM, the direction dependence has been quantified as [57, 58]

$$\frac{dR}{dE_R d\cos\gamma} \propto \exp\left[-\frac{\left[\left(v_{orbit}^{\parallel} + v_{sun}\right)\cos\gamma - v_{min}\right]^2}{v_0^2}\right], \qquad (2.39)$$

where, as usual, $v_0^2 \equiv 2\sigma^2/3$, $v_{orbit}^{\parallel}$ is the component of the Earth's speed around
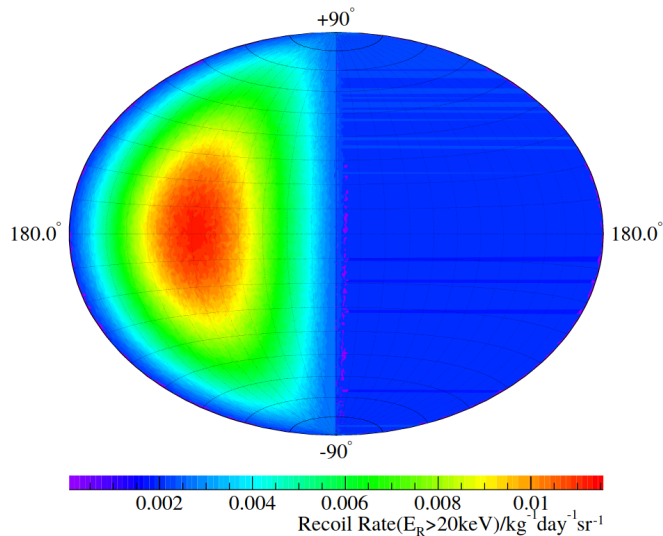
Figure 2.4. Directional recoil rate as a function of the nuclear recoil direction, for a standard halo of DM particles with mass $m_{DM} = 100 \text{ GeV}/c^2$ and for $E_R > 20$ keV. Figure from Ref. [57].

the Sun parallel to the Cygnus direction, and γ is the angle between the Cygnus-Sun line and the observed nuclear recoil. The angular distributions of the nuclear recoils peaks in the direction of the Cygnus constellation, and the event rate in this direction is approximately one order of magnitude larger than that in the opposite direction, see Fig 2.4. With a directional detector capable of measuring the 3D nuclear recoil direction of the event, it would be therefore possible to distinguish the DM signal from the flat isotropic expected background.

2.2 Underground detectors: backgrounds and signatures

This section will provide a general description of the typical backgrounds and signatures in the modern direct detection experiments, with a focus on the underground detectors. In fact, being the DM event rate so low, $\mathcal{O}(1 \text{ evt kg}^{-1} \text{ y}^{-1})$, and due to the fact that the rate has no peculiar features - i.e. peaks, knees, and so on - to be exploited for the detection, the experimental challenge is to reduce the background events as much as possible. It is therefore a natural choice to locate these experiments underground, where the cosmic ray background is reduced. In addition to possible external sources, the detector components or even the active material used as a target itself can be slightly contaminated by radioactive isotopes that in turn can mimic events during the experiment. It is therefore important to develop experimental techniques to reduce these background sources, as well as to determine in an accurate way the expected background event rate.
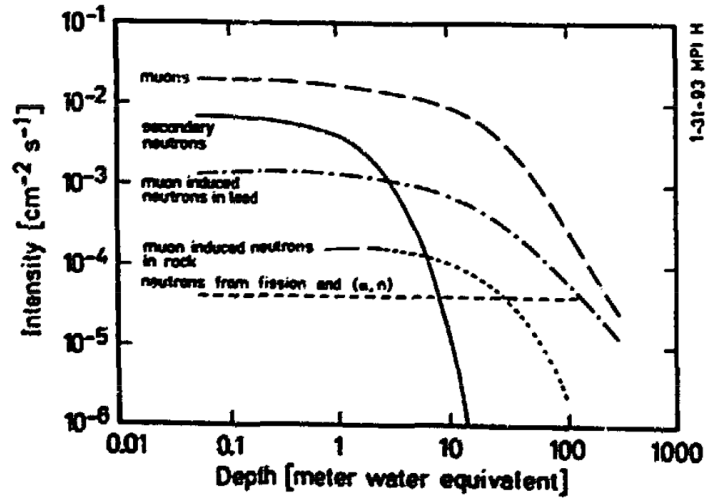
Figure 2.5. Cosmic ray flux of secondaries and neutrons produced by natural activity and muons as a function of the underground depth. The depth is expressed in terms of the equivalent distance if the crust material was made of water. Figure from Ref. [59].

2.2.1 Cosmic rays and environmental backgrounds

The first background source the experimentalist has to deal with consists of cosmic rays. These rays are made primarily of high energy particles - mostly protons and light nuclei - coming from astrophysical sources and secondaries - X-rays, muons, pions, electrons, protons, neutrons - produced in their interactions with the particles in the atmosphere. However, the rate of cosmic rays decreases by approaching the Earth surface, and it decreases even more below the Earth surface. Indeed, the materials constituting the Earth's crust act as a shield and can protect the detector from being exposed to the cosmic rays - see Fig. 2.5.

Many DM direct detection experiments are therefore located in underground caves, and many of them are also equipped with additional shielding and active veto systems capable of detecting external incoming radioactivity and therefore remove the external background. In fact, the environmental background from natural radioactivity of the rock of the cave has to be reduced as much as possible. For instance, a neutron inside the active volume of the experiment can indeed mimic a $1\,\mathrm{GeV}/c^2$ DM particle.

2.2.2 Internal Radioactivity

Besides the external radiation coming from the cosmic rays or for instance from radioactive materials inside the walls of the cave, the internal radioactivity, i.e. the radioactivity of the detector components themselves, generates background events. Indeed, the unavoidable contamination from radioactive unstable isotopes of Argon, Krypton, Cesium, Cobalt, Potassium, Thorium or Uranium atoms contained in the constituents of the detector, including the target material itself, can decay, and their decay products can produce events. This contamination is caused by either the natural abundance of these isotopes in the detector material or by cosmogenic

activation. In fact, during the different phases of the detector construction and design before moving to the final shielded location, the cosmic rays can activate unstable isotopes via nuclear interactions.

This explains why, in the modern and very sensitive dark matter direct detection experiments, a lot of work is spent in the detailed characterization of the employed materials. The internal background is usually determined by dedicated calibration measurements combined with the development of accurate MC simulations of the experimental geometry, materials and response. In order to reduce the internal radioactivity, however, a lot of detector-dependent techniques are developed by the different experimental collaborations. In fact, the materials used in the construction have to be the most radiopure, namely with very low unstable isotope contamination, and dedicated facility can be constructed to this purpose.

Let's just make an example related to the DarkSide-50 experiment [1]. This experiment is a dual phase Time Projection Chamber (TPC) experiment that uses Argon as a target material for DM detection, and a detailed description of the whole apparatus is given in Sec. 5. Argon is a noble gas that constitutes approximately 1% of the Earth's atmosphere. Atmospheric Argon (AAr) is mainly composed of three stable isotopes, that is ^{40}Ar (99.600%), ^{38}Ar (0.063%) and ^{36}Ar (99.336%). However, due to cosmogenic activation, there exists a small fraction of ^{39}Ar, which is an unstable isotope with a half-life of 269 years produced through the reaction

$$^{40}\text{Ar} + n \rightarrow {}^{39}\text{Ar} + 2n. \tag{2.40}$$

The β decays of ^{39}Ar can generate events in the same energy window of the DM signal, and therefore great effort has been made by the DarkSide collaboration to reduce the ^{39}Ar amount inside the active volume of the detector. The first step was to use underground Argon (UAr), namely argon extracted from underground caves, which, due to the shielding provided by the Earth's crust, has a lower ^{39}Ar content. After the extraction, the UAr was sent to Fermilab for an additional purification procedure by isotopic distillation. At the end of the procedure, the ultrapure UAr was inserted inside the DarkSide-50 detector, with a final concentration of ^{39}Ar $(1.4 \pm 0.2) \times 10^3$ times lower than AAr [60], see Fig. 2.6.

2.2.3 Event signatures: electronic and nuclear recoil

The radiation can interact with the active material of the detector and can produce two classes of signatures:

- the electronic recoil (ER), which is the interaction of photons, electrons, or positrons, constituting radioactive γ or X-ray emissions or β decays, with the atomic electrons. This kind of radiation can induce electronic recoils near the surface of the active volume of the experiment, but also in the core of this volume, depending on the penetrating power of the incoming particle. However, during an ER event a significant amount of energy, in the form of free electrons, can be released, and the event can in general be distinguished from a DM induced event.

- the nuclear recoil (NR), which is the interaction of neutrons, protons, alpha particles and eventually DM particles with the atom nuclei. Since they are
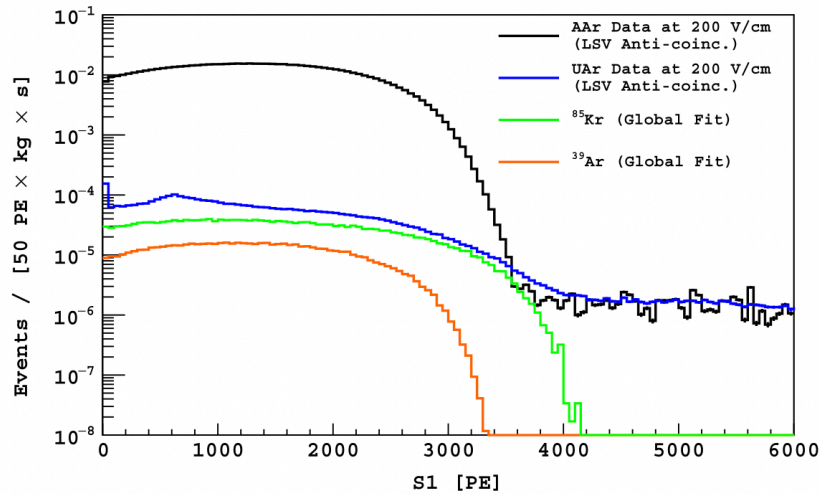
Figure 2.6. Internal background spectra as a function of the scintillation light, called "S1" signal, produced inside the detector during an event. The black line represents data observed with AAr, while the blue line correspond to data observed with UAr. The green and orange lines correspond to the ^{85}Kr and the ^{39}Ar contributions as a result of a global fit to the spectrum. Figure from Ref. [60].

electrically neutral, neutrons and WIMP-like particles interact almost entirely via NR. According to the standard assumptions, the electron cloud follows instantaneously the recoiling nucleus after the collision. The scattered atom can subsequently release energy, e.g. in the form of electrons or photons, inside the detector. However, the amount of energy released by NR events in the detector is generally smaller with respect to the ER case.

To be more specific, the responses of different materials to ER or NR can be different, and this feature can be exploited by choosing in an opportune way the active material of the experiment in order to better distinguish the two signatures. In particular, ER and NR events can produce:

- light: the interaction with the incoming particle can induce scintillation. By appropriately designing the detector target material and readout system, the light produced by scintillation can be collected, providing a measurement of a fraction of the energy release [1, 32, 61–64].

- ionization: the electrons extracted from atoms, in the presence of an appropriate electric field, can be collected, and the corresponding signal can be eventually amplified and revealed [1, 61, 62, 65–67]. From the number of extracted electrons produced in the event, with opportune calibrations, the energy release can be extracted.

- heat: in solid state materials or, more in general, in regular structured systems, phonons can be induced during the interaction and revealed [63, 66, 68]. These kinds of experiments are usually characterized by very low operating temperatures, needed to remove the thermal noise. The energy released in the form of phonons is a high fraction of the event released energy, depending

on the experimental efficiency. On the other hand, there exists another class of experiments [69, 70] which employ high pressure liquid systems in which, under appropriate conditions, an ER or NR event can induce a local boiling, and small bubbles can be observed.

As a general behavior, the detector responses to NR and ER events are different. In this context, the so-called "quenching factor" (QF) is an interesting parameter, and describes the ratio in detected energy between the two signatures. More precisely, the QF is usually defined in literature as

$$E_{detected}(\text{keVee}) = QF \, E_R(\text{keVnr}) \tag{2.41}$$

where $E_{detected}(\text{keVee})$ indicates the revealed energy in "keV electron equivalent" (keVee), namely the energy associated to an equivalent free electron, and $E_R(\text{keVnr})$ is the recoil energy in "keV nuclear recoil" (keVnr), namely the energy of the recoiling atom. The QF is therefore a measurement of the difference between the ER and NR energy response.

2.3 Direct detection experiments: sensitivity today

An interesting parameter space which is typically studied in DM direct detection searches is the SI cross-section as a function of the mass m_{DM} of the DM candidate, and it is shown in Fig. 2.7. The plot shows indeed the results from the various experiments in constraining the DM cross-section and, since it depends on the value assumed for m_{DM}, these constraints are represented by open lines. The green shaded region corresponds to the 90% C.L. excluded region. Even if in Fig. 2.7 it is not shown, the exclusion region does not extend up to an infinite cross-section. In fact, if the DM cross-section were too high, the DM particles from the galactic halo would be absorbed by the atmosphere and the Earth, without having the possibility to reach the detector. This effect is also known in literature as "Earth shielding" and in general its intensity is different in each experiment [72–76].

The most stringent constraints on the DM cross-section in the $m_{DM} \gtrsim 1 \, \text{GeV}/c^2$ mass region come from noble liquids experiments, such as XENON [41, 61] or LUX [62], which are both experiments based on a dual phase TPC filled with Xenon. The 2018 DarkSide-50 preliminary results [1] are represented by the magenta line in Fig. 2.7. For masses $m_{DM} \lesssim 1 \, \text{GeV}/c^2$ there's a general loss of sensitivity, due to the fact that, in this mass region, the nuclear recoil DM rate typically goes under the experimental thresholds $E_{th} = \mathcal{O}\left(1 \, \text{keV}\right)$. In order to explore this "low mass" region there are therefore two way to proceed: lowering the threshold, by developing more and more refined experimental techniques, or considering new possible signatures that can contribute to the signal rate. This is the case for the Migdal effect [3, 77], that, if taken into account, can improve the sensitivity to lower DM masses. This is what we showed in a recent work [2] in the Argon case, and we will see all the details in the next Chapter. The lines corresponding to experiments already employing the Migdal effect are denoted with an "(M)" flag in Fig. 2.7. With the exception of the very recent XENON1T excess [41], whose origin is still under debate, in noble gas direct detection detectors no signal has been observed so far.
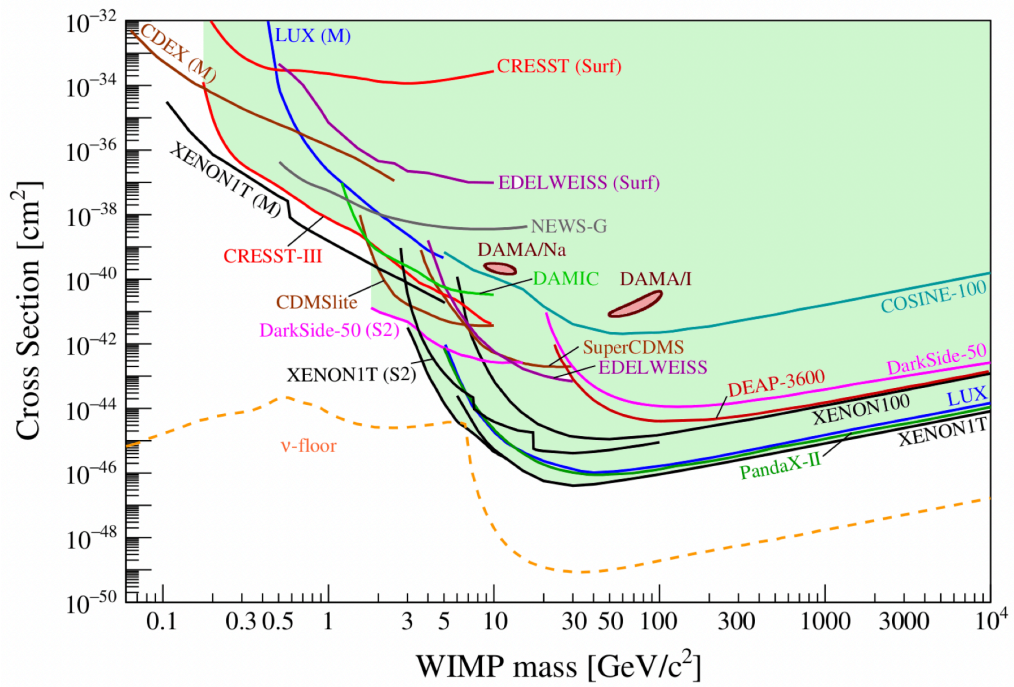
Figure 2.7. The state-of-the-art of direct DM searches for spin-independent DM cross-section. The curves are obtained assuming, as standard parameters for the DM halo, $\rho_0 = 0.3\ \mathrm{GeV\,cm^{-3}}$, $v_0 = 220\ \mathrm{km\,s^{-1}}$, $v_{esc} = 544\ \mathrm{km\,s^{-1}}$. The open lines represent the exclusion limits by the different experiments, as well as the green shaded area represent the correspondent excluded region. The closed lines correspond to possible signal observations. Figure from Ref. [71].

Fig. 2.7 shows also a pair of closed brown lines: closed lines in the plot represent the constraints coming from possible signal observations, and, in particular, these two refer to the DAMA/LIBRA experiment [31, 32]. DAMA/LIBRA is a direct detection experiment that observes scintillation induced in NaI crystals. The current results from DAMA/LIBRA show a time modulation in the observed rate with a period of 1 year and a phase consistent with the expectations for a WIMP induced modulation. However, since these results are not easy to relate to the ones of other experiments, the DM interpretation is still under debate [4]. More details with an example of this discussion will be given in Chap. 4.4.

The orange dashed line in Fig. 2.7 represents the so-called "neutrino floor". In fact, in order to improve the overall sensitivity of a certain experiment, in principle one can simply improve the exposure, i.e. the mass of the target material, the exposure time or both. Many collaborations, like for instance XENON and DarkSide, are going to upgrade the existing experiment to higher volume detectors. However, at a certain point, the experiment can become so sensitive as to be able to observe the coherent scattering of neutrinos on the nuclei, which has the same signature of a WIMP-like particle. In order to overcome this problem and remove the neutrino background, a possible solution is to use the directional information. In this sense, many projects are working on directional detection, and in particular during my PhD I had the possibility to contribute to the CYGNO project [78], which will be described in Chap. 7.

Chapter 3

The Migdal effect

Direct detection experiments are insensitive to nuclear recoils with energy below the keV, corresponding to sub-GeV/c^2 dark matter scattering. This relies on the assumption that the electron cloud around the nucleus follows instantaneously the nucleus itself, keeping the atom neutral. However, the sudden acceleration of a nucleus after a collision may lead to excitation and ionization of atomic electrons. This is an old idea from neutron-nucleus scattering experiments [79–84]. In other words, a nuclear recoil can induce the production of ionization electrons: this is known as Migdal effect [3,85,86], and it has been already exploited by experimental collaborations [62, 66, 87–92]. The advantage of taking into account the Migdal emission is that, as we will discuss in detail in this chapter, in the sub-GeV/c^2 dark matter mass region, it is easier to detect the Migdal electron rather than the nuclear recoil that originated it [2].

The Migdal effect event rate is given in Sec. 3.1: here, an analytical calculation for hydrogen and helium atoms is presented. Up to our knowledge, this represents the first explicit calculation of the rate of the Migdal process for a helium target. Then, the present techniques to compute the Migdal effect rate for heavier nuclei are summarized, with particular attention to the liquid argon case, which is the target material used for the DarkSide-50 experiments. Finally, in Sec. 3.2, following our phenomenological study of Ref. [2], the role of this effect in direct detection searches based on an argon target is explored.

3.1   Migdal electrons emission rate

The rate of ionization due to the Migdal effect for a nuclear recoil energy E_R accompanied by a ionization electron with energy E_e is given by the standard DM-nucleus differential recoil rate multiplied by the ionization rate [3]

$$\frac{d^2R}{dE_R dv} = \frac{d^2R_{NR}}{dE_R dv}|Z_{\text{ion}}(E_R)|^2, \tag{3.1}$$

where the ionization rate is given by

$$|Z_{\text{ion}}(E_R)|^2 = \frac{1}{2\pi}\sum_{n,\ell}\int dE_e \frac{dp_{q_e}^c(n\ell \to E_e)}{dE_e}. \tag{3.2}$$

Here, n and ℓ are the initial quantum numbers of the emitted electron, $q_e = m_e\sqrt{2E_R/m_N}$ is the electron momentum in the nucleus rest frame immediately after the DM collision, m_e is the electron mass, m_N is the nucleus mass, and $p_{q_e}^c(n\ell \to E_e)$ is the probability to emit an electron with final energy E_e. An approximate estimate of the total energy deposited in the detector is given by $E_d = E_e + E_{n\ell}$, where $E_{n\ell}$ is taken to be the binding energy of the (n, ℓ) state. This takes into account the fact that the emitted electron may come from an inner orbital and the remaining excited state will release further energy in the form of photons or additional electrons in order to return to the ground state.

For light nuclei, such as H and He, the ionization probability can be easily analytically computed by means of non-relativistic quantum mechanics - see Sec. 3.1.1. For heavier nuclei, due to the fact that the electron cloud wave function is in general very complicated, a numerical approach can be used [3]. In addition, a recent work demonstrated that the Migdal transition probability can be expressed in terms of the photoabsorption cross-section [93]. These approaches will be briefly described in Sec. 3.1.2.

3.1.1 Migdal ionization probability in light atoms

In this section, we will show how to compute the Migdal ionization probability in the simplest possible case: a DM particle scattering on a H or He atom. We will do the following assumptions

- for the atom's electrons, the DM-nucleus collision can be considered as instantaneous, i.e. the interaction duration, approximately of the order of $\tau_I \sim R_N/v_{DM}$, is smaller than the period of the electrons around the nucleus, approximately $\tau_e \sim a_0/v_e$, where R_N is the radius of the nucleus, a_0 is the Bohr radius and v_e is the electronic orbital velocity. This condition, which is satisfied in our case since $v_{DM} > R_N v_e/a_0 \sim 10^{-7}$, implies that the wave function of the atom's electrons does not change significantly during the event.

- the interaction is instantaneous for the nucleus itself, too. This is a good approximation if, during the collision, the displacement of the nucleus $L \sim \tau_I v_N$ is smaller if compared with the nucleus radius R_N. This second condition holds if $v_N/v_{DM} = 2\cos\theta\mu/m_N < 1$, where we used Eq. 2.10. The simplistic approach of this section will therefore be valid if $m_{DM} \lesssim m_N$.

The second condition implies that the electronic wave function in the frame O of the nucleus before the collision, where the nucleus is at rest, can be considered as the same as the one in the frame O' of the nucleus after the collision, where the nucleus has a velocity v_N. Even if this holds, however, in the frame of the nucleus after the collision, the velocity of the electrons are shifted by $-v_N$, due to the fact the nucleus acquires a velocity v_N. Therefore, in this frame, the wave function of the atom's electrons can be written as [80]

$$\psi'(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots) = e^{-\frac{i}{\hbar}\boldsymbol{q}_e \cdot \sum_i \boldsymbol{r}_i}\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots), \tag{3.3}$$

where ψ' is the wave function in the frame of the nucleus after the collision, ψ is the wave function in the frame of the nucleus before the collision, $\boldsymbol{r}_i$ is the

spatial coordinate for the i-th electron and $\boldsymbol{q}_e = m_e \boldsymbol{v}_N$, with m_e being the mass of the electrons. Eq. 3.3 can be simply derived by shifting in an opportune way the electrons momenta in the Fourier space.

Migdal effect in a H atom

As a first case, let's consider a DM particle scattering on a H atom in the ground state. The initial electronic state in the O' frame is therefore

$$\psi_0'(\boldsymbol{r}) = \frac{1}{\sqrt{\pi a_0^3}} e^{-r/a_0} e^{-\frac{i}{\hbar} \boldsymbol{q}_e \cdot \boldsymbol{r}}. \tag{3.4}$$

We are interested to the ionization probability, therefore the final state is a free particle, i.e. a plane wave

$$\psi_f(\boldsymbol{r}) = \psi_{free}(\boldsymbol{r}) = \frac{1}{(2\pi\hbar)^{3/2}} e^{\frac{i}{\hbar} \boldsymbol{p} \cdot \boldsymbol{r}}, \tag{3.5}$$

where $\boldsymbol{p}$ is the momentum of the free electron in the O' frame. The transition amplitude is therefore given by

$$\langle \psi_f | \psi_0' \rangle = N \int d\boldsymbol{r}\, e^{-\frac{i}{\hbar} \boldsymbol{p} \cdot \boldsymbol{r}} e^{-r/a_0} e^{-\frac{i}{\hbar} \boldsymbol{q}_e \cdot \boldsymbol{r}}, \tag{3.6}$$

where

$$N = \frac{1}{\sqrt{\pi a_0^3}} \frac{1}{(2\pi\hbar)^{3/2}}. \tag{3.7}$$

If we define $\boldsymbol{k} = \boldsymbol{p} + \boldsymbol{q}_e$, we obtain

$$\begin{aligned}
\langle \psi_f | \psi_0' \rangle &= 2\pi N \int_0^\infty dr \int_{-1}^1 d\cos\theta\, r^2 e^{-r/a_0} e^{-\frac{i}{\hbar} kr \cos\theta} \\
&= 2\pi N \int_0^\infty dr \frac{ir\hbar}{k} e^{-r/a_0} \left(e^{-\frac{i}{\hbar} kr} - e^{\frac{i}{\hbar} kr} \right) \\
&= 2\pi N \frac{i a_0^2 \hbar}{k} \left(\frac{1}{\left(\frac{i}{\hbar} ka_0 + 1 \right)^2} - \frac{1}{\left(-\frac{i}{\hbar} ka_0 + 1 \right)^2} \right) \\
&= 2\pi N \frac{i a_0^2}{k} \left(\frac{-4ika_0}{\left(\frac{k^2 a_0^2}{\hbar^2} + 1 \right)^2} \right) \\
&= N \frac{8\pi a_0^3}{\left(\frac{k^2 a_0^2}{\hbar^2} + 1 \right)^2}.
\end{aligned} \tag{3.8}$$

We are interested in the transition probability and therefore, squaring the amplitude and explicating N, we obtain

$$\begin{aligned}
\left| \langle \psi_f | \psi_0' \rangle \right|^2 &= \frac{8a_0^3}{\pi^2 \hbar^3} \frac{1}{\left(\frac{k^2 a_0^2}{\hbar^2} + 1 \right)^4} \\
&= \frac{8a_0^3}{\pi^2 \hbar^3} \frac{1}{\left[\frac{a_0^2}{\hbar^2} \left(p^2 + q_e^2 + 2pq_e \cos\alpha \right) + 1 \right]^4},
\end{aligned} \tag{3.9}$$

where α is the angle between $\boldsymbol{q}_e$ and $\boldsymbol{p}$. At this stage, since we are only considering the Migdal emission at a fixed v_N, $\boldsymbol{q}_e$ can be regarded as a fixed parameter. The $\cos\alpha$ term is just telling us that the electron is preferentially emitted, in the O' frame, in the direction which is opposite to the nucleus velocity direction. In the O frame, this can be interpreted as if the electron is "left behind" by the nucleus after the collision with the DM incoming particle. However, the $\cos\alpha$ dependence is very weak: for $p \ll q_e$ or $p \gg q_e$ the cosine term is negligible with respect to q_e^2 or p^2, respectively; when $p \sim q_e$, being $v_N \sim v_{DM} \sim 10^{-3}$, since $a_0^2 q_e^2/\hbar^2 \sim 10^{-2} \ll 1$, the $p^2 + q_e^2 + 2pq_e\cos\alpha$ term in the denominator of Eq. 3.9 is negligible with respect to 1.

Migdal effect in a He atom

The H atom, even if it is a useful example to understand some of the general aspects of the Migdal effect, is a very simplistic case, since there is only one electron in the electron cloud. However, an interesting and more realistic example for DM direct detection experiments is the He atom. With respect to the H case, now the system is composed of two interacting electrons feeling the nucleus $Z = 2$ Coulomb potential. The repulsion interaction between the electrons makes the Schrödinger equation for the system impossible to solve analytically. However, we can take into account this interaction term in a simple way by means of the Mean-Field approximation and the Variational principle: the ground state of the He atom can be in fact approximated to [94]

$$\psi_0\left(\boldsymbol{r}_1, \boldsymbol{r}_2\right) \approx \psi_0^{Z_{eff}}\left(\boldsymbol{r}_1\right)\psi_0^{Z_{eff}}\left(\boldsymbol{r}_2\right)\left|\chi\right\rangle \tag{3.10}$$

with $\left|\chi\right\rangle$ being the spin singlet state, and $\psi_0^{Z_{eff}}\left(\boldsymbol{r}\right)$ being defined as

$$\psi_0^{Z_{eff}}\left(\boldsymbol{r}\right) = \sqrt{\frac{Z_{eff}^3}{\pi a_0^3}}e^{-Z_{eff}r/a_0}, \tag{3.11}$$

where in turn $Z_{eff} = 27/16 < 2$. For our purpose we are interested in the ground state in the O' frame, which is given by

$$\psi_0'\left(\boldsymbol{r}_1, \boldsymbol{r}_2\right) \approx e^{-\frac{i}{\hbar}\boldsymbol{q}_e\cdot\boldsymbol{r}_1}e^{-\frac{i}{\hbar}\boldsymbol{q}_e\cdot\boldsymbol{r}_2}\psi_0^{Z_{eff}}\left(\boldsymbol{r}_1\right)\psi_0^{Z_{eff}}\left(\boldsymbol{r}_2\right)\left|\chi\right\rangle, \tag{3.12}$$

which is our initial state. The final state corresponding to the Migdal emission is such that one of the two electrons is free, while the other is bound to the nucleus. The system made of the nucleus and the bound electron is just a hydrogenoid atom with $Z = 2$. In addition, since the interaction is spin independent, the spin state of the final state is equal to the spin state of the initial state. The $\left|\chi\right\rangle$ state is antisymmetric, therefore the spatial wave function must be symmetric. Collecting all these considerations together, we can express the final state in the O' frame as

$$\psi_f\left(\boldsymbol{r}_1, \boldsymbol{r}_2\right) = \frac{1}{\sqrt{2}}\left[\psi_0^Z\left(\boldsymbol{r}_1\right)\psi_{free}\left(\boldsymbol{r}_2\right) + \psi_{free}\left(\boldsymbol{r}_1\right)\psi_0^Z\left(\boldsymbol{r}_2\right)\right]\left|\chi\right\rangle, \tag{3.13}$$

where

$$\psi_0^Z\left(\boldsymbol{r}\right) = \sqrt{\frac{Z^3}{\pi a_0^3}}e^{-Zr/a_0}, \tag{3.14}$$

with $Z = 2$. For simplicity, let's define

$$\psi^A\left(\boldsymbol{r}_1, \boldsymbol{r}_2\right) = \psi_0^Z\left(\boldsymbol{r}_1\right)\psi_{free}\left(\boldsymbol{r}_2\right) \tag{3.15}$$

$$\psi^B\left(\boldsymbol{r}_1, \boldsymbol{r}_2\right) = \psi_{free}\left(\boldsymbol{r}_1\right)\psi_0^Z\left(\boldsymbol{r}_2\right), \tag{3.16}$$

$$\tag{3.17}$$

so that

$$\langle\psi_f|\psi_0'\rangle = \frac{1}{\sqrt{2}}\Big(\langle\psi^A|\psi_0'\rangle + \langle\psi^B|\psi_0'\rangle\Big). \tag{3.18}$$

Let's focus on the first term on the right-hand side of the above equation, which corresponds to

$$
\begin{aligned}
\langle\psi^A|\psi_0'\rangle &= \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2 \left[\psi_0^Z\left(\boldsymbol{r}_1\right)\right]^*\left[\psi_{free}\left(\boldsymbol{r}_2\right)\right]^* e^{-\frac{i}{\hbar}\boldsymbol{q}_e\cdot(\boldsymbol{r}_1+\boldsymbol{r}_2)}\psi_0^{Z_{eff}}\left(\boldsymbol{r}_1\right)\psi_0^{Z_{eff}}\left(\boldsymbol{r}_2\right) \\
&= N'\int d\boldsymbol{r}_1\int d\boldsymbol{r}_2 e^{-Zr_1/a_0}e^{-\frac{i}{\hbar}\boldsymbol{p}\cdot\boldsymbol{r}_2}e^{-\frac{i}{\hbar}\boldsymbol{q}_e\cdot(\boldsymbol{r}_1+\boldsymbol{r}_2)}e^{-Z_{eff}(r_1+r_2)/a_0}, \\
&= N'I_{A1}I_{A2}
\end{aligned}
\tag{3.19}
$$

with

$$N' = \sqrt{\frac{Z^3}{\pi a_0^3}}\frac{1}{\sqrt{2\pi\hbar^3}}\frac{Z_{eff}^3}{\pi a_0^3} \tag{3.20}$$

$$I_{A1} = \int d\boldsymbol{r}_1 e^{-Zr_1/a_0}e^{-\frac{i}{\hbar}\boldsymbol{q}_e\cdot\boldsymbol{r}_1}e^{-Z_{eff}r_1/a_0} \tag{3.21}$$

$$I_{A2} = \int d\boldsymbol{r}_2 e^{-\frac{i}{\hbar}(\boldsymbol{q}_e+\boldsymbol{p})\cdot\boldsymbol{r}_2}e^{-Z_{eff}r_2/a_0}, \tag{3.22}$$

$$\tag{3.23}$$

where, as for H, $\boldsymbol{p}$ is the momentum of the free electron in the O' frame. The two integrals I_{A1} and I_{A2} have the same form as the integral in Eq. 3.6, with just different momenta and Z values. Using the result of Eq. 3.8, the integrals I_{A1} and I_{A2} are equal to

$$I_{A1} = \frac{8\pi a_0^3\left(Z + Z_{eff}\right)}{\left[\frac{a_0^2}{\hbar^2}q_e^2 + \left(Z + Z_{eff}\right)^2\right]^2} \tag{3.24}$$

$$I_{A2} = \frac{8\pi a_0^3 Z_{eff}}{\left[\frac{a_0^2}{\hbar^2}|\boldsymbol{p} + \boldsymbol{q}_e|^2 + Z_{eff}^2\right]^2}. \tag{3.25}$$

$$\tag{3.26}$$

Due to the symmetric property of the spatial wave functions, it can be easily found that

$$\langle\psi^B|\psi_0'\rangle = N'I_{A1}I_{A2} = \langle\psi^A|\psi_0'\rangle. \tag{3.27}$$
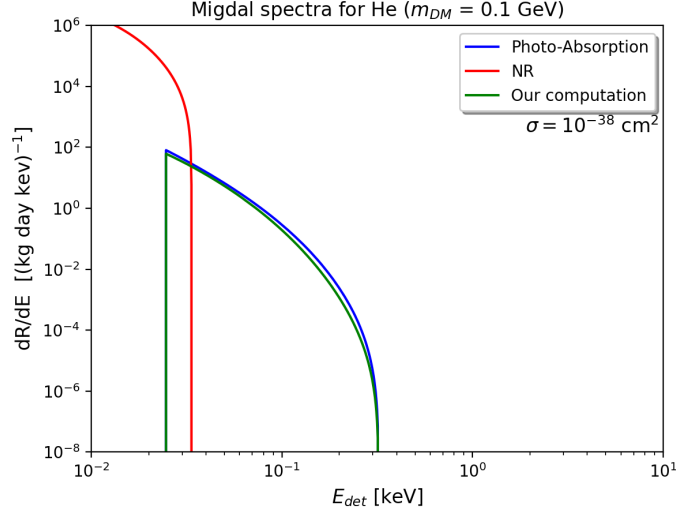
Figure 3.1. Comparison between the analytical approximated calculation of Eq. 3.28 of the Migdal rate (green), the photoabsorption [93] Migdal rate (blue) and the NR rate (red) for a He atom. The plot is computed assuming a DM mass $m_{DM} = 0.1 \text{ GeV}/c^2$ and a DM cross-section $\sigma_{SI}^{DM} = 10^{-38} \text{ cm}^2$.

Putting all the definitions together, we obtain, squaring the transition amplitude,

$$\left|\langle \psi_f | \psi_0' \rangle\right|^2 = \frac{1024}{\pi^2} \frac{a_0^3}{\hbar^3} \frac{Z^3 Z_{eff}^8 \left(Z + Z_{eff}\right)^2}{\left[\frac{a_0^2}{\hbar^2} q_e^2 + \left(Z + Z_{eff}\right)^2\right]^4 \left[\frac{a_0^2}{\hbar^2} |\boldsymbol{p} + \boldsymbol{q}_e|^2 + Z_{eff}^2\right]^4}. \quad (3.28)$$

In this case there is a dependence on the angle α between $\boldsymbol{p}$ and $\boldsymbol{q}_e$, too. However, for the same reasons of the H case, this dependence is very weak. The correspondent Migdal rate as a function of the detectable energy E_d of the Migdal electron in the O laboratory frame is shown in Fig. 3.1, together with the calculation coming from the more accurate photoabsorption derived calculations [93] - see Sec. 3.1.2. The figure shows how the Migdal rate extends up to higher E_d if compared to the NR rate. This is a general property of the Migdal effect when the DM mass is small enough, and, as we will see in detail in Sec. 3.2, this feature can be exploited in DM direct detection searches to extend the sensitivity to lower DM candidates.

We computed for the first time, as far as we know, the ionization probabilities for He.

3.1.2 Migdal ionization probability in heavy atoms

The approach used for the theoretical approximated calculation of the previous section is known as the "Migdal approach". In this approach, we treat the final state and the NR process in a separate way, and the energy-momentum conservation is imposed as an input. However, a more complete and theoretically clean method has been recently proposed in Ref. [3], showing how to treat the atom-DM system in a more coherent way, even when dealing with heavy nuclei with many electrons such as Xe, Ge or Ar. In any case, for these nuclei a numerical tool is needed to compute
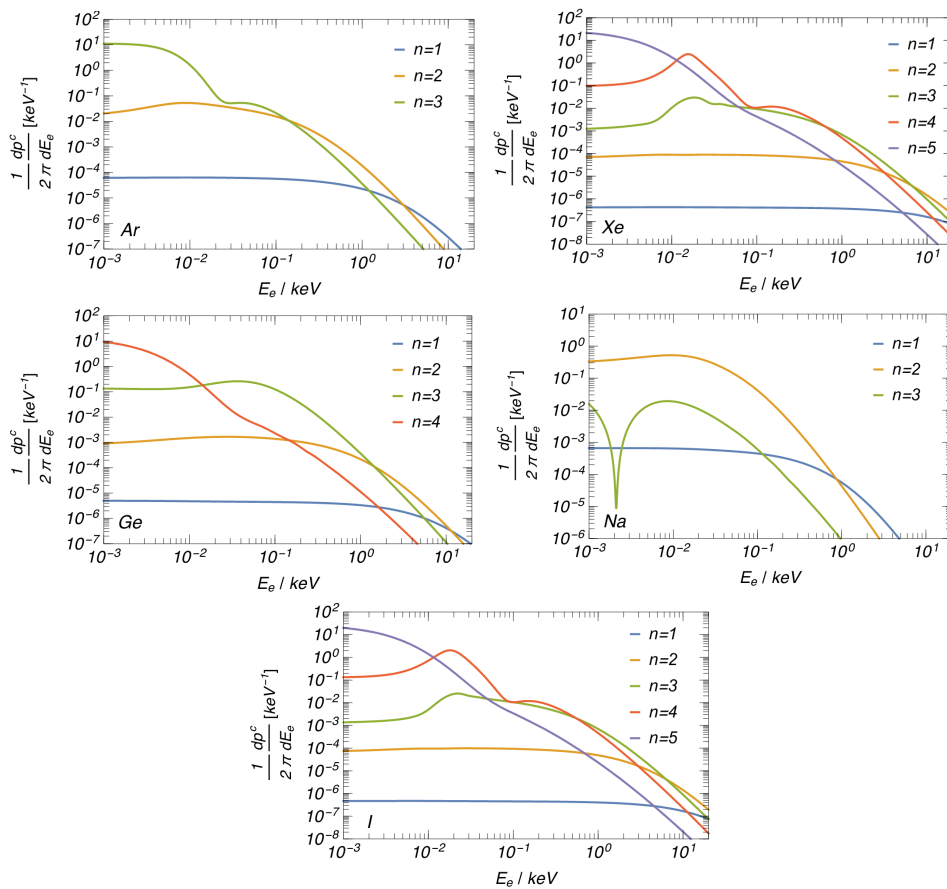
Figure 3.2. Differential Migdal ionization probabilities for Ar, Xe, Ge, Na and I. Figure from Ref. [3].

the electronic cloud wave functions, as for example the `Flexible Atomic Code` (`FAC`) [95] as it has been done in Ref. [3]. The differential ionization probabilities for different heavy nuclei as a function of the energy E_e of the emitted electron are shown in Fig. 3.2. The figure is obtained assuming $q_e = m_e v_N$, with $v_N = 10^{-3}$. Each colored line corresponds to the contribution of the electrons having a given initial principal quantum number n, in turn obtained by summing over all the possible initial electronic angular momenta ℓ. From a quantitative point of view, by integrating these probabilities, it has been found that, assuming $v_N = 10^{-3}$, the probability of the emission of an electron from the inner shells is of $\mathcal{O}(10^{-2})$, while the valence electrons can be emitted with a probability of $\mathcal{O}(10^{-1})$ [3].

The accuracy of these differential ionization probabilities relies on the fact that the numerically computed wave functions can reproduce the binding energies for the different levels with an accuracy of $\sim \mathcal{O}(20\%)$ and it should provide a correct estimate of the expected signal rate for inner-shell electrons. On the other hand, the prediction for the valence electron shells should be taken as an order of magnitude estimate.[1] However, a new relation between the Migdal process and photoabsorption [93] gives results comparable with the one obtained by Ref. [3] including also

---

[1]Private communication with M. Ibe.

the valence shell. In fact, we already showed in Sec. 3.1.1 that the matrix element describing the Migdal ionization process can be expressed as

$$M_{FI} = \left\langle \psi_f \left| e^{-\frac{i}{\hbar} \boldsymbol{q}_e \cdot \sum_{i=1}^{Z} \boldsymbol{r_i}} \right| \psi_i \right\rangle, \qquad (3.29)$$

where the index i denotes the i-th electron. Recalling that $a_0^2 q_e^2 / \hbar^2 \sim 10^{-2} \ll 1$, this matrix element can be well approximated by

$$M_{FI} \simeq -\frac{i}{\hbar} \boldsymbol{q}_e \cdot \left\langle \psi_f \left| \sum_{i=1}^{Z} \boldsymbol{r_i} \right| \psi_i \right\rangle \equiv -\frac{i}{\hbar} \boldsymbol{q}_e \cdot \boldsymbol{D}_{FI}, \qquad (3.30)$$

where $\boldsymbol{D}_{FI}$ is the dipole matrix element. As shown in Ref. [93], this matrix element is the same as the one appearing in the total cross-section of the photoabsorption process, which, in the electric dipole approximation, can be expressed as

$$\sigma_\gamma(E_e) \simeq 4\pi^2 \alpha_{EM} E_e D_{FI}^2, \qquad (3.31)$$

where $\alpha_{EM} = \frac{e^2}{4\pi\epsilon_0} \frac{1}{\hbar c}$. It is therefore straightforward to recognize that

$$\frac{d\sigma^{Mig}}{dE_R dE_e} \simeq \frac{m_e}{\mu^2 v_{DM}^2} \tilde{\sigma}_N(q_e) \frac{E_R}{E_e} \frac{\sigma_\gamma(E_e)}{4\pi^2 \alpha_{EM}}, \qquad (3.32)$$

where σ^{Mig} is the Migdal effect cross-section, E_R is the nuclear recoil energy, E_e is the energy of the emitted electron and $\tilde{\sigma}_N$ is the DM-nucleus cross-section. This relation is very useful since $\boldsymbol{D}_{FI}$ is the source of the main theoretical uncertainties (its computation requires approximations as well as many-body calculations). The photoabsorption cross-sections, instead, can be extracted directly from experiments. Figure 3.1 shows the comparison between the theoretical result for He and the one obtained with the photoabsorption method. An equivalent comparison with a heavier atom, namely Ar, is shown in Fig. 3.3. The agreement between the two methods is stable over several DM masses and DM-nucleus cross-sections. As for the He case, the Migdal rate extends up to higher E_d if compared to the NR rate - see Sec. 3.2 for more details.

Finally, let's focus in more detail on the Migdal rate for Ar, which is the active material of the DarkSide-50 experiment. The differential probability rates were computed so far without taking into account the shifts in electronic energy levels because atoms could actually be in a liquid (such as in argon or xenon targets) or crystal state (such as in germanium, silicon or sodium iodide detectors). This effect should decrease the ionization energy, and thus, if neglected, should lead to a conservative ionization yield estimate [96]. Figure 3.4 shows the differential ionization probabilities as a function of the detected energy $E_d = E_e + E_{n\ell}$ for isolated argon atoms. The different curves of Fig. 3.4 show the contributions for different principal quantum numbers n, where the contributions for different orbital angular momenta in the initial state ℓ and all possible final states are summed, and $q_e = 1$ eV/c. This figure shows that given the thresholds of 0.85 keV (4 reconstructed ionization electrons) or 1.47 keV (7 reconstructed ionization electrons) of a hypothetical liquid Argon (LAr) experiment, the contribution of the valence electrons can maximize
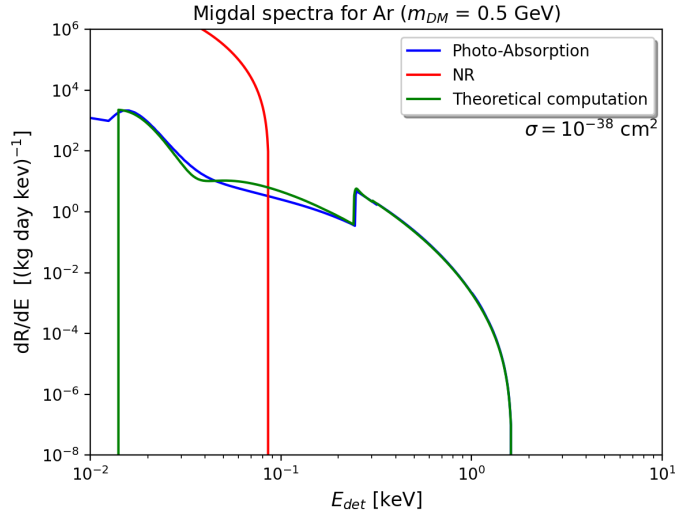
Figure 3.3. Comparison between the theoretical calculation of the Migdal rate (green), the photoabsorption Migdal rate (blue) and the NR rate (red) for a Ar atom. The plot is computed assuming a DM mass $m_{DM} = 0.5 \text{ GeV}/c^2$ and a DM cross-section $\sigma_{SI}^{DM} = 10^{-38} \text{ cm}^2$.
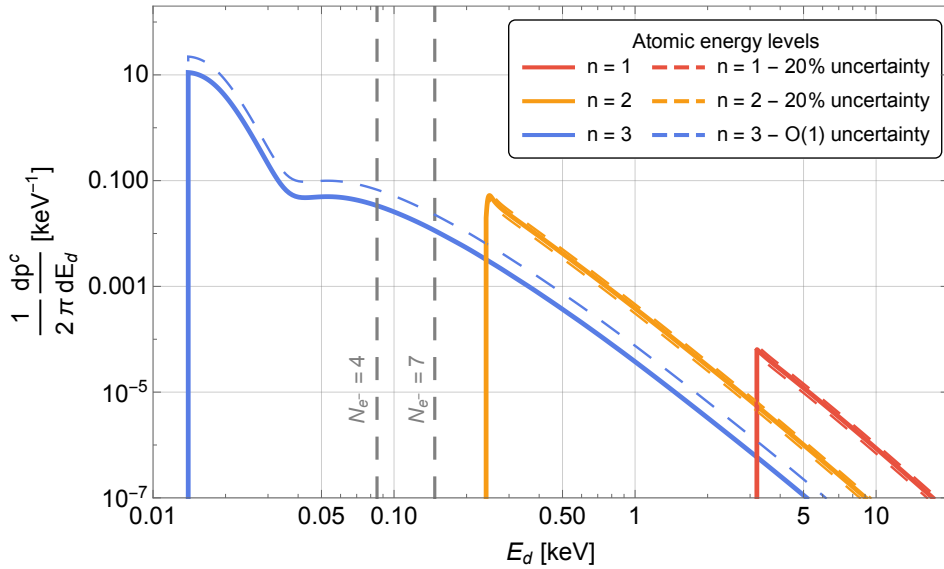
Figure 3.4. Differential ionization probabilities and related uncertainties as a function of the detected energy E_d for isolated argon and different principal quantum numbers n. We also show the 0.85 keV (4 reconstructed ionization electrons) or 1.47 keV (7 reconstructed ionization electrons) thresholds for DarkSide-50.

the sensitivity to nuclear scattering. This is in contrast with the reported results presented for xenon [62, 90] and germanium detectors [66, 87], where the detector thresholds are higher than the one needed to see a dominant signal from outer shells. As a consequence, it would be crucial to have a reliable computation of the transition probabilities in the case of LAr.

We stress that the same considerations discussed here are applicable to neutron scattering and should be taken into account by the experimental collaborations when estimating the radiogenic background contributions. In fact, the original idea was applied to neutron nucleus scattering [79–83].

3.2 Migdal spectrum vs nuclear recoil

In the limit of low momentum transfer, the Migdal effect shares the same kinematics as inelastic dark matter models [97], where the DM mass splitting δm is replaced by the total electronic energy E_d. In particular, we have that

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

(3.33)

where δ corresponds to E_d in our notation. The maximum nuclear and electronic recoil energy for a given DM mass are

$$E_{R,\max} = \frac{2\mu_N^2 v_{\max}^2}{m_N}, \qquad \delta_{\max} = \frac{\mu_N v_{\max}^2}{2}.$$

(3.34)

This shows that $\delta_{\max} > E_{R,\max}$ for $m_{DM} \ll m_N$ due to the suppression factor μ_N/m_N. Indeed, for $v_{\max} \sim 800\,\mathrm{km/s} \sim 2.7 \cdot 10^{-3}\,c$, $m_N \simeq 40$ GeV/c^2 (the approximate argon mass) and a DM mass of 0.5 GeV/c^2, we find $E_{R,\max} \sim 0.09$ keV, while $\delta_{\max} \sim 1.8$ keV. As a result, there is a range of DM masses for which it is easier to detect the electronic energy originating from the Migdal process rather than nuclear recoils, as a consequence of the fact that more energy can be carried off by light or massless particles for a given momentum transfer.

In Fig. 3.5 we show the behavior of $E_{R,\max}$ (solid lines) and $\delta_{\max}$ (dashed lines) as a function of the DM mass for different possible nuclei. The red horizontal line represents the typical experimental threshold of 1 keV. In all cases we see that there exists a region, in the low mass regime, in which the Migdal effect is over the red line, i.e. the experimental threshold, while the NR is below the red line. This region is wider for heavier nuclei like Xe and Ar. From the experimental point of view, this behavior can therefore be exploited to extend the experimental sensitivity to lower masses. In this sense, Fig. 3.5 shows that the gain in terms of sensitivity to lower DM masses is greater in heavier nuclei such as Xe and Ar. In fact, the Migdal effect in Xe nuclei has been already employed to set the exclusion limit on the DM cross-section by the XENON1T collaboration [90].

In a recent publication [2], we studied the impact of the Migdal electron on the sensitivity of LAr experiments to light DM. Fig. 3.6 shows the sensitivity of our simulated LAr experiment, called Toy Experiment Analysis of Liquid Argon Behavior, `TEA-LAB` [2]. The red curve is the result by the DarkSide collaboration,
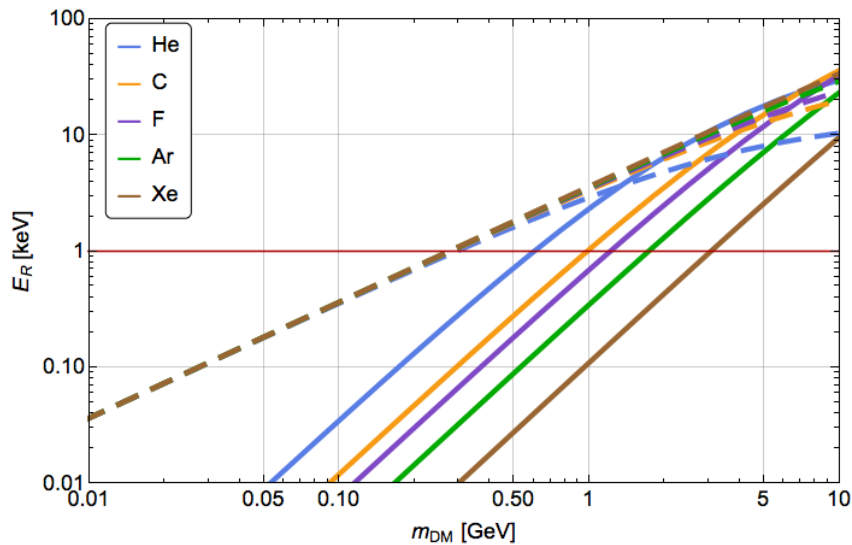
Figure 3.5. Kinematic limits of the nuclear recoil (solid) and Migdal effect (dashed) for different atoms: He (blue), C (orange), F(purple), Ar (green) and Xe (brown).

the green curve is the Migdal result by the XENON1T collaboration and finally the blue curve is the sensitivity of `TEA-LAB` with an exposure of 6786 kg day, chosen to match the exposure of the 2018 DarkSide-50 S2-only analysis [1]. It is possible to notice that this result is comparable with the XENON1T's one, and it even surpasses its sensitivity for dark matter masses below ~ 0.1 GeV.

There is therefore great interest in the Migdal effect in the field of DM direct detection. However, this effect has only been observed for α and β decays, lacking experimental confirmation in nuclear scattering. A compelling question is then if direct detection experiments are able to measure this effect in nuclear scattering, for instance by using a neutron source. More details on the possible signatures and detection techniques will be given in Sec. 8, where we will investigate the possibility of measuring it using a CYGNO-like detector.
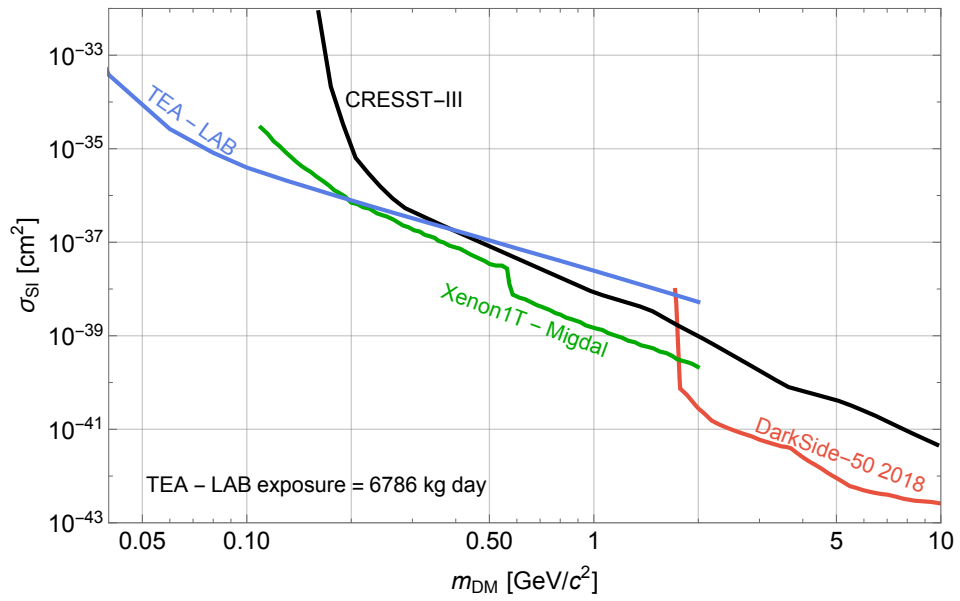
Figure 3.6. Expected sensitivity on the σ_{SI} exploiting the Migdal effect for the TEA-LAB
simulated experiment using an exposure of 6786 kg day (blue). The red curve is the
current DarkSide-50 bound [1], the green is the bound from the XENON1T Migdal
search [90] and the black one is the bound by the CRESST-III experiment [64].

Chapter 4

Bayesian analysis approach

In this thesis, we adopt a Bayesian approach to infer the upper bound and to estimate the expected experimental sensitivities to the interaction of DM candidates with ordinary matter. Similar approaches for the analysis of DM experimental data have already been deployed [4, 98–105], although they are not frequent among analyses carried out by the experimental collaborations, as for example [1, 64, 66, 87, 90, 106].

The possibility to associate a probability value to any uncertain quantity is at the basis of the Bayesian approach to data analysis. This may apply to the outcome of a measurement before it is carried out, to a parameter of interest within a given model, as well as to the model itself. Theoretical models, parameters of interests, and results of experiments before they are carried out are intended as uncertain propositions connected by the rules of probability. Within this approach we can compute, at least in principle, the probability of any specific proposition given some state of information. Exploiting the Bayes theorem, we can update the initial probability for a model or a parameter after new information is available in the form of experimental observations. The experimental information is fully contained in the so-called likelihood function. This term refers to the conditional probability for the data given the model when it is regarded as a function of the model's parameters while keeping the data fixed to the experimental observations.

We will use the following notations:

- $D = \{x_i\}$ represents the data, possibly organized in different classes i;

- E is the experimental exposure given in terms of the duration time T of the data-taking period and the fiducial mass M_{det} of the detector.

- H_i represents a specific hypothesis: H_0 is the background-only hypothesis according to which the known physics processes (backgrounds) are enough to explain the observations;

- H_{r_S} is the background-plus-signal hypothesis for which some DM signal with rate r_S is required to explain the data. We note that the two hypotheses H_0 and H_{r_S} are nested, since H_0 can be obtained for H_{r_S} by setting $r_S = 0$. For what concerns the inferential problem of constraining r_S, we will always work within the hypothesis H_{r_S}, assuming its validity.

- r_S indicates the expected rate of DM interaction for a given σ_{SI}^{DM} per unit mass and time, expressed in evt/kg/day. It is also a function of the mass m_{DM} of the DM candidate through the cross-section.

- r_B indicates the rate of the total background events expressed in evt/kg/day;

- $\pi(r)$ is the prior probability density function (*pdf*) for the generic parameter r and encapsulates all the available knowledge on the parameter r before the experiment is carried out;

- $\mathcal{L}(r; D)$, or simply $\mathcal{L}(r)$ is the likelihood for the generic parameter r of the hypothesis H_r, and coincides with $p(D \,|\, r,\, H_r)$;

- $p(r \,|\, D)$ is the posterior *pdf* for the generic parameter r given the data D;

- $\boldsymbol{\theta} = (r_B, \dots)$ is the list of parameters necessary to describe the experimental conditions or theoretical assumptions which are not exactly known but can vary according to their prior *pdf* $\pi(\boldsymbol{\theta})$, which for example can be the result of a set of calibration measurements. These are the so-called nuisance parameters, as we are not explicitly interested in inferring their posterior values;

- Ω is the nuisance parameters space.

The posterior p.d.f for the parameters of the model, which the result of the Bayesian inference, can be computed by means of the Bayes theorem as

$$p(r_S,\, \boldsymbol{\theta} \,|\, \{x_i\},\, H_{r_S}) = \frac{p(\{x_i\} \,|\, r_S,\, \boldsymbol{\theta},\, H_{r_S})\pi(r_S,\, \boldsymbol{\theta} \,|\, H_{r_S})}{\int_\Omega \int_0^\infty p(\{x_i\} \,|\, r_S,\, \boldsymbol{\theta},\, H_{r_S})\pi(r_S,\, \boldsymbol{\theta}|H_{r_S})dr_S d\boldsymbol{\theta}}. \tag{4.1}$$

The marginal *pdf* of the parameter of interest r_S is given by

$$p(r_S \,|\, \{x_i\},\, H_{r_S}) = \int_\Omega p(r_S,\, \boldsymbol{\theta} \,|\, \{x_i\},\, H_{r_S})d\boldsymbol{\theta}, \tag{4.2}$$

and similarly for any other parameter of the model.

We would like to remark that, especially in a complex model with many nuisance parameters and non-linear relations, the marginalization is a solid way of propagating the uncertainties. Other approaches, based for instance on the profiling of the likelihood function, only work under appropriate conditions, usually relying on the "gaussianity" of the likelihood or the linearity of the relations between the parameters - see Chap. 40 of Ref. [27] for a summary of these alternative approaches.

This chapter is structured as follows: Sec. 4.1 describes the ways in which the experimental sensitivity can be assessed in the Bayesian approach; in Sec. 4.2 the graphical formalism of the Bayesian networks for inference is outlined; in Sec. 4.3 the tools for model comparison in the Bayesian approach are described, and the case study of the analysis of the DAMA residuals we carried out in Ref. [4] is illustrated in Sec. 4.4.

4.1   Upper bounds and experimental sensitivity

4.1.1   90% Credible Interval upper bound

We compute the upper bound for the DM signal as the 90% Credible Interval (C.I.). This is defined as the value of $\sigma_{SI}^{DM}(m_{DM})$ corresponding to the 90% quantile of the posterior *pdf* for r_S:

$$r_S(90\%\,\text{C.I.}) : \int_0^{r_S(90\%\,\text{C.I.})} p(r_S \,|\, \{x_i\},\, H_{r_S})\, dr_S = 0.9. \tag{4.3}$$

In the Bayesian approach, the upper bound is a statement on the true value of the parameter of interest. The quantity $r_S(90\%\,\text{C.I.})$ has to be interpreted as the value below which we believe at 90% probability level the true value of r_S lies, given the present experimental information.

4.1.2   Prior choice

It is evident from Eq. (4.1) that the posterior *pdf* depends on the priors on all parameters. However, we have to distinguish the effect due to the priors on nuisance parameters from the one due to the prior on the parameter of interest. The former has the effect of averaging the posterior over the nuisance parameters space, which is an elegant way of propagating systematic effects on the parameter of interest. In addition, the prior on the nuisance parameters is often a parametrization of calibration measurements. The latter, although indispensable to invert the probability and get the posterior, has a degree of 'subjectivity' with potentially a significant impact on the posterior. The prior $\pi(r_S)$ represents the knowledge on r_S before the experiment is carried out, and gets updated by a factor proportional to the likelihood of the observed data. It is a critical term in many respects, and it should reflect the researcher's state of knowledge. Especially for searches where the sought quantity is unknown, and the search is pushed to the limit of the experimental sensitivity, the input from $\pi(r_S)$ might have a sizable effect on the posterior. The prior has thus to be well justified, and the posterior sensitivity to different prior choices needs to be explored.

In our case, r_S depends on m_{DM} and on σ_{SI}^{DM}. In all the analyses we will perform, m_{DM} will be fixed to an appropriate value. For what regards σ_{SI}^{DM}, in order to explore the sensitivity of the upper bound to the prior choice, in Ref. [2] we studied its behavior for a Migdal-only signal at $m_{DM} = 1$ GeV/c^2. We tested, in different possible domain ranges, four different prior choices: a uniform prior, a wider gamma prior with a shape $k = 1$ and a scale $\theta = 10^{-37}$ cm^2, a narrower gamma prior with a shape $k = 1$ and a scale $\theta = 10^{-38}$cm^2 and a uniform prior in $\log(\sigma_{SI}^{DM}/\text{cm}^2)$ (we will call this prior "loguniform"). In principle, there is no need to restrict the domain to a finite interval, but both the uniform and the loguniform distributions are not normalizable otherwise. In addition, there could be physical motivations that define a reasonable interval. In our opinion, for the DM direct detection experiments based on noble atoms Time Projection Chambers (TPCs), the natural constraints come from upper bounds imposed by previous experiments as for example XENON1T [90] or CRESST-III [64]. From below, we can use two arguments: the first is that below a

certain value - usually $\sigma_{SI}^{DM} = O\left(10^{-44}\,\text{cm}^2\right)$, depending on the DM mass - the rate would be dominated by neutrino coherent scattering [107] (the so-called 'neutrino floor'); the second is that the experimental sensitivity is not infinite, but it is limited to a certain cross-section. Therefore, the choice of the domain can be driven either by physical considerations about the previous experiments, or by the maximum experimental sensitivity that an experiment like ours can reach before hitting the neutrino floor. The result we draw in Ref. [2] is that, in a LAr TPC experiment, we can quantify the dependence of the bound from the prior choice in a factor as big as 10, and this confirms the importance of choosing the prior in a reasonable and coherent way. For simplicity, to avoid possible bias coming from the prior choice and to make our results comparable with other statistical approaches, in the rest of the thesis we will report upper bounds obtained using a flat prior.

4.1.3 Experimental sensitivity and $\mathcal{R}$ function

A meaningful way to report the experimental sensitivity to the sought phenomenon which is as much as possible independent of the priors is given by the Bayes factor.

The posterior *pdf* for a signal rate of r_S given a background rate of r_B, and x observed events can be normalized to the posterior for $r_S = 0$, obtaining

$$\frac{p(r_S \mid x, r_B)}{p(r_S = 0 \mid x, r_B)} = \frac{\mathcal{L}(r_S \mid r_B)}{\mathcal{L}(r_S = 0 \mid r_B)} \cdot \frac{\pi(r_S)}{\pi(r_S = 0)}, \tag{4.4}$$

where we notice that the first factor on the right-hand side is independent of priors, and it is simply given by the likelihood ratio of the two hypotheses

$$\mathcal{R}(r_S \mid x, r_B) = \frac{\mathcal{L}(r_S \mid r_B)}{\mathcal{L}(r_S = 0 \mid r_B)}. \tag{4.5}$$

The properties of the $\mathcal{R}$ function have been discussed in great detail for a similar case study in Ref. [108]. Here we only mention that $\mathcal{R}$ has the probabilistic interpretation of hypotheses' belief updating ratio. It is equal to 1 in the limit $r_S \to 0$, in this limit the experimental sensitivity is lost, and thus the experiment does not change the relative belief. While $\mathcal{R} \to 0$ for large r_S, where the posterior density for r_S vanishes, no matter how strong it was before. In addition, the quantity $\mathcal{R}$ is used as test statistic in the frequentist approach to limit settings, for details see Ref. [27] (section 39-Statistics) and Ref. [109, 110].

In the simple case of a Poisson process of intensity $(r_S + r_B)E$, where E is the exposure, and observed counts x the likelihood is proportional to the quantity $e^{-(r_S+r_B)E}\left[(r_S+r_B)E\right]^x$, thus

$$\mathcal{R}(r_S \mid x, E, r_B) = e^{-r_S E}\left(1 + \frac{r_S}{r_B}\right)^x \tag{4.6}$$

In Ref. [2] we developed a simulated experiment of a LAr TPC with an exposure of $6786.0\,\text{kg}\,\text{d}$, which is the exposure of the 2018 DarkSide-50 dataset. As for the analysis of the previous section, we evaluated the $\mathcal{R}$ function for a Migdal-only signal at $m_{DM} = 1\,\text{GeV}/c^2$. The results are reported on the left side of Figure 4.1 (black lines), where the green lines represent the corresponding results for a greater value of
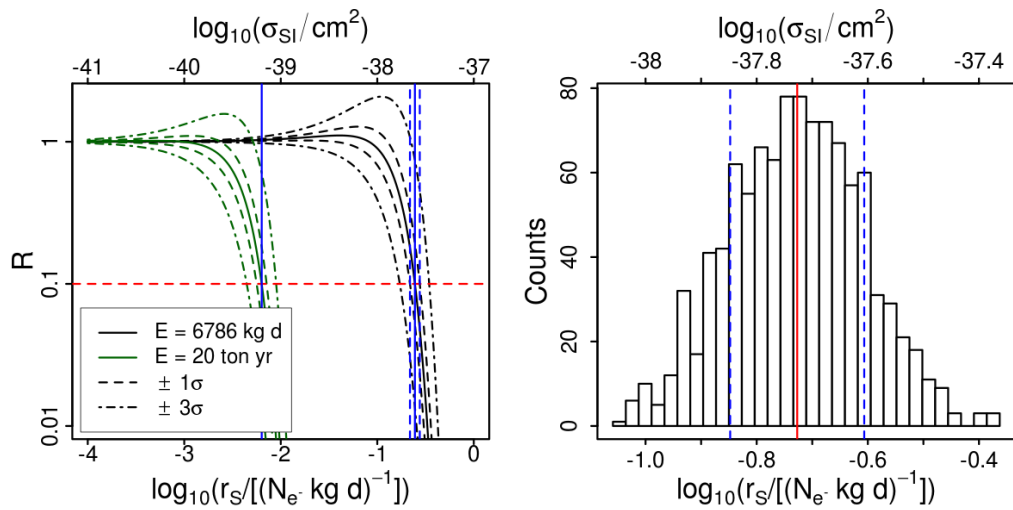
Figure 4.1. **Left**: $\mathcal{R}$ function for a Migdal-only signal at $m_{DM} = 1 \text{ GeV}/c^2$ for the atomic shells $n = 1, 2$, obtained using pseudo-datasets generated from the background-only likelihood with the exposure $E = 6786 \text{ kg d}$ (black lines) and an exposure $E = 20 \text{ton y} \sim 7.3 \times 10^6 \text{ kg d}$ (green lines). The dashed lines and the dashed-dotted lines represent the Bayes factor computed varying the expected rate by $\pm\sigma$ and $\pm3\sigma$, respectively. **Right**: Histograms of $r_S(90\% \text{ C.I.})$ for a Migdal-only signal at $m_{DM} = 1 \text{ GeV}/c^2$ for the atomic shells $n = 1, 2$ obtained using pseudo-datasets generated from the background-only likelihood with an exposure $E = 6786 \text{ kg d}$.

the exposure (i.e. 20 ton yr which is roughly 10^3 times the exposure $E = 6786 \text{ kg d}$). From this figure we see that the informative region where the simulated experiment has sensitivity and $\mathcal{R} \to 0$ is starting from $r_S \sim 10^{-39} \text{ cm}^2$. Any conventional value of r_S in this region would be representative of the experimental sensitivity. The statistical uncertainty associated with this value is computed using pseudo-data generated varying the expected rate by $\pm\sigma$. A possible value of r_S representative of this region could be such as that $\mathcal{R}(r_S) = 0.10$, which corresponds to a probability update ratio of 10% with respect to the null hypothesis.

As well described in Ref. [108], in order to extract any probabilistic statement on r_S from $\mathcal{R}$ one has to add the information about the prior. We would like to stress that there is a conceptual difference in using the 90% C.I. upper bound or taking the r_S such that $\mathcal{R} = 0.1$: the former is the cumulative of the posterior *pdf* and then it takes into account all the possible values of r_S from 0 to $r_S(90\% \text{ C.I.})$ as well as the prior choice; the latter is the likelihood ratio, and it is a punctual comparison, namely it takes into account only one single possible value of r_S (the one which solves $\mathcal{R}(r_S) = 0.1$), and it is prior independent.

For our case study, however, in Ref. [2] we showed that the two methods give very similar results.

4.2 Graphical method for Bayesian Inference

In the context of the Bayesian approach, in the view of determining the likelihood function, which in general for realistic experiments could be complicated, it is use-

ful to adopt a graphical method, the so-called Bayesian network, to represent the probabilistic connection between variables. According to this method, the *pdf* of a collection of random variables can be represented by a network made of arrows and nodes where:

- the nodes represent the variables;

- a solid arrow starting from a certain node x pointing to another node y represents a probabilistic link between the two variables;

- a dashed arrow starting from a certain node x pointing to another node y represents a deterministic link between the two variables.

- a gray box labeled with n means that a particular part of the network is repeated n times.

4.2.1   Bayesian Networks from examples

In order to better understand this approach, in the rest of this paragraph we will show how to build the Bayesian networks in some simple cases.

Let's consider at first two random variables A and B. The joint *pdf* of the two variables is given by the multiplication rule of probability as

$$p(A,B) = p(A|B)P(B) = p(B|A)p(A), \tag{4.7}$$

where $p(A|B)$ indicates the *pdf* of the variable A conditioned to the value of the variable B. This joint *pdf* can be graphically represented as
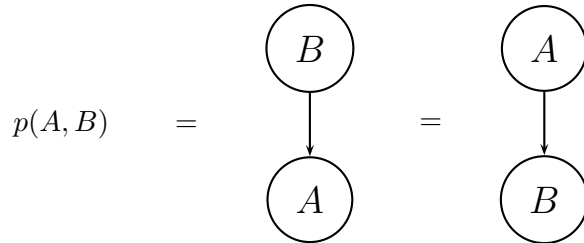
where the nodes represent the two variables A and B, and the (solid) arrows represent the conditional *pdf*s appearing in Eq. 4.7.

If we now consider a third variable C, the joint *pdf* of the three events is

$$p(A,B,C) = p(A|B,C)p(B,C) = p(A|B,C)p(B|C)p(C). \tag{4.8}$$

In this case, following the conditional relations on the right-hand side of the above equation, the joint *pdf* of the three events can be graphically represented as
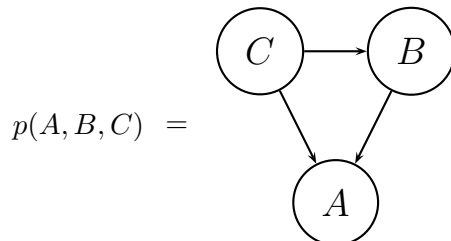
No assumption has been made on the independence of A, B and C. In the case in

which all the variables are independent then $p(A,B,C) = p(A)p(B)p(C)$: the corresponding network will be therefore completely disconnected, i.e. no arrow would connect the three nodes because between them there is no probabilistic or deterministic connection. In the case in which one of the three variables, for example C, is independent of the other two and vice-versa, then $p(A,B,C) = p(A|B)p(B)p(C)$: the node C would be therefore disconnected from the other two, and the remaining nodes are connected in the same way as in the case of Eq. 4.7. Therefore, the remaining interesting cases are three:

- *head-to-head*: $p(B|C) = p(B)$. Then the joint probability of the three variables can be depicted as
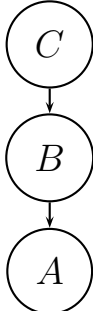
$$p(A,B,C) = p(A|B,C)p(B)p(C) \quad =$$

This configuration gives us the opportunity to introduce the standard terminology: we will refer to C and B as the "parent" nodes with respect to A, we will refer to the A as the "child" node with respect to B and C, and we will refer to C (B) as the "co-parent" node with respect to B (C) - for a more detailed discussion see Ref. [111]. If we suppose to observe the variable A, and thus to know the status of the correspondent node, the two nodes C and B are no more independent. In fact, in this configuration the joint probability of C and B, conditioned to the observed value of A, is given by

$$p(B,C|A) = \frac{p(A,B,C)}{p(A)} = \frac{p(A|B,C)p(B)p(C)}{p(A)}, \tag{4.9}$$
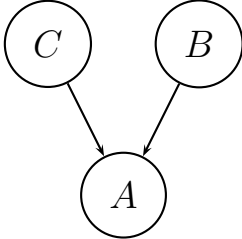
which in general cannot be factorized in the form $p(B|A)p(C|A)$. In this case, the two variables B and C are called "conditionally dependent".

- *head-to-tail*: $p(A|B,C) = p(A|B)$. Then the joint probability of the three variables can be depicted as

$$p(A,B,C) = p(A|B)p(B|C)p(C) \quad =$$

If we suppose to observe the variable B, and thus to know the status of the correspondent node, the two nodes A and C become independent. In fact, in this configuration the joint probability of A and C, conditioned to the observed value of B, is given by

$$p(A,C|B) = \frac{p(A,B,C)}{p(B)} = \frac{p(A|B,C)P(C|B)p(B)}{p(B)} = p(A|B)P(C|B), \tag{4.10}$$

where in the last equality we used the *head-to-tail* hypothesis $p(A|B,C) = p(A|B)$. In this case the two variables A and C are called "conditionally independent".

- *tail-to-tail*: $p(A|B,C) = p(A|C)$. Then the joint probability of the three variables can be depicted as

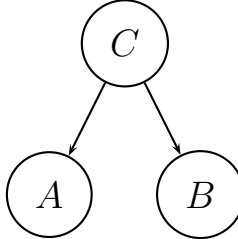$$p(A, B, C) = p(A|C)p(B|C)p(C) \quad =$$

If we suppose to observe the variable C, and thus to know the status of the correspondent node, the two nodes A and B are independent. In fact, in this configuration the joint probability of A and B, conditioned to the observed value of C, is given by

$$p(A, B|C) = \frac{p(A, B, C)}{p(C)} = \frac{p(A|B,C)P(B|C)p(C)}{p(C)} = p(A|C)P(B|C), \quad (4.11)$$

where in the last equality we used the *tail-to-tail* hypothesis $p(A|B,C) = p(A|C)$. As in the previous case, the two variables A and B are called "conditionally independent".

We illustrated how it is possible to represent, in three simple cases, the joint probability of a set of random variables, showing how to easily decompose the joint conditional probability of a certain node and introducing the notion of conditional independence. However, more complex networks can always be treated as the composition of the previous three interesting cases. In this respect, of particular interest is the notion of "Markov blanket" [112]: let's suppose to consider a certain node X in a particular Bayesian network. The Markov blanket of the node X is defined as the set of all the nodes that are parents, co-parents or children with respect to X. It can be easily demonstrated that the *pdf* of X conditioned to all the nodes of the network is equal to the probability of X conditioned just to the nodes of its Markov blanket. In other words, the Markov blanket is the minimal set of nodes that separates X from the rest of the network [111] from a probabilistic point of view.

Finally, let's consider the more realistic case of a parametric fit. Let's suppose to have performed some measurement $D = \{(x_i, y_i)\}$, where i is an index running over the number N of experimental data points. We know that, according to a certain theoretical model H, the true values (μ_i^x, μ_i^y) of each point (x_i, y_i) are connected by a certain deterministic relation $\mu^y = g(\mu^x, \vec{\theta})$, where $\vec{\theta}$ are the parameters that we want to determine with the fit on the experimental data. The likelihood for this model can be written as

$$\mathcal{L} = p\left(D \middle| \{\mu_i^x\}, \{\mu_i^y\}, \vec{\theta}, H\right) = \prod_{i=1}^{N} p\left(x_i|\mu_i^x, H\right) p\left(y_i|\mu_i^y = g(\mu_i^x, \vec{\theta}), H\right) \quad (4.12)$$

In terms of Bayesian network, this likelihood can be drawn as
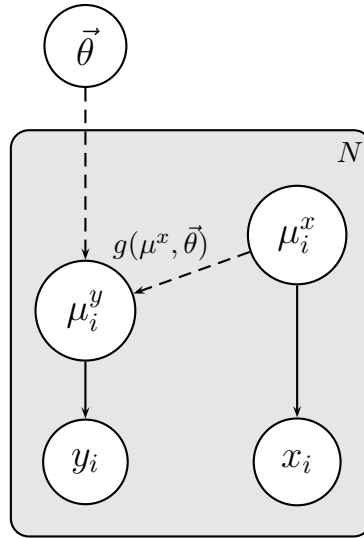
where the dashed arrows represent the deterministic relation given by $g(\mu^x, \vec{\theta})$ and the gray box labeled with N stands for the fact that, because of the product in Eq. 4.12, we have N copies of the portion of the network that is inside the box. This approach is very intuitive: the N measured points $\{(x_i, y_i)\}$ depend in a probabilistic way on their true values $\{(\mu_i^x, \mu_i^y)\}$. The true values and the model parameters $\vec{\theta}$ are in turn linked by the deterministic relation $\mu^y = g(\mu^x, \vec{\theta})$. Then, since in this approach the dependencies between the variables become clear, in the next sections we will often give the likelihood also in terms of the corresponding Bayesian network.

4.2.2 Technical implementation using JAGS and BAT

In the Bayesian approach, once the likelihood $\mathcal{L}$ is written and the priors π on the parameters of the model H are chosen, the goal of the inferential process is to compute the posterior on the parameters of interest $\vec{\theta}_I$, namely the *pdf* of the parameters of interest given the observation of the data D. In order to do that, an integration over the nuisance parameters $\vec{\theta}_N$ is needed, in fact

$$p(\vec{\theta}_I | D, H) = \frac{\int \mathcal{L}(D|\vec{\theta}_I, \vec{\theta}_N, H)\pi(\vec{\theta}_I, \vec{\theta}_N|H)d\vec{\theta}_N}{\int \mathcal{L}(D|\vec{\theta}_I, \vec{\theta}_N, H)\pi(\vec{\theta}_I, \vec{\theta}_N|H)d\vec{\theta}_I d\vec{\theta}_N} \tag{4.13}$$

This integral is usually computed by means of numerical methods, i.e. a Markov Chain Monte Carlo (MCMC). When using the Bayesian Networks, the most straightforward way to proceed is to implement directly the network in `JAGS` (Just Another Gibbs Sampler) [113], a software which implements a MCMC algorithm called Gibbs Sampling and is interfaced with `R` [114] in the package `rjags` [115]. However, in those cases in which the model is very complex, in order to have better performances we decided to use the Metropolis-Hasting MCMC algorithm as implemented in `BAT` (Bayesian Analysis Toolkit) [116], a `C++` set of libraries to implement a Bayesian analysis.

4.3 Model comparison in the Bayesian approach

In the following, we summarize how models can be compared in the Bayesian approach. Let's suppose to have a dataset $D = \{x_i\}$ composed of n measurements x_i, a set of models or hypotheses H_i, where each hypothesis could in general depend on a vector of parameters $\vec{\theta}_i$. Within a model, we assume to know the prior probability $\pi(H_i)$, the likelihood function $\mathcal{L}(H_i, \vec{\theta}_i; D)$, and the prior *pdf* of all the model's parameters $\pi(\vec{\theta}_i|H_i)$.

Suppose we want to compare two different models H_i and H_j. To compare these models we can compute the ratio of the posterior probabilities as

$$O_{ij} \equiv \frac{p(H_i|D)}{p(H_j|D)}. \tag{4.14}$$

This ratio gives the odds in favor of one model over the other. To express the odds as a function of the likelihoods and priors, we shall apply the Bayes theorem; that is, for example for the numerator of Eq. (4.14),

$$p(H_i|D) = \frac{p(D|H_i)\,\pi(H_i)}{\pi(D)}. \tag{4.15}$$

The denominator $\pi(D)$ is the prior probability of data, and it can be more easily interpreted if expressed in terms of the likelihoods by imposing that $\sum_i p(H_i|D) = 1$. To do this, we need to have a complete class of hypotheses, and this is not often the case. However, if we are only interested in the odds ratio, we can neglect this term as it cancels out in the odds. The ratio between the probabilities of the two models becomes

$$O_{ij} \equiv \frac{p(D|H_i)}{p(D|H_j)} \times \frac{\pi(H_i)}{\pi(H_j)}. \tag{4.16}$$

The data dependent part is the so called "Bayes factor" (BF), and it is defined as

$$BF_{ij} \equiv \frac{p(D|H_i)}{p(D|H_j)} = \frac{\mathcal{L}(H_i; D)}{\mathcal{L}(H_j; D)}, \tag{4.17}$$

where $\mathcal{L}(H_i; D)$ is the so-called marginal likelihood defined as

$$\begin{aligned} \mathcal{L}(H_i; D) &\equiv p(D|H_i) \\ &= \int p(D, \vec{\theta}_i|H_i)\, d\vec{\theta}_i \\ &= \int p(D|H_i, \vec{\theta}_i) p(\vec{\theta}_i|H_i)\, d\vec{\theta}_i \\ &= \int \mathcal{L}(\vec{\theta}_i, H_i; D)\pi(\vec{\theta}_i|H_i)\, d\vec{\theta}_i. \end{aligned} \tag{4.18}$$

In the hypothesis in which, due to our a priori ignorance, all the model's priors are equal, considering the odds ratio is equivalent to considering the Bayes factor.

The Bayes factor BF_{ij} encapsulates all the new information associated to the measurements, although, for parametric models, it could critically depend on the choice of the priors $\pi(\vec{\theta}_i|H_i)$. In any real situation, the integral of Eq. (4.18) is

not solvable analytically and can only be estimated numerically. Nevertheless, in the case where the information content of the measurements is such that the priors can be considered sufficiently vague in the range of the parameters space where the likelihood is sizable, we can consider the prior as a constant and use the Laplace approximation to estimate the integral. This is a very useful approximation to get sense of the different contributions to the marginal likelihood. For simplicity, let's assume the model depends just on a single parameter θ_i, the best fit to the data occurs for $\theta = \hat{\theta}$ (in the following we will use $\hat{x}$ to refer to the best fit value of the quantity x), the likelihood is reasonably normal, and the prior is flat over a range $\Delta\theta_i$. Under these assumptions the integral can be approximated as

$$
\begin{aligned}
p\left(D|H_i\right) &= \int \mathcal{L}\left(\theta_i, H_i; D\right) \pi(\theta_i|H_i)\, d\theta_i, \\
&\simeq \frac{1}{\Delta\theta_i}\, \sqrt{2\pi}\, \sigma(\hat{\theta}_i)\, \mathcal{L}(\hat{\theta}_i, H_i; D).
\end{aligned}
\tag{4.19}
$$

The last expression makes it clear that the marginal likelihood corresponds to the likelihood evaluated at the best fit value of θ_i weighted by a volume factor corresponding to the parameter's uncertainty. In this oversimplified example, the proportional factor between the global and the best fit likelihood is given by the ratio of two quantities, namely $\Delta\theta_i$ and $\sigma(\theta_i)$, which control respectively the possible range of θ_i before and after the data are seen. This factor is often referred to as the Ockham's factor as it naturally penalizes a model by a quantity proportional to the necessary complexity of the model to get a good fit with respect to the a priori model complexity; the ideal case where the model is a good one and the a priori variability of the parameter is just enough to fit the data, the Ockham's factor is of order one. The Ockham's factor enters the BF through the marginal likelihood, and thus, when we compare two models, we automatically account for the potential difference in their complexity.

In order to visualize that, let's make a useful example. Let's consider two models:

- *Model A*: this models has only one parameter, let's call it θ_A, and the prior over this parameter is a uniform distribution in the range $[\theta_A^1, \theta_A^2]$ of width $\Delta\theta_A$, and thus the *pdf* is

$$
p\left(\theta_A|H_A\right) = \frac{1}{\Delta\theta_A}
\tag{4.20}
$$

- *Model B*: this models has two parameters, θ_{B1} and θ_{B2}, that are uncorrelated and the prior over these parameters is a uniform distribution of width $\Delta\theta_{B1}$ and $\Delta\theta_{B2}$, respectively

$$
p\left(\theta_{B1}, \theta_{B2}|H_B\right) = p\left(\theta_{B1}|H_B\right) p\left(\theta_{B2}|H_B\right) = \frac{1}{\Delta\theta_{B1}} \frac{1}{\Delta\theta_{B2}}
\tag{4.21}
$$

Assuming normal likelihoods, and using the Eq. (4.19), the BF can be estimated as

$$
BF_{AB} \simeq \frac{p(D|\hat{\theta}_A, H_A)}{p(D|\hat{\theta}_{B1}, \hat{\theta}_{B2}, H_B)} \times \frac{\sqrt{2\pi}\sigma(\hat{\theta}_A)}{\Delta\theta_A} \left/ \left(\frac{(2\pi)\sigma(\hat{\theta}_{B1})\sigma(\hat{\theta}_{B2})}{\Delta\theta_{B1}\Delta\theta_{B2}} \right) \right. = LR_{AB} \times OF_{AB},
\tag{4.22}
$$

where we defined the likelihood ratio (LR_{AB})[1] and the Ockham's factor (OF_{AB}) as

$$LR_{AB} = \frac{p(D|\hat{\theta}_A, H_A)}{p(D|\hat{\theta}_{B1}, \hat{\theta}_{B2}, H_j)},$$

$$OF_{AB} = \frac{\sqrt{2\pi}\sigma(\hat{\theta}_A)}{\Delta\theta_A} \bigg/ \left(\frac{(2\pi)\sigma(\hat{\theta}_{B1})\sigma(\hat{\theta}_{B2})}{\Delta\theta_{B1}\Delta\theta_{B2}} \right) \qquad (4.23)$$

The message behind Eq. (4.22) is that even though a model with more parameters can be more flexible and thus better fit the data producing a higher likelihood, one has to pay the price of having a more complex model. In other words: even if the likelihood ratio pushes towards a more complex model which often better fits the data, on the other hand the Ockham's factor penalizes it.

As already said, for parametric models the Bayes factor critically depends on the choice of priors, that, in this case, is represented by the choice of the widths $\Delta\theta_i$. For models with the same number of parameters, especially if the parameters have the same physical meaning, the Ockham's factor should be of order one, and the BF should fairly be insensitive to priors' choice.

For models with a different number of parameters, one has to pay particular attention to avoid introducing biases due to an 'unreasonable choice of priors'. In the next section we will describe the case study of the modulation of the DAMA residuals, and we will show how to deal with the prior choice issue, starting by comparing models with just one parameter and then moving to more complex and realistic scenarios exploring the sensitivity of our conclusions to the prior's choice.

Finally, two other useful criteria to quantify the ability of a model to describe the observations are the Bayesian Information Criterion (BIC) [117] and the Akaike Information Criterion (AIC) [118], a review of which is given for example in Ref. [119]. Given a dataset of size n, and a model A, with a number k_A of parameters $\vec{\theta}_A$ whose value that maximize the likelihood is $\hat{\vec{\theta}}_A$, the BIC is defined as

$$BIC_A = k_A \ln(n) - 2\ln\left(\mathcal{L}(\hat{\vec{\theta}}_A, H_A; D)\right); \qquad (4.24)$$

and the AIC is defined as

$$AIC_A = 2\,k_A - 2\ln\left(\mathcal{L}(\hat{\vec{\theta}}_A, H_A; D)\right); \qquad (4.25)$$

According to these criteria, the smaller the value of the BIC (AIC) is, the better the description of the observations.

---

[1]Note that for normal likelihoods the LR_{AB} can be expressed in terms of $\chi^2_{A(B)} = \sum_i \left[\frac{(x_i - \mu_{A(B)}(x_i))^2}{\sigma_i^2} \right]$ as: $LR_{AB} = \exp\left[-\frac{\chi^2_A - \chi^2_B}{2} \right]$, where $\mu_{A(B)}(x_i)$ represents the true value of the measurement x_i in the model $A(B)$. This means that the LR_{AB} of the two models evaluated at their best fit values is simply given by $\exp[-\Delta\hat{\chi}^2_{AB}/2]$. In a frequentist approach to model comparison, the $\Delta\hat{\chi}^2_{AB}$ is often taken as a test-statistic and its probability distribution used to compute p-values. For nested models, where the more general model contains the simpler plus k additional parameters, the Wilks' theorem says that $pdf(\Delta\hat{\chi}^2_{AB})$ is itself a χ^2 distribution with k degrees of freedom. For non-nested models one would have to estimate $pdf(\Delta\hat{\chi}^2_{AB})$, possibly by sampling it with pseudo-experiments.

When comparing two models, the ΔBIC_{AB} and ΔAIC_{AB} can be defined as

$$
\begin{aligned}
\Delta BIC_{AB} &= BIC_A - BIC_B = \ln(n)(k_A - k_B) - 2\ln(LR_{AB}). &(4.26)\\
\Delta AIC_{AB} &= AIC_A - AIC_B = 2(k_A - k_B) - 2\ln(LR_{AB}). &(4.27)
\end{aligned}
$$

The previous equations show how the ΔBIC_{AB} and ΔAIC_{AB} are closely related to the LR_{AB}, plus a penalty term that penalizes the model which is more complex.

For the interpretation of the odds ratio or Bayes factor, we refer to the criterion based on Jeffreys' scale [120]: a value of > 10 represents strong evidence in favor of model A, and a value of > 100 represents decisive evidence. Similar evaluation can be done based on the ΔBIC and the ΔAIC.

4.4 Bayesian model comparison in a case study: the DAMA residuals

In a recent publication [4] we applied the Bayesian model comparison to a specific physics problem: the interpretation of the DAMA modulation as induced by a time dependent slowly varying background. The results of the DAMA/NaI and DAMA/LIBRA experiments are interpreted by the DAMA collaboration as a strong evidence of the presence of dark matter particles in the galactic halo [31,32,121–125]. These conclusions derive from the compatibility of the annually modulated rate with a sinusoidal signal, characterized by a phase and a period in agreement with those expected from the interaction of DM particles.

In a recent paper [126] it has been discussed the possibility that, due to the analysis procedure followed by the DAMA collaboration, the observed annual modulation could be reproduced by a slowly varying time-dependent background. This possibility has been used to interpret the modulation of residuals of the single-hit scintillation rate as a function of time, published by the DAMA collaboration [32,121,124].

The argument proposed in Ref. [126] goes as follows: since the residual rate is computed by subtracting the average total rate in every cycle of data-taking, if the background rate is not constant over time this procedure can generate a time modulated rate. More precisely, in the hypothesis of the actual presence of a DM signal, the total rate $r(t)$ is given by:

$$
r(t) = r_0(t) + A\cos\left(\frac{2\pi t}{T} - \phi\right), \tag{4.28}
$$

where $r_0(t)$ is the time-dependent background, A is the amplitude of the oscillating DM signal, T is the period of the oscillation, which is equal to 1 year, and ϕ is a phase such that the peak of the oscillation is around the 2$^{\text{nd}}$ June, as expected in DM galactic halo model. If r_0 is constant over time, then one can choose a time window of width Δ equal to any multiple of T and average the rate in that window to isolate the background contribution. Under such an hypothesis, the residuals obtained by subtracting window by window this average are:

$$
S(t) \equiv r(t) - \langle r(t) \rangle_\Delta = A\cos\left(\frac{2\pi t}{T} - \phi\right) \tag{4.29}
$$

and then the signal is isolated from the background. This is what has been done by the DAMA collaboration, with the time windows Δ chosen to be of the order of 1 year.

On the other hand, as it has been pointed out, if the background is slowly varying over time, for example linearly, the choice of using time windows of the order of the period of the wanted signal can produce residuals with a modulation of the same period. The possibility of a time-varying background is supported by the fact that in underground detectors, especially in the keV energy range, the features of the background are not completely modeled. In particular, in Ref. [126] explanations due to out-of-equilibrium physical or instrumental effects are considered. The authors of Ref. [126] conclude that the available data could not safely exclude the extreme possibility that the modulation in the residual rate is produced by a slowly varying background only. Such a conclusion is based on the debatable argument that the DAMA residuals could be fitted by a linearly growing background with a $\chi^2/$ d.o.f $\simeq 1$.

In Ref. [104] a Bayesian comparison between the cosine and the null hypothesis on the DAMA (2-6) keVee energy window dataset has been performed, leading to a decisive evidence against the null hypothesis. The aim of our work [4] was to perform a thorough comparison of the two hypotheses of a cosine modulation and a slowly varying background, and give a quantitative conclusion on their performance by studying the likelihood ratio, the model complexity, and the posterior probabilities.

4.4.1 The DAMA and DAMA/LIBRA data

The DAMA and DAMA/LIBRA experiments use ultra-radiopure NaI(TI) scintillating crystals as active target, coupled with photomultipliers to measure the amount of deposited energy. The single-hit scintillation events rate is used to look for a possible signal of DM interactions with matter over the large backgrounds due to natural radioactivity of the detector and surrounding environment. The event rate, expressed in cpd/kg/keVee (where "ee" stands for electron equivalent), has only been published [32, 121, 124] after the subtraction of its time-average over each cycle (of roughly one-year duration) following the procedure outlined in the previous section. Each cycle starts every year roughly around the beginning of September.

The residual rates are available in different energy windows: for the DAMA/NaI and the DAMA/LIBRA Phase I experiments, the energy windows are (2-4), (2-5), and (2-6) keVee; for the DAMA/LIBRA Phase II the energy windows are (1-3), (1-6), and (2-6) keVee. The (2-6) keVee energy window is the one with the smallest uncertainties, and it is common to the three phases. For this reason, to give a quantitative and fair comparison of the different hypotheses in the three stages, we decided to study only the residuals in the (2-6) keVee energy window.

In principle, our analysis could be easily extended to other energy bins, although we believe our conclusions would not change significantly.

4.4.2 A possible bias in the signal subtraction

The algorithm used by the DAMA collaboration to extract the time-dependent residual rate from the data has a two-fold objective: account for any constant or slowly
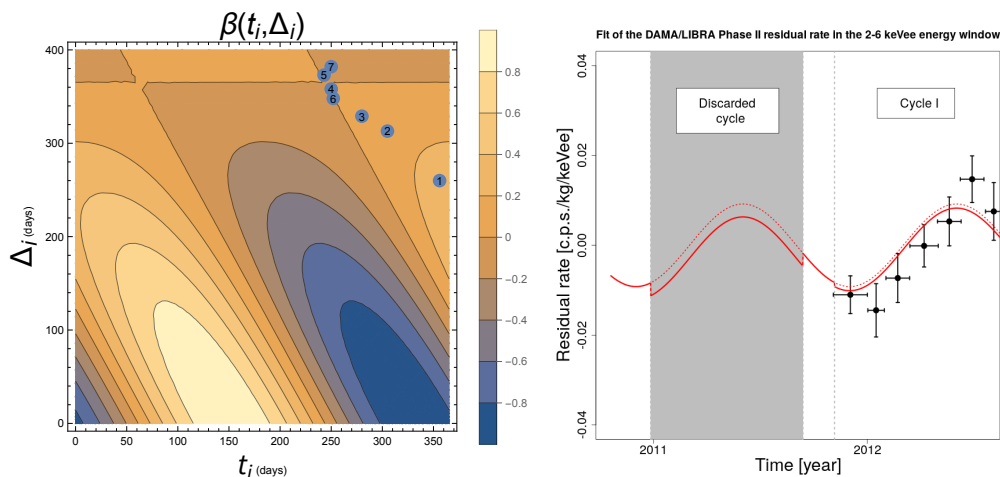
Figure 4.2. **Left**: Contour plot of the function β given in Eq. (4.30); the points represent
the data-taking cycles given in table 1 of Ref. [32]. In particular, the point number 1
corresponds to the first and discarded data-taking cycle. **Right**: Fit of a sinusoidal
model (red) to the DAMA/LIBRA Phase II residual rate in the (2-6) keVee energy
window, obtained taking into account the subtraction-bias effect. The dotted red line
is the corresponding sinusoidal modulation without the subtraction. The gray are is
the region discarded as explained in Ref. [32]: since, assuming the sinusoidal signal,
the effect is perfectly quantifiable, in principle also this region could be used for the
residuals' analysis.

varying component while keeping the sinusoidal time structure of a hypothetical
DM signal. These requirements put constraints on the optimal time intervals of the
data-taking cycle.

The variance of the sinusoidal signal is given by the quantity $(\alpha - \beta^2)$, with the
time averages $\alpha = \langle \cos^2 2\pi/T(t - t_0) \rangle$ and $\beta = \langle \cos 2\pi/T(t - t_0) \rangle$. For $\Delta = T$, the
quantity $(\alpha - \beta^2)$ is 0.5 (i.e. $\beta = 0$). In Ref. [32] the value of 0.5 is taken as a
figure of merit to assess whether the detectors were operational evenly throughout
the period of the modulation of one year. To the best of our understanding, a value
of $(\alpha - \beta^2)$ much different from 0.5 implies that either the average of the signal is
not vanishing ($\beta \neq 0$), or the data-taking period is not close to one year, or both.

The DAMA collaboration used the argument that $(\alpha - \beta^2) \neq 0.5$ to remove the
first cycle of data-taking in table 1 of Ref. [32] from the analysis.

The procedure to extract the residual rate might induce a subtraction of the
signal if the time interval is not chosen carefully. In particular, for a data-taking
starting in t_i and extending to $t_i + \Delta_i$, the average of the signal is a function of t_i
and Δ_i:

$$\beta(t_i, \Delta_i) = \frac{1}{\Delta_i} \int_{t_i}^{t_i + \Delta_i} \cos\left(\frac{2\pi}{T}(t' - t_0)\right) dt'. \qquad (4.30)$$

The contour plot of $\beta(t_i, \Delta_i)$ as a function of t_i and Δ_i is given in Fig. 4.2. Inde-
pendently of t_i, if Δ_i is equal to T, as expected, $\beta(t_i, \Delta_i = T) = 0$. This is the
approximate condition used by the DAMA collaboration to include a data-taking
cycles in the modulation analysis. However, as it is shown in Fig. 4.2, there are
other combination of t_i and Δ_i that result in $\beta(t_i, \Delta_i) = 0$; of particular relevance

there is the one with t_i and $t_i + \Delta_i$ chosen symmetrically with respect to the time when the cosine is zero. Figure 4.2 also reports the values of $\beta(t_i, \Delta_i)$ for the 7 data-taking cycles reported in Ref. [32]; we note that, for the first and the second cycle, the relative contribution of the signal average is of the order of 10%. In particular, we computed the effect of such a subtraction for the first and discarded cycle and represented it graphically in the plot on the right side of Fig. 4.2. The plot shows the expected behavior in the discarded region (shaded area): the dotted line corresponds to a sinusoidal model with no subtraction (as generally used), and the continuous line is the corrected sinusoidal model. The difference between the dotted and continuous lines corresponds to the value of β in the two considered cycles. Although this effect is generally small in the cycles selected by the DAMA collaboration, we suggest to account for it in the fitting procedure and use the entire dataset without being forced to drop any data-taking cycle, even if completely asymmetrical. It would be sufficient to compute the integral in Eq. (4.30) and include it in Eq. (4.29) potentially as a function of A, T, and t_0.

Finally, we stress that, if the background is not constant as a function of time, it will have an impact on the value of the average of the rate in the relative data-taking cycle. In particular, the linear term of expansion of the background as a function of time will produce a sawtooth-shape in the residual rate.

4.4.3 Models considered in the analysis

We decided to consider, in our analysis, three different models:

- COS model: this is the pure-cosine model, where the only free parameter is the amplitude A of the modulation:

$$S_{COS}(t) = A \cos\left(\frac{2\pi}{T}\left(t - t_0\right)\right), \tag{4.31}$$

  where the period T is assumed to be fixed to 1 year and $t_0 = 152.5\,d$ such that the peak of the modulation is on the 2nd June.

- SAW model: this is the pure-sawtooth model, where the only free parameter is the slope B of the linearly varying sawtooth:

$$S_{SAW}(t) = B\left(t - t_i\right) \quad \text{with} \quad t_i - \frac{\Delta_i}{2} < t < t_i + \frac{\Delta_i}{2}, \tag{4.32}$$

  where Δ_i is the width of the time window to which t belongs, and t_i is the center of this time window.

- C+S model: this is the cosine plus sawtooth model, where this time the free parameters to be considered are two, A and B:

$$S_{C+S}(t) = S_{COS}(t) + S_{SAW}(t), \tag{4.33}$$

  where S_{COS} and S_{SAW} are defined in the Eq. (4.31) and (4.32), respectively.

What we would like to stress is that the models have been chosen in such a way that the COS and the SAW model have the same number of parameter, in order to minimize the impact of the Ockham's factor (and thus of the priors of the parameters) in the final Bayes factor. Indeed, we expect that, if the C+S model doesn't give a strong improvement in the likelihoods with respect to the other two models, the Ockham's factor will drive the Bayes factor in favor of the simpler models.

4.4.4 Implementation of the models using `JAGS` and `rjags`

For our study we used the general analysis framework `R` [114] and the MCMC algorithm called *Gibbs Sampler* as implemented in `JAGS` [113] and interfaced with `R` in the package `rjags` [115]. Our code is publicly available on `github` [127].

We assumed, as it is done by the DAMA collaboration [122], that the measurements at different times are independent and follow a normal likelihood in each time bin t_i with known standard deviation σ_i given by the experimental uncertainty. The total likelihood is then given by:

$$\mathcal{L}(\{\mu_i\}, \{\sigma_i\}; \{D_i\}) = \prod_{i=i}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - \mu_i)^2}{2\sigma_i^2}\right], \qquad (4.34)$$

where the true parameters $\{\mu_i\}$ are given as a function of $\{t_i\}$ by implementing one of the three models outlined above, for example for the COS model:

$$\mu_i = A \cos\left(\frac{2\pi}{T}(t_i - t_0)\right) \qquad (4.35)$$

with the only unknown parameter represented by the amplitude A. For each unknown parameter in the model, we have to give a prior distribution probability. In the following, we used flat priors for all parameters. However, these priors could be different in principle, and in Ref. [4] we checked, by changing both the width of the priors and their functional form, that the conclusions of our analysis wouldn't change with other reasonable choices.

The Monte Carlo simulation gives the unnormalized posterior probability of the parameters of interest sampled using the Gibbs algorithm. The results reported in this work are obtained with a single Markov chain with $5 \cdot 10^5$ steps. Finally, to estimate the marginal likelihood and the Bayes factor, we used the `bridgesampling` package [128]. This package uses the same Markov chain used to sample the posterior probability for the integration of the marginal likelihood.

4.4.5 Results

As already said, we decided to analyze the residual rate in the (2-6) keVee energy window. Since the precision of the measurements and the size of the modulation are different in the various phases, as a first step we analyzed the data individually on each of the three datasets, testing only the two hypotheses of pure-cosine (COS) and pure-sawtooth (SAW) contribution. We tested on the most informative dataset also the third model (C+S), and then we performed the analysis on the whole dataset. Finally, we evaluated the performance of the cosine model, promoting the period and the phase to free parameters.

DAMA phase	$BF_{C,S}$ [dB]	$LR_{C,S}$ [dB]	$OF_{C,S}$ [dB]	ΔBIC	ΔAIC
DAMA/NaI	-16.7	-18.1	1.4	8.34	8.34
LIBRA Phase I	14.0	12.0	2.0	-5.53	-5.53
LIBRA Phase II	86.5	84.7	1.8	-39.0	-39.0
All	88.8	64.7	24.1	-39.7	-33.8

Table 4.1. Bayes factor (BF), likelihood ratio (LR), Ockham's factor (OF), Bayesian Information Criterion (BIC) difference and Akaike Information Criterion (AIC) difference between the COS and the SAW model for the three stages of DAMA experiments. For the first three rows, the ΔBIC and the ΔAIC values are equal because the two models have the same number of parameters. The last row refers to the comparison between COS and a SAW model with 3 different slopes.

COS versus SAW models

In table 4.1 we show the results of the comparison between the COS and the SAW models in terms of Bayes factor, likelihood ratio, Ockham's factor, Bayesian Information Criterion difference and Akaike Information Criterion difference in every experimental stage. Since some of these quantities can eventually be very large or very small, BF, LR and OF are given in decibels[2]. This table also shows how the contribution of the Ockham's factor, which is the component of the Bayes factor critically dependent on the choice of the priors of the parameters, is marginal. For each of the three datasets, we chose a flat prior on B as well as on A. However, for all the three datasets we tested different possible priors (uniform, normal and gamma distributions), but in all cases the Ockham's factor contribution is always under control, namely it changes of at the most 4-5 units in decibels with respect to our final choice of the prior. The results for the different stages can be summarized as follows:

- For the DAMA/NaI dataset, the Bayes factor, which is $10^{-1.67}$ in normal units, disfavors the pure-cosine model in favor of the pure-sawtooth one. However, this is the less informative dataset, namely it is the one with relative uncertainties on each point larger than in the other phases. The fit is shown in Fig. 4.3.

- For the DAMA/LIBRA Phase I dataset, the Bayes factor is of the order of $10^{1.40}$, and this means that already at this stage, where the relative uncertainties are smaller than in the previous case, the cosine model starts to win on the pure-sawtooth model. The fit is shown in Fig. 4.4.

- For the DAMA/LIBRA Phase II dataset, the Bayes factor is $10^{8.65}$. This means that in the region where the data give the maximum of the information, in the sense that we explained before, the pure-cosine model is greatly preferred to the pure-sawtooth model. In Fig. 4.5 the result of the fit is presented.

[2]Given a quantity x, the decibel d is defined as $d = 10\log_{10}(x)$.

The details of the three fits are described in table 4.2, in which all the reported values are obtained by computing the expected value and the standard deviation of the posterior probability of each parameter.

Fit of the DAMA/NaI residual rate in the 2-6 keVee energy window

Figure 4.3. Fit to the DAMA/NaI residual rate in the (2-6) keVee energy window. The regions in which the blue line is constant (and equal to zero) are the regions between two non-contiguous data-taking cycles.

Fit of the DAMA/LIBRA Phase I residual rate in the 2-6 keVee energy window
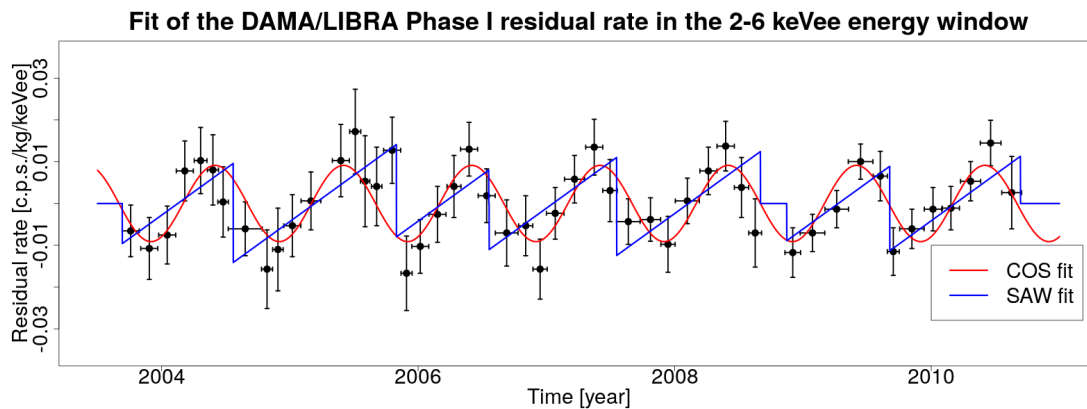
Figure 4.4. Fit to the DAMA/LIBRA Phase I residual rate in the (2-6) keVee energy window. The regions in which the blue line is constant (and equal to zero) are the regions between two non-contiguous data-taking cycles.

Performance of the COS, SAW, and C+S models

Since the third dataset is the most informative one, as already shown in Fig. 4.5, we decided to test not only the COS and the SAW model, but also the C+S model. The results of the fit are given in table 4.3 together with the results of the comparisons between the various models. The value of B obtained in the C+S model is consistent with zero, $B = (-0.0035 \pm 0.0042)$ cpd/kg/keVee/yr, and the value of A is consistent with the value of A obtained in the COS model. Indeed, looking at the Fig. 4.5, the green (C+S fit) and the red (COS fit) lines are very close: in fact the χ^2/dof of the C+S fit is very similar to that of the COS fit, or, in other words, the LR of

Fit of the DAMA/LIBRA Phase II residual rate in the 2-6 keVee energy window
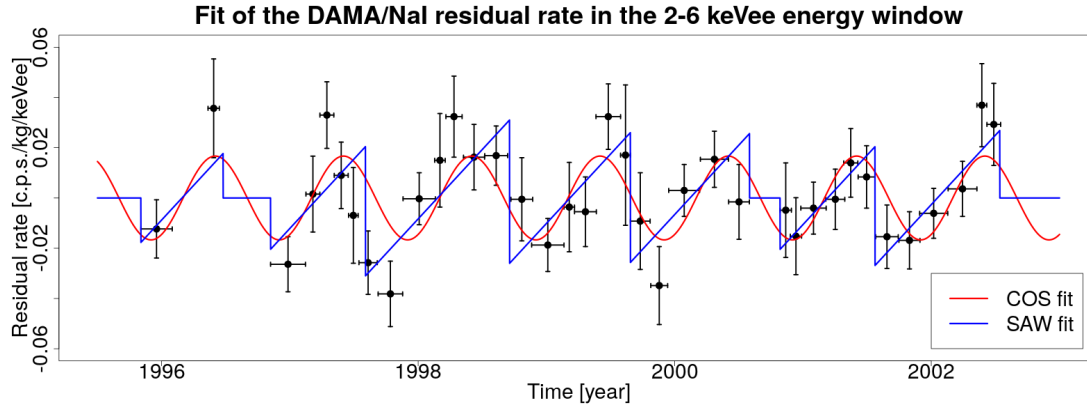

Figure 4.5. Fit to the DAMA/LIBRA Phase II residual rate in the (2-6) keVee energy window. The regions in which the blue line is constant (and equal to zero) are the regions between two non-contiguous data-taking cycles.
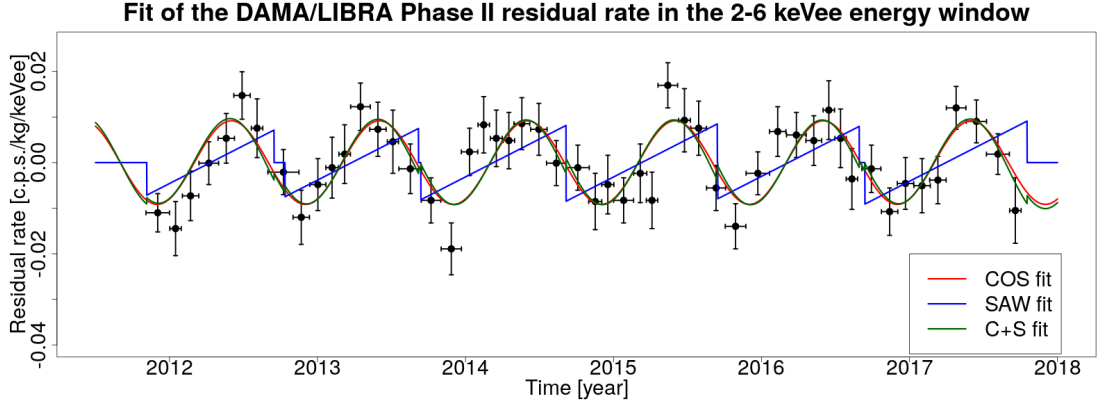
DAMA phase	Fit to COS model		Fit to SAW model	
	A [cpd/kg/keVee]	χ^2_{cos}/dof	B [cpd/kg/keVee/yr]	χ^2_{saw}/dof
DAMA/NaI	0.0168 ± 0.0029	$36.7/36$	0.0552 ± 0.0085	$28.3/36$
LIBRA I	0.0092 ± 0.0013	$29.6/49$	0.0222 ± 0.0032	$35.1/49$
LIBRA II	0.0092 ± 0.0011	$43.3/51$	0.0166 ± 0.0029	$82.3/51$

Table 4.2. Results of the fit for the cosine amplitude A of the COS model and the sawtooth coefficient B of the SAW model obtained on the DAMA residuals in the (2-6) keVee energy window during the different phases, together with the corresponding χ^2/d.o.f.

the models is very close to one (it is $10^{-0.14} = 0.72$ in normal units). Therefore, the Bayes factor in the COS vs C+S case is driven by the Ockham's factor, and since the number of parameters is different in the two models, the COS model wins against the C+S model. The C+S model is an extension of the COS model but, even if it's more complex, it doesn't improve the goodness of fit, and the price to pay for the greater complexity is reflected in the Bayes factor. Since in this case the value of BF depends critically on the priors, we tried to use different possible priors for the parameters as before, but in all cases the BF favored the COS model. On the other hand, for the SAW and C+S models, although the Ockham's factor pushes in the direction of the simpler SAW model, the contribution of the LR drives the BF in favor of the C+S model. The final message of this analysis is that the most informative dataset available to us can be better represented by models that contain a dominant cosine component, as the COS and C+S models, with respect to the SAW model. In addition, the sawtooth component of the C+S model, quantified by the B parameter, is compatible with zero.

Model comparison on the whole dataset

Since adding the sawtooth component to the cosine modulation in the most informative dataset didn't produce any change with respect to the pure-cosine model,

Fit results		A [cpd/kg/keVee]	B [cpd/kg/keVee/yr]		χ^2/dof	
C+S fit		0.0102 ± 0.0016	-0.0035 ± 0.0042		42.7/50	
Model comparison		BF [dB]	LR [dB]	OF [dB]	ΔBIC	ΔAIC
COS vs SAW		86.5	84.7	1.8	-39.0	-39.0
COS vs C+S		11.3	-1.4	18.0	-3.31	-1.36
SAW vs C+S		-69.9	-86.2	16.3	35.7	37.7

Table 4.3. **Top**: Results of the fit of the cosine amplitude A and the sawtooth coefficient B of the C+S model obtained on the DAMA residuals in the (2-6) keVee energy window during the DAMA/LIBRA Phase II, together with the corresponding χ^2/d.o.f. **Bottom**: Comparison between the various models in the DAMA/LIBRA Phase II dataset in terms of Bayes factor (BF), likelihood ratio (LR), Ockham's factor (OF) and difference of Bayesian Information Criterion (ΔBIC) and Akaike Information Criterion (ΔAIC). For all these four metrics (BF, LR, ΔBIC, ΔAIC) the SAW model is largely disfavored.
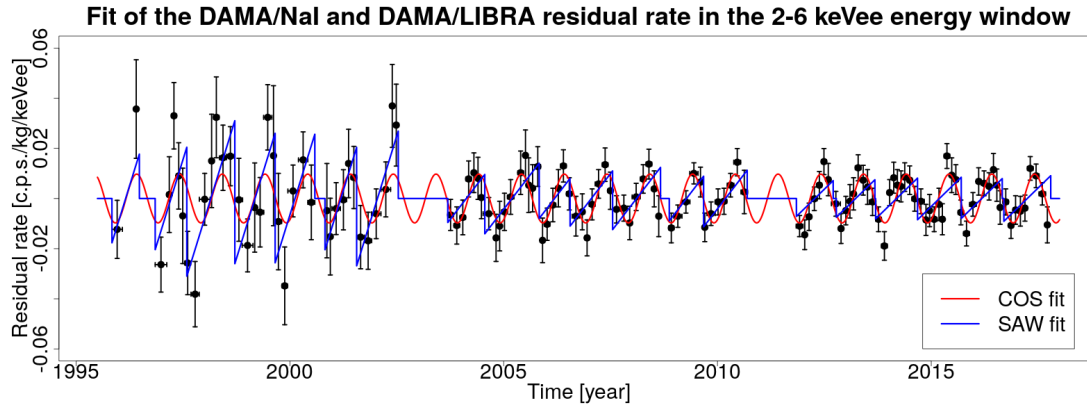
Figure 4.6. Fit to the DAMA residual rate in all the three stages in the (2-6) keVee energy window.

the next step of our analysis was comparing the SAW and COS model on all the available data. In principle in the various phases the size of the sawtooth variation could be different, thus we allowed to have three different B parameters. The results of the global fit is shown in Fig. 4.6 and the details can be summarized as follows:

- For the COS fit:

$$A = (0.00973 \pm 0.00078) \, \text{cpd/kg/keVee} \qquad \chi^2/dof = 116.0/138 \qquad (4.36)$$

- For the SAW fit:

$$\begin{cases} B_{NaI} & = (0.0553 \pm 0.0085) \, \text{cpd/kg/keVee/yr} \\ B_{LIBRAI} & = (0.0222 \pm 0.0032) \, \text{cpd/kg/keVee/yr} \qquad \chi^2/dof = 145.8/136 \\ B_{LIBRAII} & = (0.0166 \pm 0.0028) \, \text{cpd/kg/keVee/yr} \end{cases}$$

$$(4.37)$$

As a cross-check, for the SAW model in each of the phases we obtain for the parameters B the same results of table 4.2. The comparison between COS and SAW

model gives

$$BF = 88.8 \text{ dB},$$
$$LR = 64.7 \text{ dB},$$
$$OF = 24.1 \text{ dB},$$
$$\Delta BIC = -39.7 \,,$$
$$\Delta AIC = -33.8 \,. \tag{4.38}$$

In this case the Ockham's factor is still quite marginal with respect to the LR, and the final Bayes factor, which is $10^{8.88}$ in normal units, pushes again strongly towards the COS model.

Extraction of phase and period of the cosine model

Finally, we decided to fit all the available data with a pure-cosine model in which the phase and the period of the modulation are not fixed, that is to say that t_0 and T defined in Eq. (4.31) are now treated as parameters. In this case, choosing uniform priors for the two new parameters, we obtain:

$$\begin{cases} A & = (0.00981 \pm 0.00079) \text{ cpd/kg/keVee} \\ t_0 & = (0.382 \pm 0.037) \text{ yr} \\ T & = (1.0008 \pm 0.0023) \text{ yr} \end{cases} \qquad \chi^2/dof = 112.8/136 \tag{4.39}$$

These results are compatible with a period of 1 year and a $t_0 = 152.5\,d = 0.418\,yr$. The comparison between cosine and SAW model this time gives

$$BF = 62.4 \text{ dB},$$
$$LR = 71.7 \text{ dB},$$
$$OF = -9.3 \text{ dB},$$
$$\Delta BIC = -33.0 \,,$$
$$\Delta AIC = -33.0 \,. \tag{4.40}$$

Therefore, again, even if now the SAW model enters the game with the same number of parameters of the cosine model (in fact now the OF slightly favors the SAW model), the BF, which is $10^{6.24}$ in normal units, still pushes strongly towards the cosine model.

A suggested model to interpret the DAMA residuals

All our studies indicate that the time modulation of the DAMA residual rate in the $(2-6)$ keVee energy window cannot be described by an artifact due to the algorithm to subtract a slowly varying background.

Nevertheless, we believe that such an algorithm could potentially have an impact on the extraction of the parameter of interest of the signal. In particular, the definition of the time window used to average the rate can introduce the following problems in the residual rate:

	F1	F2
A [cpd/kg/keVee]	0.0084 ± 0.0011	0.0084 ± 0.0012
B1 [cpd/kg/keVee/yr]	0.0371 ± 0.0089	0.0381 ± 0.0090
B2 [cpd/kg/keVee/yr]	0.0078 ± 0.0038	0.0080 ± 0.0038
B3 [cpd/kg/keVee/yr]	-0.0006 ± 0.0035	-0.0003 ± 0.0038

Table 4.4. Results of the F1 and F2 fit in terms of bet-fit values of the parameters.

- a non-zero contribution due to the possible presence of a signal, as discussed in Sec. 4.4.2;

- a sawtooth time modulation due to the presence of a slowly varying background; such a modulation can enhance or reduce the amplitude of a sinusoidal signal, as well as affect its phase.

For this reason, we suggest using a model that takes these effects into account.

Let's consider a data-taking cycle that starts at t^* and extends for a period of time Δ, the true value μ_i of the observed rate in the time bin t_i is:

$$\mu_i = A \cos\left(\frac{2\pi}{T}(t_i - t_0)\right) - \frac{A}{\Delta} \int_{t^*}^{t^*+\Delta} \cos\left(\frac{2\pi}{T}(t' - t_0)\right) dt' + B\left(t_i - \frac{\Delta}{2}\right) \quad (4.41)$$

where A is the amplitude, T the period, and $(2\pi\, t_0/T)$ the phase of the sinusoidal signal, while B is the slope of a linearly varying background. For a different time dependence of the background, a linear model is however the first order contribution in a time power series.

For the sake of completeness, we deployed such a model to fit the three phases of the experiment in the (2-6) keVee energy window. Like we did in Section 4.4.5, since in the various phases the size of the sawtooth variation could be different, we allowed to have three different B parameters. We performed a first fit (F1) keeping both T and t_0 fixed to the values expected for a DM signal, as well as a second one (F2) allowing them to vary. To better understand the differences with respect to the COS model fit, we decided to show in the Fig. 4.7 and Fig. 4.8 the marginal posteriors *pdf*s of all the fit parameters for F1 and F2, respectively.

As shown in table 4.4, except for $B1$ which is positive, the other two are compatible with zero in both cases. The value of A extracted in this way is slightly smaller with respect to the previous cases because of the expected anti-correlation with the Bi parameters. The small correlation between the Bi parameters, assumed to be *a-priori* independent, is an induced effect of their common anti-correlation with the A parameter. For the F2 fit, the values of t_0 and T are consistent with those expected for a DM signal. Finally, we remark that even though we included into the model the correction for a signal "double-counting" as described in Section 4.4.2, this effect is not relevant with the current choice of time intervals for the experimental cycles.

4.4.6 Conclusions

In Ref. [4] we highlighted that, with an exception for the DAMA/NaI dataset, which is however the less informative one, in all the comparisons we performed the
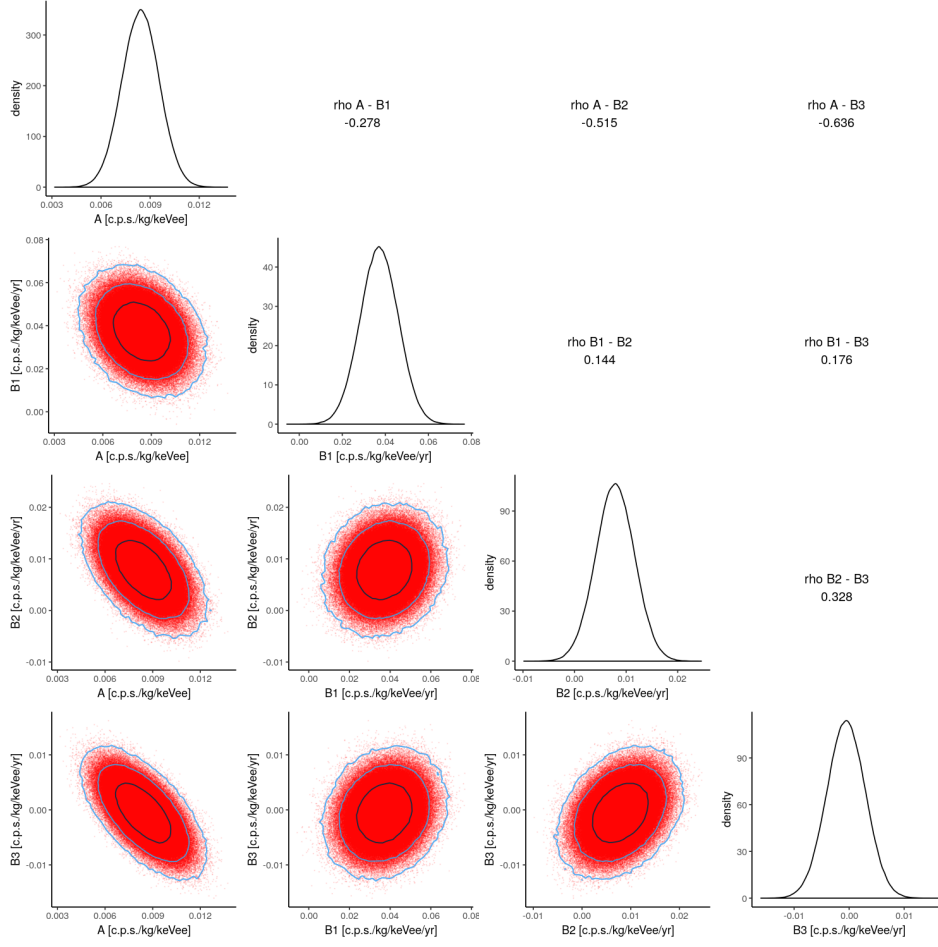
Figure 4.7. Marginal posterior *pdf*s for the free parameters of the proposed model at fixed T and t_0 for the fit of the whole (2-6) keVee DAMA dataset. The 4 plots on the diagonal of the figure are the uni-dimensional *pdf*s of each single free parameter obtained by marginalizing on the all the others. The 10 bi-dimensional *pdf*s in the bottom-left corner of the figure give the marginal *pdf*s of each pair of parameters obtained by marginalizing on the other. The plots show also the credible regions at $0.68, 0.95, 0.997$ probability as black, green, and blue contours respectively. The correlation coefficients are given in the upper-right corner of the figure.

sawtooth hypothesis proposed in Ref. [126] is always disfavored with respect to the cosine modulation hypothesis. This is quantifiable in the Bayes factor, which is in all cases between 10^6 and 10^8. To be more specific, in the case of using the entire dataset and comparing the cosine with just the amplitude as a free parameter in the fit against the sawtooth with three independent and free to vary slope parameters for the three phases of data-taking, we obtain a Bayes factor of $10^{8.88}$ in favor of the cosine model. The effect of the priors choice has been extensively tested and can be reasonably quantified in a contribution at the most of the order 10^2 to the total BF. Finally, we used the full cosine model with free period and phase and still found a BF of order 10^6. For this case, we obtained a value of the period and phase compatible with those reported by the DAMA collaboration.
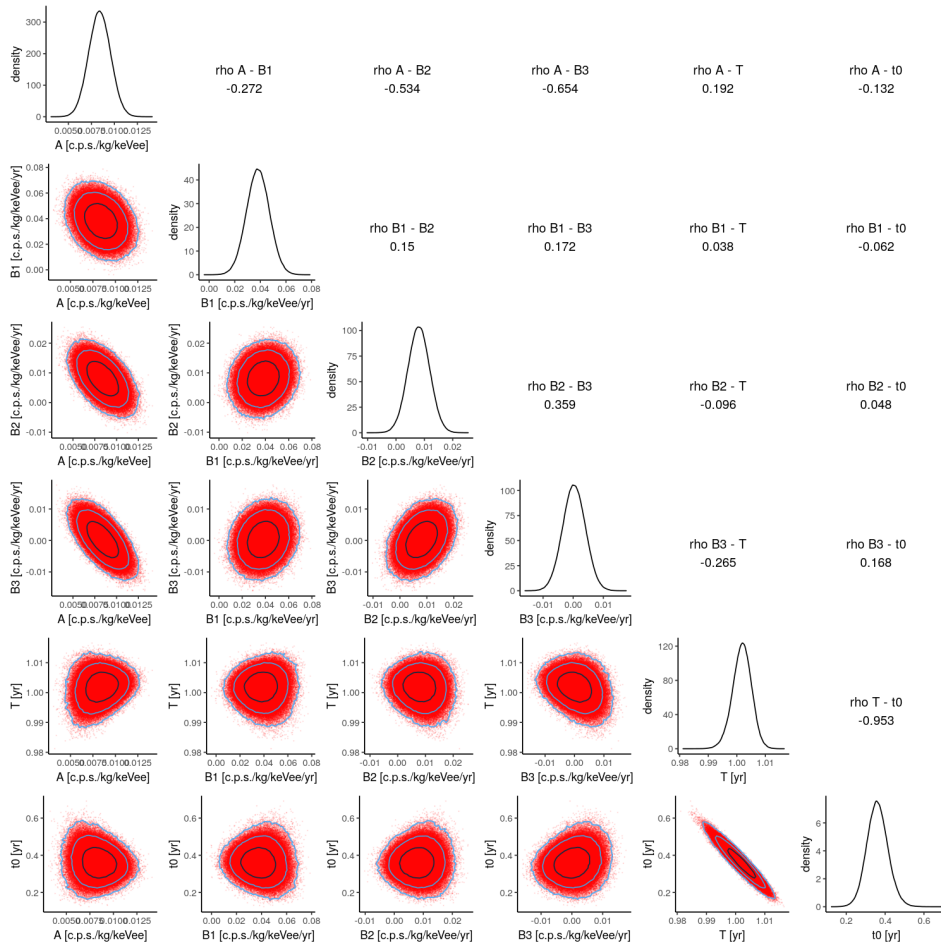
Figure 4.8. Marginal posterior *pdf*s for the free parameters of the proposed model for the fit of the whole (2-6) keVee DAMA dataset. The 6 plots on the diagonal of the figure are the uni-dimensional *pdf*s of each single free parameter obtained by marginalizing on the all the others. The 15 bi-dimensional *pdf*s in the bottom-left corner of the figure give the marginal *pdf*s of each pair of parameters obtained by marginalizing on the other. The plots show also the credible regions at $0.68, 0.95, 0.997$ probability as black, green, and blue contours respectively. The correlation coefficients are given in the upper-right corner of the figure.

Finally, we point out that the background subtraction algorithm used by DAMA can introduce a bias in the parameters of the fitted signal, produced by a residual contribution from the expected signal as well as a contribution from a slowly varying background. Although the time intervals are chosen in a such a way that this bias is small, we showed that these effects can be safely taken into account in the analysis.

Chapter 5

The DarkSide-50 experiment

This chapter will be focused on the description of the DarkSide-50 experiment. The DarkSide-50 detector, which operated underground at Laboratori Nazionali del Gran Sasso (LNGS) until 2019, is a dual phase Ar Time Projection Chamber (TPC) for the direct detection of the DM interaction, see Fig. 5.1. The active target mass of the TPC is (46.4±0.7)kg [130] and the light produced by the interaction events inside this active volume is detected by two arrays of low-radioactivity PhotoMultiplier Tubes (PMTs). The TPC is in turn located inside the Liquid Scintillator Veto (LSV), a stainless steel sphere with a diameter of 4 m which is filled with 30 t boron-loaded organic scintillating liquid and acts as a neutron veto. The sphere is placed inside a 11 m steel cylinder with a diameter 11 m of and a height of 10 m. This cylinder is called Water Cherenkov Veto (WCV) and, filled with 1000 t of ultrapure water, acts as a Cherenkov detector for the cosmic rays.

The main aspects of the interactions of ionizing particles in liquid argon are described in Sec. 5.1. The detailed description of the DarkSide-50 experimental apparatus is given in Sec. 5.2, while the detector response model, the calibration measurements, and the background model are presented in Sec. 5.3.

5.1 Interactions in liquid Argon

Before proceeding to a more detailed description of the experimental apparatus, in this section we will describe the mechanisms through which an incoming particle can transfer energy-momentum inside the LAr.

Let's consider an incoming particle releasing energy inside the TPC. During its interactions with the atoms in the active volume, electric charge and photons can be produced. Once the charge, in the form of ion-electron pairs, is produced, a fraction of it is converted into scintillation light through a process called recombination, see Sec. 5.1.1. In addition to that, a certain amount of energy is lost in the form of heat. The framework is therefore quite complex, and a detailed comprehension of the energy release mechanisms has been needed to characterize the detector response and background models.

A particle interacting in LAr will initially transfer energy either to the electrons or to the nuclei [131, 132]. The total energy E_{tot} deposited inside the LAr can be
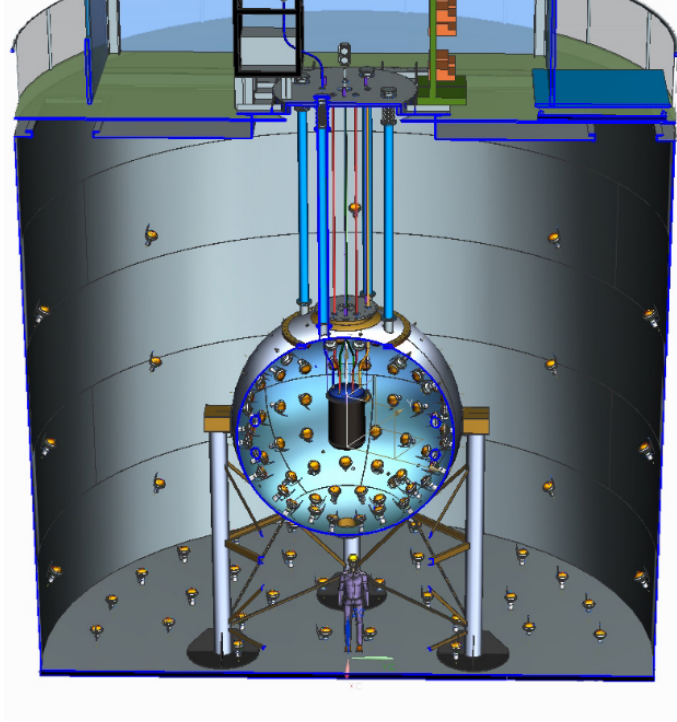
Figure 5.1. Cross-section rendering of the DarkSide-50 detector. Figure from Ref. [129].

written as

$$E_{tot} = \overline{\nu} + \overline{\eta}, \tag{5.1}$$

where $\overline{\nu}$ and $\overline{\eta}$ are the mean energy released in the nuclear channel and the electronic channel, respectively. The possible excitation and ionization induced by secondary nuclear recoils are included into $\overline{\eta}$. In the above equation E_{tot} is related to mean quantities: the amount of energy released inside the detector in the two cases is in fact regulated by a stochastic process, with the constraint, standing also for the mean quantities, that their sum must be equal to E_{tot}. The $\overline{\nu}$ term contains the energy lost in nuclear collisions, as well as dissipated as heat. For what regards the energy $\overline{\eta}$ transferred to the electrons, three different physical channels are usually considered: the contribution needed for the ionization, the energy needed to excite the electrons and the residual kinetic energy of the ionized electrons [133]. Therefore, $\overline{\eta}$ can be expressed as

$$\overline{\eta} = N_i \overline{E}_i + N_{ex} \overline{E}_{ex} + N_i \overline{\epsilon}_{heat}, \tag{5.2}$$

where: N_i is the number of ionized atoms, $\overline{E}_i$ is the mean energy spent for the ionization, N_{ex} is the number of excited atoms, $\overline{E}_{ex}$ is the mean energy of the excited atoms, and $\overline{\epsilon}_{heat}$ is the mean energy that goes into heat as it is lower than the energy needed to at least excite an electron. The ionization energy for an isolated Argon atom is $E_I = 15.75 \, \text{eV}$. In LAr the situation is more complex, as the electronic states show a band structure similar to what happens in solids. The energy of the band gap for LAr is $E_g = 14.2 \, \text{eV}$.

After a nuclear or an electronic recoil, the relative abundance of excited atoms with respect to ionized atoms is quite different, expected to be ~ 0.2 in the ER case

and ~ 1 in the NR case. However, in a liquid material these quantities may look different in data due to the process of recombination, which, in turn, is influenced by the presence of the electric field. The recombination must be therefore qualified and quantified in a proper way in order to be able to discriminate signals produced by different incoming particles.

5.1.1 The recombination process

When an incoming particle transfers sufficient energy to an atom inside the TPC, an electron can be emitted as well as a positive Ar^+ ion is produced. In LAr, in a time of $\mathcal{O}\,(ps)$ after this ionization process, a molecular Ar_2^+ ion is produced. For what regards the electrons, they undergo both elastic and inelastic scatterings, producing additional ionization, excitation and heat. However, a fraction of these electrons can recombine with the positive ions, eventually with the emission of scintillation light, see Sec. 5.1.2: this process is the so-called "recombination".

From a quantitative point of view, in LAr the mean energy transferred for the production of a single electron-ion pair can be assumed to be independent of the type of incoming particle that originated the event. To be more specific, this quantity is defined as

$$\omega = \frac{\overline{\eta}}{N_i}, \tag{5.3}$$

where $\omega = 23.6\,\text{eV}$ in LAr. What is instead dependent on the incoming particle nature is the number of pairs, namely the ionization intensity produced during the primary interaction.

The number of charges produced by the primary interaction is generally different from the total charge measured by the experimental setup. During the drift of the electrons towards the readout system, their number can be reduced mainly due to the recombination process and, to a marginal extent, by being absorbed by impurities. The recombination is a very complex process and depends on many factors: the electric field, the initial distribution of electrons and ions and therefore the nature and the kinematics of the original interacting particle, the mobility of the charges inside the LAr, the distance traveled by the electrons, and the diffusion rate.

For low energy recoils induced in LAr with an energy of the order of few keVs, a commonly accepted theory parameterizing the recombination process is the Thomas-Imel box-model [134]. This model is based on the assumption that the ion-electron pairs can be treated as isolated and that suddenly after the primary interaction the charges are uniformly distributed inside a box of side a. In the presence of an electric field, the fraction of charges surviving the recombination and being collected is given by

$$\frac{Q}{Q_0} = \frac{1}{\xi} \ln\,(1 + \xi), \tag{5.4}$$

where ξ is a free parameter to be determined experimentally during the calibration and can be expressed in terms of the number N_0 of pairs inside the box, α_{EM}, a, the mobility u_- and u_+ of the electrons and the ions, respectively, and the electric

drift field E_d by the equation

$$\xi = \frac{N_0 \alpha_{EM}}{4a^2(u_- + u_+)E_d} \tag{5.5}$$

For higher energy recoils, the recombination in LAr is well described by the "columnar recombination" model by Jaffe [135]. In fact, rather than being enclosed in a volume of size a, energetic particles can produce long ionization tracks. In this case, the distribution of charges can be assumed to be in a column around the particle trace. For a track of radius b, length l and angle θ relative to the electric drift field, the fraction of charges surviving the recombination is given by

$$\frac{Q}{Q_0} = \left[1 + \frac{N_0 \alpha_{EM}}{4\pi l D} e^{x(\theta)} K_0(x(\theta)) \right]^{-1}, \tag{5.6}$$

where: $D = D_- + D_+$ is the diffusion coefficient of the charges, $x(\theta) = [E_d b(u_- + u_+) \sin\theta / 2D]^2$, and K_0 is the modified Bessel function of the second kind. In most of the applications, the Jaffe's model is successfully approximated by the so-called Birks model [136, 137], according to which

$$\frac{Q}{Q_0} = \left(1 + \frac{k \, dE/dx}{E_d} \right)^{-1}, \tag{5.7}$$

where dE/dx is the particle stopping power, and k is a parameter depending on the material properties and to be measured from data.

In the nuclear recoil case, there is an additional physical process to be taken into account, the so-called nuclear quenching. Indeed, as already mentioned, a relevant fraction of the energy of the primary nuclear recoil is lost through secondary nuclear collisions. The traces generated by NR interactions have a cascade structure and are therefore generically very complicated, making it often necessary to use simulations to study their structure and properties. However, a phenomenological approach to quantify the nuclear quenching has been pursued by Bezrukov et al. [138], exploiting the theoretical interplay between the scintillation and the ionization yields in noble liquids. A more detailed and quantitative description of the response to nuclear recoils in LAr will be given in Sec. 5.3, where the full DarkSide-50 response and background models will be shown.

5.1.2 The scintillation process

Both recombination and excitation in LAr contribute to the production of argon bi-atomic excited molecules. As a consequence of their de-excitation, scintillation 128 nm (~ 9.7 eV) photons are emitted. However, not all the induced excitation is converted into scintillation light, due to processes reducing the number of excited argon molecules.

The light produced after the primary excitation of the atoms in LAr is called "excitation luminescence". The whole process, for instance due to an incoming

electron, can be summarized as

$$
\begin{aligned}
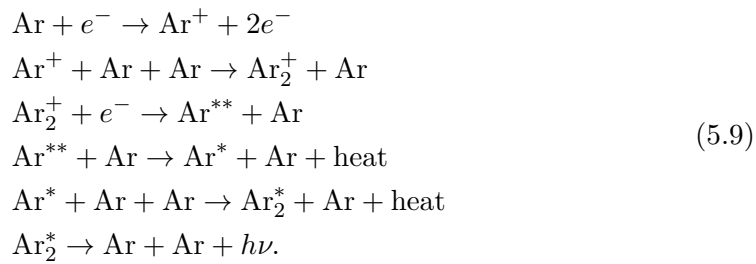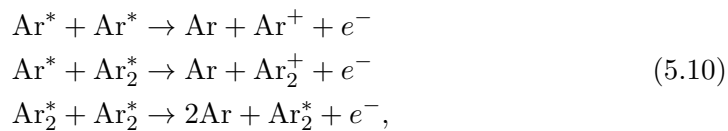\mathrm{Ar} + e^- &\rightarrow \mathrm{Ar}^* + e^- \\
\mathrm{Ar}^* + \mathrm{Ar} &\rightarrow \mathrm{Ar}_2^{*,\nu} \\
\mathrm{Ar}_2^{*,\nu} + \mathrm{Ar} &\rightarrow \mathrm{Ar}_2^* + \mathrm{Ar} \\
\mathrm{Ar}_2^* &\rightarrow \mathrm{Ar} + \mathrm{Ar} + h\nu.
\end{aligned}
\tag{5.8}
$$
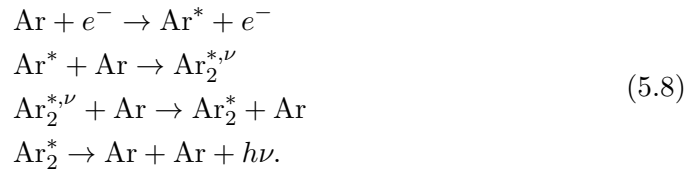
To be more specific, after its production, the excited atom interacts with the other atoms in the liquid forming a strongly bounded dimer $\mathrm{Ar}_2^{*,\nu}$ in a vibrational excited state. The production of the excited dimer is favored with respect to the direct decay of Ar^*: in other words, the $\mathrm{Ar}_2^{*,\nu}$ dimer is produced before Ar^* can decay to its ground state and emit a photon. After the $\mathrm{Ar}_2^{*,\nu}$ formation, further interactions with the Ar atoms produce a Ar_2^* purely electronic excited state. Finally, the excited dimer decays with the emission of a $128\,\mathrm{nm}$ photon and with the dissociation of the molecule in two Ar atoms. This whole process happens in a time of $\mathcal{O}(\mathrm{ps})$.

If the light is instead produced after the excitation induced during the recombination, the scintillation process is called "recombination luminescence" and can be illustrated, for instance in the case of an electron induced ionization, as

$$
\begin{aligned}
\mathrm{Ar} + e^- &\rightarrow \mathrm{Ar}^+ + 2e^- \\
\mathrm{Ar}^+ + \mathrm{Ar} + \mathrm{Ar} &\rightarrow \mathrm{Ar}_2^+ + \mathrm{Ar} \\
\mathrm{Ar}_2^+ + e^- &\rightarrow \mathrm{Ar}^{**} + \mathrm{Ar} \\
\mathrm{Ar}^{**} + \mathrm{Ar} &\rightarrow \mathrm{Ar}^* + \mathrm{Ar} + \mathrm{heat} \\
\mathrm{Ar}^* + \mathrm{Ar} + \mathrm{Ar} &\rightarrow \mathrm{Ar}_2^* + \mathrm{Ar} + \mathrm{heat} \\
\mathrm{Ar}_2^* &\rightarrow \mathrm{Ar} + \mathrm{Ar} + h\nu.
\end{aligned}
\tag{5.9}
$$

After the ionization, the Ar^+ ion can form a Ar_2^+ ion, in the ground state, by capturing another Ar atom. At this point, the recombination of one of the electrons produced by the primary ionization with the Ar_2^+ ion causes a doubly-excited Ar^{**} atom and a neutral ground state Ar atom to be produced. After dissipating some energy in heat, a Ar_2^* excited dimer is formed, and the $128\,\mathrm{nm}$ light is produced in the same way as the excitation luminescence process.

As we just showed, the scintillation light is produced as a consequence of the decay of excited dimers. However, a certain number of them can undergo processes which are different from the one described above, and which can reduce their number, in turn reducing the scintillation light. The light-reducing processes, called generally "electronic quenching", are mainly three and in particular can be written as

$$
\begin{aligned}
\mathrm{Ar}^* + \mathrm{Ar}^* &\rightarrow \mathrm{Ar} + \mathrm{Ar}^+ + e^- \\
\mathrm{Ar}^* + \mathrm{Ar}_2^* &\rightarrow \mathrm{Ar} + \mathrm{Ar}_2^+ + e^- \\
\mathrm{Ar}_2^* + \mathrm{Ar}_2^* &\rightarrow 2\mathrm{Ar} + \mathrm{Ar}_2^+ + e^-,
\end{aligned}
\tag{5.10}
$$

where the first one is called "bi-excitonic quenching", the second one is called "photo-ionization", and the last one is referred to as "Penning process". These processes are more probable where the number density of excited atoms and dimers is higher. Therefore, the electron quenching effect is higher for incoming particles with higher

stopping power. It is therefore useful to express the energy released in the form of scintillation light in units of keVee. We already introduced this quantity in Sec. 2.2.3, but now we can define it in a more precise way: 1 keVee is the energy needed to produce the same amount of scintillation light of a 1 keV electron.

More precisely, there are two excited dimer states that can decay, emitting the 128 nm scintillation light. These states are approximately degenerate, so that they cannot be distinguished by just measuring the energy of the emitted photons. However, the decay time of the two states is quite different. In order to understand the reason for this difference, we need to characterize in more detail these two states. The Ar_2^* excited dimer can be treated, in an approximate way, as being composed by the spin 1/2 excited electron and a spin 1/2 positively charged molecule. In this scheme, there are two allowed quantum states: the spin 0 singlet state and the three spin 1 triplet states. The ground state is a spin 0 state, therefore in the simplistic approximation in which we neglect the interaction between the excited electron and the remaining molecular electron cloud, only the decay of the singlet state is allowed, while the decay of the triplet states is forbidden. However, in reality, the neglected interaction is different from zero, and, as a consequence, the decay from the triplet to the ground state is not forbidden, but it is anyway disfavored with respect to the singlet decay. The result of this scenario is that the singlet-to-ground state decay has a decay time $\tau_s \sim 6$ ns (fast component) which is much smaller than the triplet-to-ground state decay time $\tau_t \sim 1.5\,\mu$s (slow component). Both τ_s and τ_t do not depend on the incoming particle nature, but this is not the case for the relative size of the fast and the slow components, that depends instead on the incoming particle stopping power. In particular, the experimental data show that the fast component grows where the energy loss density is greater, producing in turn a faster signal. This feature is important and, as we will see, is an important handle for the discrimination of ERs from NRs [130, 139].

5.2 Experimental apparatus

5.2.1 DarkSide-50 TPC

A picture of the DarkSide-50 TPC is shown in Fig. 5.2. It is contained in a stainless steel cryostat kept and sustained at the center of the LSV by means of a system of leveling rods [130]. The target material consists of an active mass of (46.4 ± 0.7) kg of liquid Ar (LAr) surrounded on the lateral walls by a 2.54 cm thick Teflon reflector and on the top and on the bottom by fused silica windows. All the surfaces in contact with the active argon mass with the exception of the extraction grid (see Fig. 5.2) are coated with a wavelength shifter, the TetraPhenyl Butadiene (TPB). This is needed to absorb the 128 nm scintillation light produced by LAr and emit 420 nm photons, which can be efficiently reflected, transmitted and detected. The two faces of the silica windows are also coated with 15 nm thick transparent conductive Indium Tin Oxide (ITO) films. The inner faces act as grounded/High Voltage (HV) anode/cathode (top/bottom) for the electron drift system. The outer faces are instead kept at the PMT photocathode potential. The top anode window is equipped with a rim which extends downwards, forming the so-called "diving bell", while the bottom cathode window is flat. The diving bell is needed to hold the 1 cm
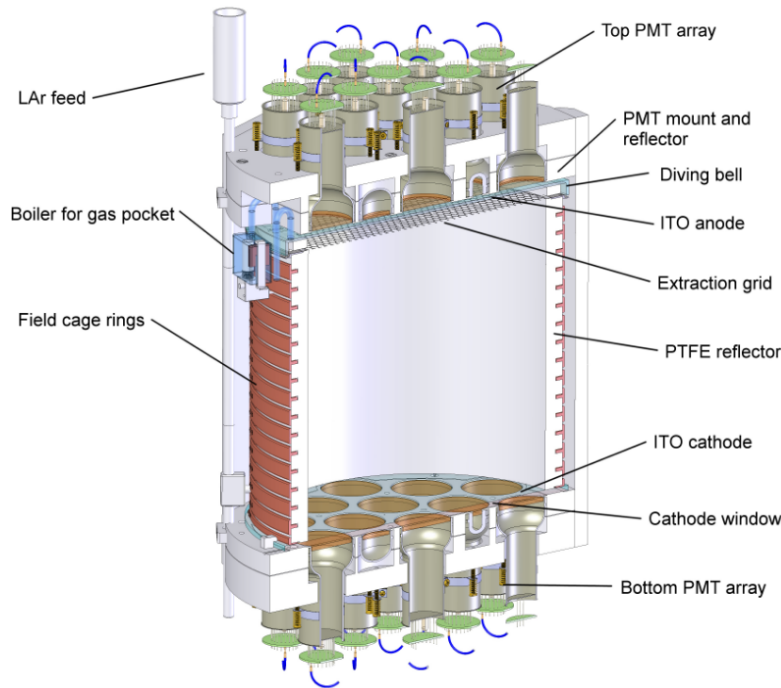
Figure 5.2. Drawing of the DarkSide-50 dual-phase liquid Ar TPC. Figure from Ref. [130].

Ar gas layer, called "gas pocket". The level of the gas pocket is kept at the desired height thanks to a bubbler that, according to the needs, extracts the gas from boiling argon produced externally inside the cryostat. The TPC is instrumented by a total of 38 Hamamatsu R11065 3" PMTs, located half (19) at the bottom and half (19) at the top of the TPC. These PMTs are characterized by low-background and high quantum efficiency, which, at room temperature and at 420 nm, is on average 34%. The PMTs, submerged in turn in LAr, are able to see the LAr active volume through the fused silica windows. Once a photon hits a PMT, an amplified charge signal is produced, collected and digitized by a dedicated electronic system. The total charge collected by the electronic system is proportional to the number of photons that originally hit the PMT. Therefore, to get this number of photons, the total charge is divided by the charge of a Single PhotoElectron (SPE). A fraction of the original photons are anyway lost due to the 85% PMT collection efficiency reported by Hamamatsu. In addition, due to the physical processes happening in the TPC after an ER or NR event, a certain number of residual photons, ranging from 1 to a few PhotoElectrons (PEs), can reach the PMTs with a certain delay with respect to the main event. These signals, called "afterpulses", are generally characterized by a time delay $t_{delay} \in [10\,\text{ns}, 20\,\mu\text{s}]$.

The electric field inside the TPC is controlled by the electron drift system. This system consists of: the ITO coating cathode/anode planes, a field cage and the extraction grid. The electric field is configured such that the electrons produced inside the TPC drift upwards to the LAr surface, are extracted in the gaseous Ar and are accelerated to produce a secondary scintillation signal. The extraction grid,

placed 5 mm below the liquid-gas separation surface, is a 50 μm thick stainless steel hexagonal mesh, with an optical transparency of 95% in the normal direction. The drift field is produced by applying an opportune voltage between the cathode and the grid and is kept uniform in the active volume of the detector by means of copper field cage rings placed outside the cylindrical TPC walls and set at graded potentials. In addition to that, another independent potential is set between the grid and the anode in order to generate the electric field that extracts the electrons from LAr and accelerates them in the gas phase. During the data-taking periods, the drift field was set to 200 $\mathrm{Vcm^{-1}}$, while the electric field at the extraction point and in the gas phase were 2.8 $\mathrm{kVcm^{-1}}$ and 4.2 $\mathrm{kVcm^{-1}}$, respectively [130]. With this configuration, the maximum electronic drift time is 376 μs and the drift speed has been measured to be (0.93 ± 0.01) mm $\mu\mathrm{s^{-1}}$.

5.2.2 The Liquid Scintillator Veto

The TPC is placed at the center of the Liquid Scintillator Veto (LSV), a 4m diameter stainless steel sphere filled with a boron-loaded liquid scintillator with high neutron stopping power. It is indeed very difficult for the TPC to distinguish nuclear recoils induced by neutrons from those induced by WIMP-like particles, since the physics of the interaction with the nuclei is approximately the same in the two cases. A neutron veto system is therefore mandatory to exclude the NR background, as well as to reduce the γ-ray ER background.

The internal surface of the sphere is equipped with 110 R5912 8" PMTs by Hamamatsu, and, to improve the photon collection efficiency, it is coated with a Lumirror reflecting film. The liquid scintillator is a mixture of three liquids:

- 1,2,4-Trimethilbenzene[1] ($\mathrm{C_6H_3(CH_3)_3}$), or PseudoCumene (PC): it is an organic liquid used as the primary scintillator. The scintillation is produced during the radiative decay of the singlet and triplet excitations induced by the energy release of an incoming particle. With a mechanism similar to the LAr, the de-excitation of the singlet state is fast (few ns), while the triplet state decays in a slower time (μs $-$ ms) [140]. The scintillation light emitted by pure PC has a wavelength of ~ 290 nm, but due to the overlap of the emission and absorption spectra of the PC at this wavelength the light collected by the PMTs on the 4 m spherical surface could be reduced, unless a wavelength shifter is added to the liquid mixture.

- 2,5-Diphenyloxazole ($\mathrm{C_{15}H_{11}NO}$), or PPO: it is the wavelength shifter. The energy deposited in the PC is indeed thermally, non-radiatively, absorbed by the PPO, which re-emits it in the form of ~ 360 nm photons, that are not absorbed by the PC. The PPO re-emission time is short enough ($\sim 10^{-8}$ s) to produce a faster light signal with respect to the pure PC, useful to have tighter prompt coincidence cuts [129]. In addition, since the energy transfer from the PC to the PPO is a very highly efficient process, only a small amount of PPO is needed to guarantee a good scintillation efficiency and to improve the absorption length of the scintillation light in the LSV.

---

[1]Here and in the following, the numbers before the name of the liquid specify the chemical structure of the compound.

- TriMethylBorate $(B(OCH_3)_3)$, or TMB: it is an organic liquid containing boron. In particular, the ^{10}B isotope (19.9% natural abundance) has a very high thermal neutron capture cross-section of 3.84×10^3 b. As a consequence of a neutron capture process on ^{10}B, three kinds of products are generates: γ-rays, α particles, and ^{7}Li. The γ-rays can deposit energy in the LSV, inducing scintillation light detectable by the PMTs. The other reaction products have high stopping power and are completely absorbed inside the LSV, but due to the high electronic quenching, a scintillation light of only $50 - 60$ keVee [129] is produced. In addition to ^{10}B, a thermal neutron can also be captured by ^{1}H and ^{12}C, with a smaller capture cross-sections if compared to the boron case.

The composition of the mixture used for the WIMP search phase was 95% of the mass fraction of PC, 5% of TMB, and 1.4 g/L of PPO. With this mixture the total thermal neutron capture time is $\sim 22\mu$s, the number of thermal neutrons capturing on ^{1} H is $\sim 8\%$ of the total, and the number of ^{12}C captures is two order of magnitudes smaller than the captures on ^{1} H. The neutrons entering the LSV transfer energy mostly to the ^{1}H and ^{12}C, undergoing thermalization in ~ 100 ns and inducing an initial fast thermalization signal. The subsequent neutron capture is a slower process, producing a second delayed signal of $\mathcal{O}(1 - 100 \ \mu$s). The delayed capture signal does not depend on the initial energy of the incoming neutron, oppositely to the initial thermalization signal: it is therefore possible to have a high neutron detection efficiency with a relatively low threshold.

5.2.3   The Water Cherenkov Veto

The LSV is placed inside the Water Cherenkov Veto (WCV), a 11 m steel cylinder with a diameter 11 m of and a height of 10 m filled with ~ 1000 t of ultra-purified water. Its internal surface is equipped with 80 ETL 9351 8" PMTs: 24 of them are located on the floor of the WCV facing upwards, while the other 56 PMTs are placed on the lateral walls of the cylinder, facing the center of the cylinder. In addition, both the internal and the external surfaces of the cylinder are coated with a Tyvek reflector, which has very high reflectivity in the desired wavelength window.

The goal of the WCV is two-fold. First of all it allows the vetoing of cosmogenic neutrons that come along with the cosmic muons by detecting the Cherenkov light produced inside the ultra-purified water by the muons. The muons can be in fact in coincidence with the cosmogenic neutrons, which in turn have a great penetrating power. The experiment has been located underground, therefore the number of cosmic muons and cosmogenic neutrons is reduced. However, additional neutron background arises from natural radioactivity, since the rocks of the cave contain radioactive U and Th isotopes. The WCV therefore acts also as a neutron shield for the LSV and the TPC.

5.2.4   DarkSide-50 dual-phase TPC working principle

The energy-momentum released in LAr by an incoming particle can be revealed by measuring the scintillation light or the charges produced via ionization. In Sec. 5.1 we explained how the response of LAr in terms of scintillation and ionization is

generally different. The dual-phase TPC technology allows us to measure both these observables at the same time, increasing the energy resolution and giving us additional handles to discriminate between signal events and background events. In fact, a WIMP-like particle is expected to induce NR, while the dominant background consists mostly of β and γ-rays, which instead induce ER.

The first signal produced by ionizing events is the scintillation signal, and it is usually called "S1". In DarkSide-50, the scintillation light is readout by the 19 PMTs placed at the bottom of the TPC. The S1 signal is proportional to the number of photons N_γ produced in LAr, and can be therefore expressed as [141]

$$S1 = g_1 N_\gamma = g_1 \left(N_{ex} + f N_i \right), \tag{5.11}$$

where N_{ex} is the number of excited atoms, N_i is the number of ion-electron pairs, f is the fraction of charges that undergo recombination, and g_1 is the so-called S1 yield.

The fraction of electrons that survives recombination drift upwards to the extraction grid thanks to the electric drift field. The higher electric field between the grid and the ITO anode is able to extract the electrons from the liquid to the gas phase, with an efficiency $> 99\%$ for the DarkSide-50 operational extraction electric field $(2.8\,\mathrm{kVcm^{-1}})$. In the gas phase the so-called "electroluminescence" process takes place: thanks to the local electric field, the electrons are further accelerated, generating scintillation light and additional ionization, resulting in an avalanche. The S2 signal is then proportional to the number of electron surviving recombination, then [141]

$$S2 = g_2(1 - f)N_i, \tag{5.12}$$

where, again, f is the fraction of charges that undergo recombination, N_i is the number of ion-electron pairs and g_2 is the S2 yield. The physical processes behind the production of light in gaseous Ar are very close to what we explained for LAr in Sec. 5.1. Even if the energy levels in gaseous Ar are slightly different from the LAr case, the scintillation light has approximately the same 128 nm wavelength. This secondary scintillation signal, usually called "S2", is collected by the 19 PMTs on the top of the TPC.

The dual-phase TPC gives access to the 3D position information. The x-y position - namely the position on the horizontal plane - can be determined from the distribution of the S2 signal on the top PMTs. The z position is instead derived from the time difference occurring between the S1 and the S2 signal, since this time delay depends on the z distance between the primary event and the extraction grid, and on the (measured) speed of the electrons inside the liquid phase.

The discrimination between the ER and the NR is achieved by studying not only the ratio between the S1 and the S2, but also by exploiting their temporal shape. In fact, we already explained how the relative contributions of the fast (singlet) and slow (triplet) components in the scintillation process depends on the nature of the primary interaction. In particular, the NR scintillation signal has a higher fast component and smaller slow component with respect to the ER case. The DarkSide-50 collaboration decided to exploit this feature by defining the so-called f_{90} parameter, namely the fraction of the light detected in the first 90 ns of the
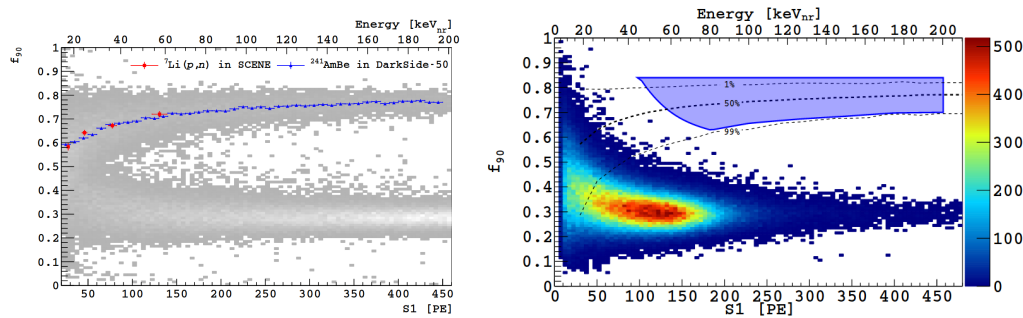
Figure 5.3. Left: f_{90} vs S1 distribution in DarkSide-50 during the AmBe calibration (gray dots). The blue dots correspond to the upper band f_{90} media, while the red dots are the measurements from the SCENE experiment [142, 143] to the DarkSide-50 size. Two horizontal bands in the gray points corresponding to the AmBe neutrons induced NR (top band) and the β/γ background (bottom band) are clearly visible. Figure from [60]. **Right**: f_{90} vs S1 distribution in DarkSide-50 after all quality and physics cuts. The DM search region is represented by the shaded blue area. The dashed lines correspond to the 1%, 50% and 99% NR acceptance contour as a result from a fit on the AmBe data. Figure from Ref. [139].

signal. It is given by

$$f_{90} = \frac{\int_0^{90\,\text{ns}} S1(t)dt}{\int_0^T S1(t)dt},\tag{5.13}$$

where T is the total time of the pulse. For NR events, f_{90} is equal to ~ 0.7, while it is equal to ~ 0.3 for ER events, as it is shown in the left plot of Fig. 5.3, where the data from a calibration run with an AmBe neutron source are showed. This information, combined with all the other observables, is used to define a region of interest for DM searches in the f_{90} vs $S1$ parameter space, as it is shown in the right plot of Fig. 5.3, which is obtained applying the analysis cuts for the background rejection [130]. During the 532 d DarkSide-50 DM search phase described in Ref. [139] no event has been observed in the desired region. Assuming the standard DM hypotheses of the SHM, the results are consistent with at most 2.3 DM induced events (90% C.L.). The exclusion limits in terms of DM-nucleon spin independent cross-section are depicted in Fig. 5.4.

5.3 Detector response, calibration and background model

My work for the DarkSide-50 collaboration was primarily focused on the so-called "low-mass" DM searches. With the "low-mass" region we indicate the sub-GeV DM mass region, which corresponds to events with an energy of the order of the keV. In this section we will describe the detector response model, the calibration measurements and the background model for the ionization signals. They are crucial for the analyses carried out in DarkSide-50 and in particular for the low-mass DM search campaign, since the observable used for this purpose is the energy of the ionization signals. This energy is measured in terms of N_{e^-}, namely the number of extracted ionization electrons. In order to reach the low threshold needed for DM searches in the low-mass region we need to consider also the S2-only events, since, in
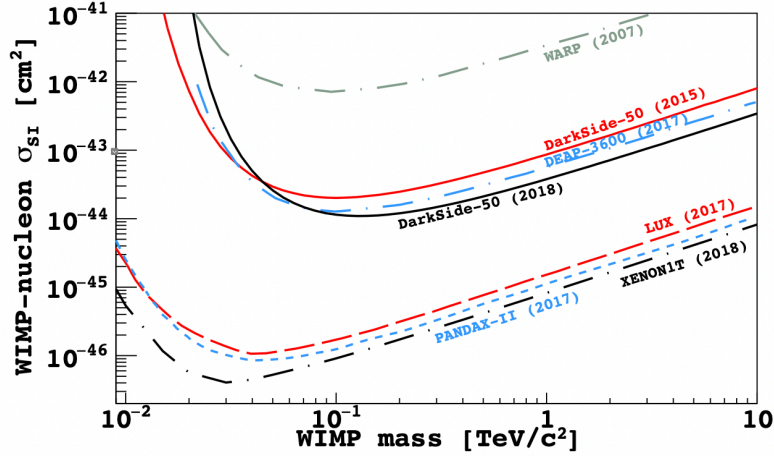
Figure 5.4. 90% C.L. exclusion limits from DarkSide-50 on DM-nucleon spin indepen-
dent cross-section (black line). The other lines correspond to the results from other
experiments using Ar or Xe. Figure from Ref. [139].

the energy region of interest, the S1 signals are not large enough to be detected [1].
Fig. 5.5 shows how the S1 signals are not visible in the low energy region. If we
use also the S2-only events we do not have a "zero event" observation, since the ER
β/γ internal background (mainly ^{39}Ar and ^{85}Kr) and external background (induced
by the radioactive isotopes inside the different components of the PMTs and the
cryostat) produce low energy events that, without the S1 information, cannot be
in principle removed. It is therefore crucial to develop an accurate background and
detector response model and the related systematic effects, since they in turn have
an impact on the sensitivity to low mass DM candidates.

5.3.1 Detector response

In order to be able to understand how the incoming events are readout by the
experiment, all the relevant instrumental effects must be taken into account. Since,
as we mentioned, we are interested in the ionization signals, let's focus on S2. The
measured value of the mean S2 yield g_2, as indirectly defined in Eq. 5.12, for events
located under the central PMT is 23 ± 1 pe/e$^-$, with a resolution of $\sim 27\%$ [144].

An additional efficiency effect must be considered due to the electron lifetime,
namely the time in which an electron could travel during the drift before being
absorbed by the electronegative impurities of LAr. This time has been measured
to be ~ 10 ms, more than an order of magnitude greater than the maximum drift
time in the TPC, which is $376\,\mu$s. Therefore, the reduction of the S2 yield due to
the electron lifetime is small, of the order of few percent.

Data show a dependence of the ionization response on the radial position of the
events, in the x-y plane parallel to the extraction grid. This is the dominant effect
in the S2 response, giving up to a factor of 4 of difference between the S2 yield at
the center of the TPC and the yield near the lateral sides of the TPC. The reasons
could be mainly two: the TPB coating of the fused silica windows could not be
sufficiently uniform, inducing variations in the conversion of the 128 nm scintillation
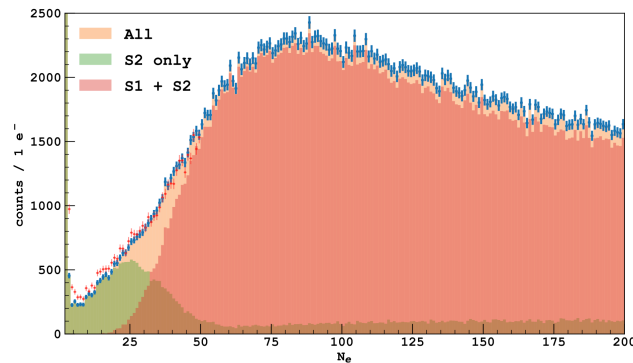
Figure 5.5. Data spectrum of DarkSide-50 as a function of N_{e^-}, the number of extracted electrons. The red dots are obtained with the data selection used for the 2018 analysis [1], while the blue dots over the orange shaded area are obtained with the latest data selection - see Tab 6.2 for more details. The green shaded area corresponds to the S2-only events with the new selection, while the red shaded area corresponds to the S1+S2 events with the new selection.

light; the sagging of the fused silica windows, that could induce variations in the thickness of the gas pocket and, therefore, in the amplification of the gas phase. In order to characterize in a quantitative way this effect throughout the TPC, a ^{83m}Kr gas source has been added to the active mass for calibration. A radial dependent efficiency has been therefore measured using the ^{83m}Kr 41.5 keV line and applying a position reconstruction algorithm [141] which is able to determine the position of the events with a resolution < 1 cm. Nevertheless, the low-mass analysis reaches very low energies, down to ~ 100 eV, and at those energies the reconstruction algorithm is not sufficiently efficient. To overcome this issue, an additional channel-based correction has been applied to the S2 signals [1, 144].

In order to validate this approach, all these efficiency and resolution effects have been implemented in a Monte Carlo and applied to the ^{37}Ar data. Indeed, a small quantity of ^{37}Ar radioactive isotope is naturally present in LAr. Its half life is $35.01(2)$ d and it decays mainly in two ways: single electron capture from the K shell, releasing 2.83 keV, and single electron capture from the L1 shell, releasing 277 eV. The correspondent branching ratios are $\sim 90.4\%$ and $\sim 8.4\%$, respectively, and has been computed by means of the `BetaShape` code [145]. In both cases, the energy is released in the form of Auger electrons, UV photons, and X-rays. In order to simulate in a detailed way the ^{37}Ar events, including the emitted cascades of secondaries, the `RELAX` software [146] and the `EADL2017` library of atomic transition data [147] have been used. A simulated spectrum of the ^{37}Ar decays in DarkSide-50 has been therefore obtained by generating ^{37}Ar events uniformly inside the TPC and by applying the detector response.

In order to isolate the two lines from ^{37}Ar, the data from the latest ~ 500 d of the DM search campaign, where all the ^{37}Ar is approximately decayed, are properly normalized and subtracted to the first ~ 100 d data. At this point, the resulting spectrum is fitted with the Monte Carlo simulated spectrum. The average number of detected electrons of the two ^{37}Ar lines are treated as two independent parameters. The intrinsic fluctuations in the number of particles produced subsequently to an
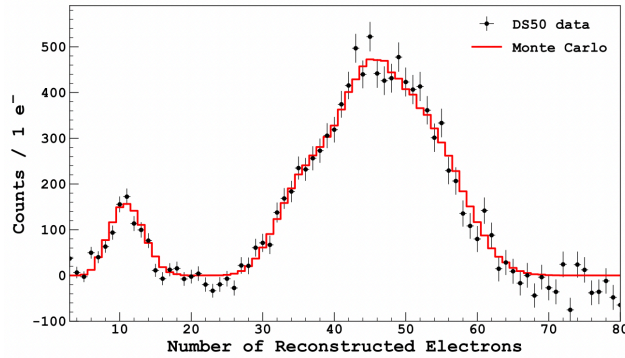
Figure 5.6. Best fit of ^{37}Ar data spectrum (black dots) with the simulated ^{37}Ar spectrum (red line). The spectra are expressed in terms of the number of reconstructed electrons. Figure from Ref. [144].

ionizing ER event is parameterized by a Fano factor [148]. Therefore, adding the Fano factor, the fit has 3 parameters. Fig. 5.6 shows the best χ^2 fit on the ^{37}Ar data, which, in numerical terms, gives: 0.10 ± 0.03 for the Fano factor, $12.0 \pm 0.1(stat.) \pm 0.5(syst.)$ electrons for the L1 shell line, and $48.2 \pm 0.2(stat.) \pm 2.1(syst.)$ for the K shell line, for a total χ^2 of 82.4 over 64 degrees of freedom [144]. The relative amplitude between the two lines (0.10 ± 0.01) is consistent both with the `BetaShape` calculation (~ 0.093) and the other existent measurements (e.g. 0.103 ± 0.003 in Ref. [149]). The Fano factor is in agreement with the predictions from both the Shockley and the Alkhazov models (0.107 and 0.116, respectively [150]) for the fluctuation of ionization yield in LAr.

In order to determine the number of original ionization electrons of the two lines, we need to subtract the number of electrons produced directly by the K and L1 shells decays from the number of extracted electrons. For the L1 shell, the computation of the number of emitted primaries is straightforward [144]. Taking into account both statistical, systematic and theoretical uncertainties, the resulting number of original ionization electrons, which for the L1 shell have an energy of 179 eV, is 8.2 ± 1.3, equivalent to an ionization yield of $45.7 \pm 7.0\,e^-/$keVee. For the 2.414 keV electrons emitted by the K-shell decay, instead, the situation is complicated by the complex event topology, which in turn makes it very difficult to build a model for the recombination. Therefore, for the next steps involving the calibration, the K-shell ^{37}Ar line has not been taken into account.

5.3.2 Electronic recoil ionization yield calibration

The ER ionization yield has been calibrated using not only the ^{37}Ar L1 shell line but also the ^{39}Ar data acquired during a run with AAr in 2013-2014, with the same drift field configuration of the subsequent DM search campaign. In fact, the β emission of ^{39}Ar is the dominant background when using AAr, with an activity of ~ 1Bq/kg [130]. Even if the ^{39}Ar β emission produces events uniformly in the whole TPC, only the events in a small fraction of the active volume, corresponding to a central cylinder of 21.6 cm height and 2 cm radius, has been selected for this analysis. This has been done in order to minimize the contribution of the background coming

from the detector materials themselves: indeed, the chosen fiducialization cylinder is 16.8 cm far from the TPC walls and 7 cm far from the top and the bottom of the TPC. As a result, the selected events are readout only by the central PMT, and the radial dependence of the detector response can be safely neglected. Finally, events with multiple S2 pulses, incompatible with the ^{39}Ar β decay topology, are removed by the data sample.

The total kinetic energy E_{er} of the event is measured using both the S1 and S2 signals, since it can be expressed as

$$E_{er} = w \left(\frac{S_1}{g_1} + \frac{S_2}{g_2} \right), \tag{5.14}$$

where $w = (19.5 \pm 1.0)$ eV is the average energy needed to produce a detectable quantum in LAr, and $g_1 = (0.16 \pm 0.01)$ [141] is the S_1 collection efficiency. The energy defined by means of Eq. 5.14 is also referred to as "rotated energy". On the other hand, the number $N_{i.e.}$ of ion-electron pairs surviving recombination is

$$N_{i.e.} = \frac{S_2}{g_2} - 1, \tag{5.15}$$

where we subtracted the primary β electron producing the event. The mean ER ionization yield can be therefore computed as

$$Q_y^{ER} = \frac{N_{i.e.}}{E_{er}}, \tag{5.16}$$

where $N_{i.e.}$ and E_{er} are computed, bin-by-bin, by means of Eq. 5.15 and Eq. 5.14, respectively, with a bin width of 0.2 keV. The resulting measurements, together with the ^{37}Ar L1-shell data, are represented by the points in Fig. 5.7, where the uncertainty is mostly coming from the systematic effect induced by the uncertainty on g_2. The bins corresponding to an energy $E_{er} < 1.7$ keV are not considered, since in the low energy region the S1 pulse identification efficiency is not 100%.

The number of ion-electron pairs $N_{i.e.}$ surviving recombination are a $1 - f$ fraction of the total number N_i of ion-electron pairs produced after the electronic recoil, where f is the recombination probability. The ER ionization yield is therefore given

$$Q_y^{ER} = \frac{(1-f)N_i}{E_{er}}. \tag{5.17}$$

In the $\mathcal{O}$ (keV) low ER energy region the recombination probability is described by the Thomas-Imel model - see Eq. 5.4. In particular, it is given by

$$1 - f = \frac{1}{\gamma N_i} \ln (1 + \gamma N_i), \tag{5.18}$$

where the free parameter $\gamma = \xi/N_i$, and ξ is defined in Eq. 5.5. In the low ER energy range, we can safely assume that N_i is proportional to E_{er}, and introduce a new free parameter $\rho \equiv N_i/E_{er}$. The validity of this assumption has been tested repeating the calibration analysis with more refined parameterizations [151], with
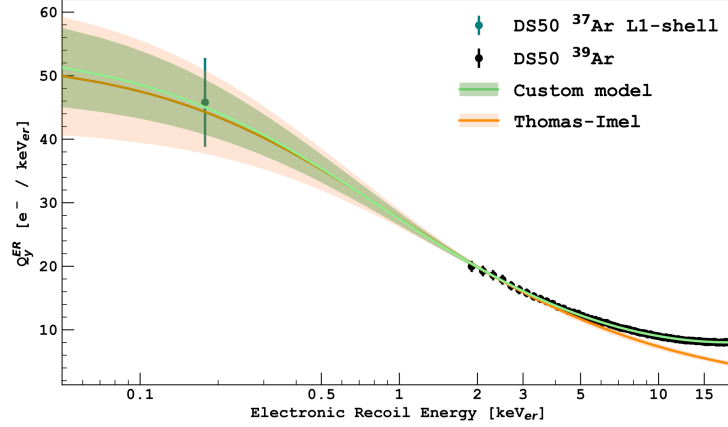
Figure 5.7. The ER ionization yield measured from the ^{37}Ar L1-shell data (teal) and AAr data (black). The orange line is the best fit of data, up to 3 keV, with the Thomas-Imel model of Eq. 5.19. The green line is the best fit obtained with the custom empirical model of Eq. 5.20, extending the Thomas-Imel model up to 20 keV. The colored bands are the 1σ regions, including also the g_2 uncertainty propagation. Figure from Ref. [144].

no significant changes in the results with respect to the constant ρ assumption [144]. Therefore, the ER ionization yield can be expressed as

$$Q_y^{ER} = \frac{1}{\gamma E_{er}} \ln\left(1 + \gamma \rho E_{er}\right). \tag{5.19}$$

The Thomas-Imel model, as shown in Fig. 5.7, is in good agreement with the data up to 3 keV. The fit performed in the $E_{er} < 3$ keV region yields $\rho = (52.7 \pm 10.9)$ keV^{-1} and $C_{box} = (8.6 \pm 1.5)$ V/cm, where $C_{box} = \gamma \mathcal{E}$ and $\mathcal{E} = 200$ V/cm is the drift electric field. For energies greater than 3 keV, the Thomas-Imel model is no more valid. The empirical Doke-Birks model [151], which is in agreement with the ARIS experiment data [152] up to 40 keV, is not in agreement with the ^{37}Ar and ^{39}Ar data of DarkSide-50 in the $E_{er} = [3, 20]$ keV ER energy region. This issue is overcome by means of a custom empirical model that extends the Thomas-Imel thanks to an additional term, as follows

$$Q_y^{er} = \left[\frac{1}{\gamma} + p_0 \left(\frac{E_{er}}{\text{keV}}\right)^{p_1}\right] \frac{\ln\left(1 + \gamma \rho E_{er}\right)}{E_{er}}, \tag{5.20}$$

where p_0 and p_1 are two additional positive free dimensionless parameters. Data in Fig. 5.7 are in excellent agreement with this model, and the results of the fit are: $C_{box} = (9.2 \pm 0.9)$ V/cm, $\rho = (54.4 \pm 7.3)$ keV^{-1}, $p_0 = 0.11 \pm 0.03$ and $p_1 = 1.71 \pm 0.08$. The parameters C_{box} and ρ are consistent with the results of the fit with the Thomas-Imel model. Moreover, the results for the p_0 and p_1 parameters show that the additional custom term is negligible with respect to the $1/\gamma$ term in the $E_{er} < 3$ keV region, reproducing in the low energy region the Thomas-Imel model.

5.3.3   Nuclear recoil ionization yield calibration

In the NR case we can proceed similarly to the ER: the NR ionization yield is given by

$$Q_y^{NR} \equiv \frac{N_{i.e.}}{E_{nr}} = \frac{(1-f)N_i}{E_{nr}}, \tag{5.21}$$

where E_{nr} is the total kinetic energy of the nuclear recoil, N_i is the number of ion-electron pairs produced after the nuclear recoil, and f is the recombination probability, described in Eq. 5.18 by the Thomas-Imel model. According to the Bezrukov model [138], N_i can be computed as

$$N_i = \beta \, \kappa(\varepsilon)$$
$$\kappa(\varepsilon) = \varepsilon \frac{s_e(\varepsilon)}{s_e(\varepsilon) + s_n(\varepsilon)}, \tag{5.22}$$

where:

- β is a free parameter, fixing the normalization. The other parameter of the model is C_{box}, which is contained in the definition of f - see Eq. 5.18;

- κ is a dimensionless factor which has the physical interpretation of the energy resealed via electronic excitations, in turn producing ionization and scintillation signals;

- s_e and s_n, also known as electronic and nuclear screening functions, are the rate of excitation due to inelastic collisions and the rate of the energy transfer to other nuclei via elastic collisions, respectively;

- ε is a dimensionless quantity defined as

$$\varepsilon = \frac{a}{2e^2 Z^2} \left(\frac{E_{nr}}{\text{keV}} \right), \tag{5.23}$$

  where Z is the Ar atomic number, e is the electron charge and $a = 0.626 a_0 Z^{-1/3}$ is the Thomas-Fermi screening length [138], with a_0 being the Bohr radius.

The quantity s_e is given by [138]

$$s_e(\varepsilon) = \frac{0.133 \, Z^{2/3}}{A^{1/2}} F\left(\frac{v}{v_0} \right) \sqrt{\varepsilon}, \tag{5.24}$$

where A is the Ar atomic mass number, and $F(v/v_0)$, with v being the nuclear speed and $v_0 = e^2/\hbar$, is a correction factor depending on the velocity of the nucleus. Since there are no theoretical calculations supporting the possibility that F is either a suppression or an enhancement factor, it will be conservatively assumed to be $F(v/v_0) = 1$ [144]. Finally, according to the Ziegler et al. model [153], s_n is given by

$$s_n(\varepsilon) = \frac{\ln\left(1 + 1.1383 \, f_Z \varepsilon \right)}{2 \left[f_Z \varepsilon + 0.01321 \, (f_Z \varepsilon)^{0.21226} + 0.19593 \, (f_Z \varepsilon)^{0.5} \right]}, \tag{5.25}$$

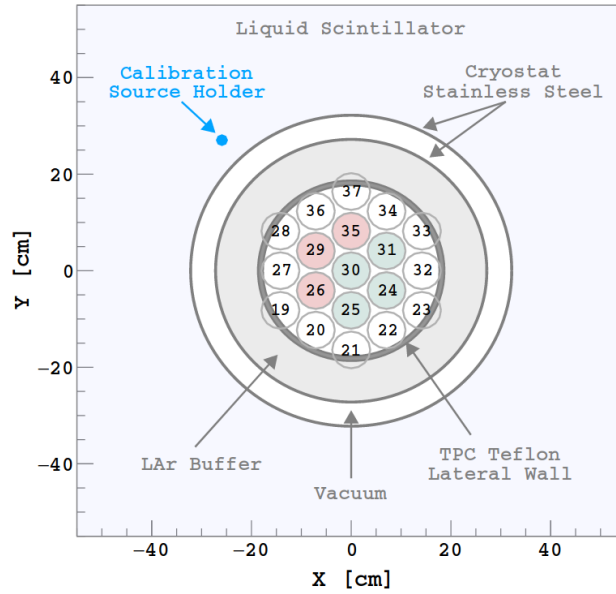where, as a consequence of our definition of ε, $f_Z = 0.953$.

Figure 5.8. Sketch of the top view of the DarkSide-50 detector. The numbered circles represent the top PMTs. Only the events whose maximum S2 yield is measured by the colored PMTs are considered for the AmBe analysis. The AmC analysis, instead, considers only the events whose maximum S2 yield is measured by the green PMTs, in order to suppress the γ background coming from the source itself. Figure from Ref. [144].

The two parameters of the NR ionization yield model outlined above have been determined by performing a fit on calibration data. These data consists in: calibration runs with an $^{241}\text{Am}^{13}\text{C}$ (AmC) or an $^{241}\text{Am}^{9}\text{Be}$ (AmBe) neutron source, and "external" data from the ARIS [152], SCENE [143], and Joshi et al. [154] experiments. The AmC and AmBe neutron sources were placed in both cases inside the LSV, as shown in Fig. 5.8. These calibration datasets underwent independent data selections [144], in order to optimize the background rejection in the two cases. The AmC and AmBe data spectrum are shown in Fig. 5.9.

The AmC data are selected by keeping only the events corresponding to the PMTs 24, 25, 30, and 31, depicted in green in Fig. 5.8, in order to suppress the residual γ background induced by the source which is not absorbed by the LAr buffer and the other materials before entering the TPC. However, a small fraction of γs survives and reaches the four fiducial PMTs. The AmC data spectrum is obtained by subtracting the normalized UAr data spectrum and the expected residual γ background from the observed data. This expected background is computed via MC simulations [144], in turn validated on and constrained by the higher energy $N_{e^-} = [290, 900]$ control region, where no NR is expected.

For what regards the AmBe data, characterized by a different and more energetic γ spectrum with respect to the AmC source, a more refined data selection, requiring also the S1 signal, is applied [144]. The events corresponding to all the colored PMTs of Fig. 5.8 are taken into account. The requirement of an S1 signal is the reason why the AmBe spectrum in Fig. 5.9 is suppressed in the low N_{e^-} region. An ER contamination is however still present in the AmBe dataset, and it has been
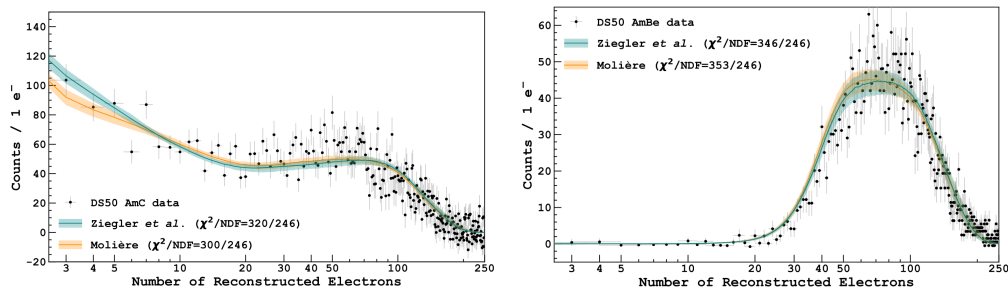
Figure 5.9. AmC (**left**) and AmBe (**right**) data (black points). The blue curve corresponds to the best fit MC simulated N_{e^-} spectra using the Ziegler et al. [153] screening function, see Eq. 5.25. The orange curve corresponds to an alternative screening function by Molièr [155]. The colored bands are the $1\,\sigma$ regions. Figure from Ref. [144].

accounted for validating the procedure on the control region $N_{e^-} = [250, 500]$ and $f_{90} < 0.4$, where no NR is expected. Also in this case, the normalized UAr spectrum is subtracted.

Finally, further external measurements of neutron induced NRs from other LAr detectors has been considered. The ARIS experiment [152] measured the NR ionization yield for 8 different NR energies in the $E_{nr} = [7.1, 117.8]$ keV range. In order to account for the differences in the experimental setup between ARIS and DarkSide-50, the data from ARIS has been appropriately renormalized [144]. The SCENE experiment [143] measured the NR ionization yield for 4 different energies in the $E_{nr} = [16.9, 57.3]$ keV range. Also in this case data has been properly renormalized in order to be compared to DarkSide-50. Lastly, the Joshi et al. [154] experiment produced a measurement of the NR ionization yield at $E_{nr} = 6.7$ keV. However, this measurement has not been taken into account in the analysis because the data point has been corrected using the Joshi et al. 2.82 keV K-shell ^{37}Ar line and the one from DarkSide-50 itself as a cross-calibration point [144]. The data from the external datasets are shown in Fig. 5.10.

The global fit to the data has been performed with a toy MC approach, in which the AmC and the AmBe spectra to be fit on data are generated thanks to a Geant4 simulation for each value of the two parameters β and C_{box}. For the analysis, the AmC and AmBe data in the $N_{e^-} = [3, 250]$ range has been taken into account. The χ^2 global fit yields $\beta = (6.8)^{+0.1}_{-0.3} \times 10^3$ and $C_{box} = 8.1^{+0.1}_{-0.2}$ V/cm, and the correspondent best fit curves are shown in Fig. 5.9 and Fig. 5.10. This is the lowest NR ionization yield calibration ever achieved in LAr, since the AmC data constrains the model up to 3 ionization electrons, corresponding to $E_{nr} = 435^{+47}_{-34}$ eV [144].

As it is shown in Fig 5.9 also different models for the s_n screening function have been used. In particular, the NR ionization yield analysis has been performed also with the Moliere [155] and the Lenz-Jensen [156,157] models, finding that they both are in agreement with the data. Fig. 5.11 shows the results of the three analyses in terms of Q_y^{NR}.
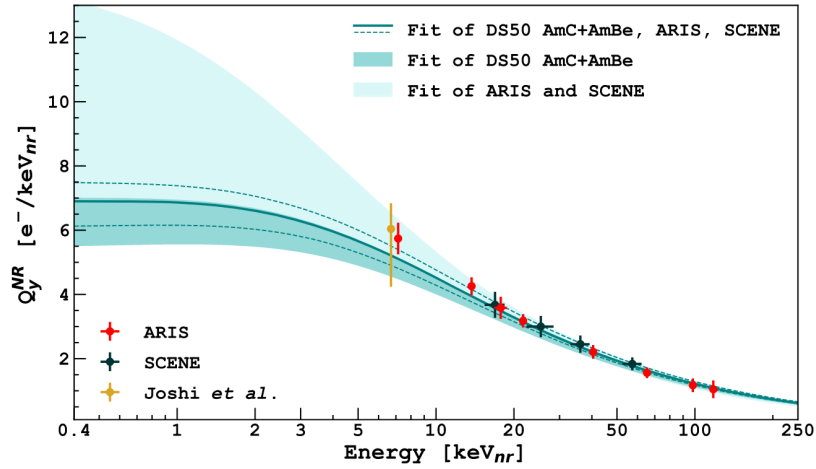
Figure 5.10. Data from the external datasets. The blue curve is the fit performed on the
AmC, AmBe and the renormalized external datasets from ARIS [152] and SCENE [143]
at the same time. The dashed curves correspond to the 1σ bands. The Joshi et al. point,
which has been excluded from the fit, is reported for comparison. The colored shaded
areas correspond to 1σ bands of the fits performed singularly on the DarkSide-50and
the external datasets, respectively. Figure from Ref. [144].

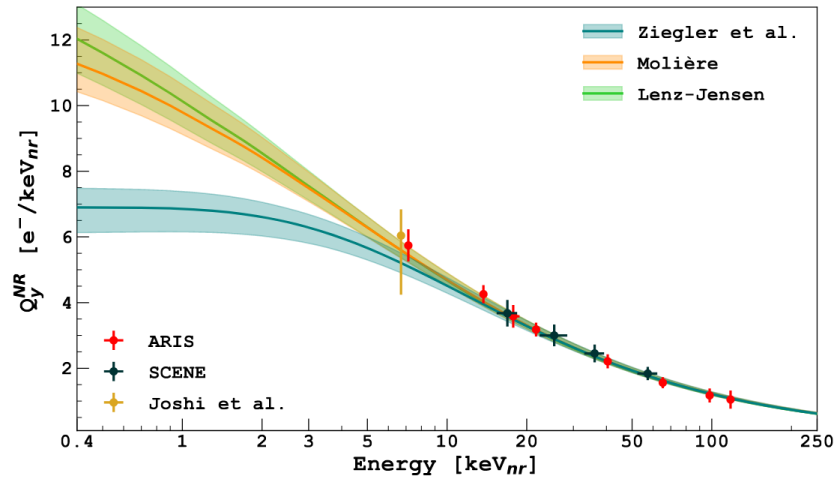**Figure 5.11.** Global fit of AmC and AmBe DarkSide-50 data and ARIS [152] and
SCENE [143] datasets for different nuclear screening functions. The blue curve cor-
responds to the Ziegler et al. [153] model, while the orange and the green curves cor-
respond to the Moliere [155] and the Lenz-Jensen [156, 157] models, respectively. The
shaded regions correspond to the 1σ bands. Figure from Ref. [144].

5.3.4 Background Sources

The relevant background sources for the low-mass DM searches can be divided into two categories: the "internal" background, due to the intrinsic β emissions from ^{39}Ar and ^{85}Kr contamination of the UAr inside the TPC, and the "external" background, due to the natural radioactivity originating in the PMTs and the cryostat, the two most relevant TPC components.

Thanks to the analysis of the S1 spectrum above 50 pe, the activity of the ^{39}Ar isotope has been measured to be (0.73 ± 0.10) mBq/kg [60]. Regarding instead the ^{85}Kr activity, a new analysis on the same dataset of Ref. [60], after fixing a bug in the MC simulations and considering the ^{85}Kr decay (with an half life of 10.76 yr) during the data taking, yielded (1.89 ± 0.12) mBq/kg. This result has been validated with two additional independent measurements. The first one has been obtained by identifying β-γ fast coincidences from the 0.43% decay branch to ^{85m}Rb with 1.46 μs mean lifetime: this resulted in an activity of 1.85 ± 0.15 mBq/kg. The second one has been obtained by fitting the ^{85}Kr decay time in the low energy $N_{e^-} = [50, 200]$ region, which yielded (1.75 ± 0.23) mBq/kg, in agreement with the other measurement. The weighted mean of the three measurements is 1.86 ± 0.09 mBq/kg. The above measurements of the activity of ^{39}Ar and ^{85}Kr will be used as prior *pdf*s for the normalization of these two background components in the analysis outlined in Chap. 6.

With respect to the preliminary analysis of DarkSide-50 in 2018 [1], a major improvement has been achieved thanks to the new calculation of the ^{39}Ar and ^{85}Kr spectral shapes. The new calculation, performed with an improved version of the `BetaShape` code [145], takes now into account the atomic exchange effect to forbidden unique transitions. This effect consists in the probability that the electron involved in the atomic β emission is produced in a bound state, resulting in a small enhancement of the spectrum at low energies [158]. The new corrections were validated on measured ^{63}Ni and ^{241}Pu spectra. These measurements have a lower threshold of 200 eV: therefore, below this threshold, a 10% uncertainty on the ^{39}Ar and ^{85}Kr spectra, referred to as "screening uncertainty", has been included[2]. Another source of uncertainty on the theoretical spectrum is the uncertainty on the β Q-value, estimated to be 1% for ^{39}Ar and 0.4% for ^{85}Kr. The left plot of Fig. 5.12 shows the internal background theoretical spectra with their associated uncertainties. The dashed lines are the spectra without the corrections, which instead enhance the β spectral shape in the low energy region.

The activities of the PMTs and the cryostat background contributions are assumed from the material assay measurements. The dominant components are the ^{238}U, ^{235}U, and ^{232}Th chains and the ^{60}Co, ^{54}Mn, and ^{40}K isotopes. The theoretical spectra are computed by means of a very high statistics Geant4 simulation, generating these isotopes in the cryostat and the PMTs. For what regards the PMTs, the contaminant location has been determined to be the stems in the back of the PMTs and the ceramic in the PMT body. The original positions of the isotopes do not alter the shape of the background spectra, but have an impact on the relative event rate: therefore, in addition to the uncertainties from the material assays mea-

---

[2]X. Mougeot, author of `BetaShape`, suggested quantifying this uncertainty as a $\sim 10\%$ uncertainty on the spectral shapes.
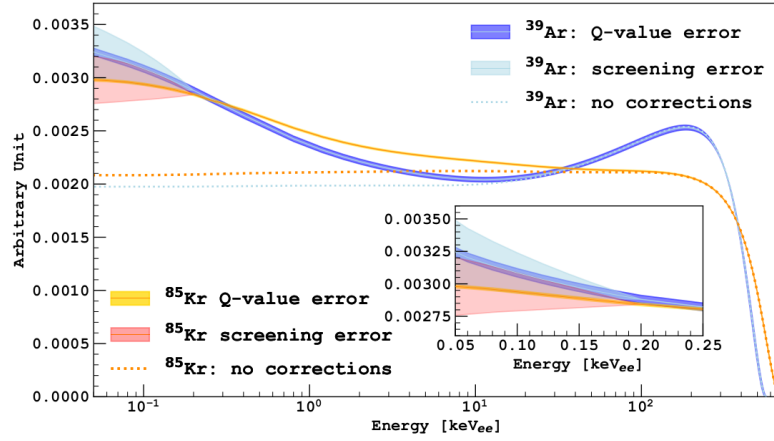
Figure 5.12. ^{39}Ar and ^{85}Kr β decay spectra in keV. The different bands correspond to different uncertainty contributions. The dashed lines correspond to the theoretical spectra without the atomic exchange effect corrections.
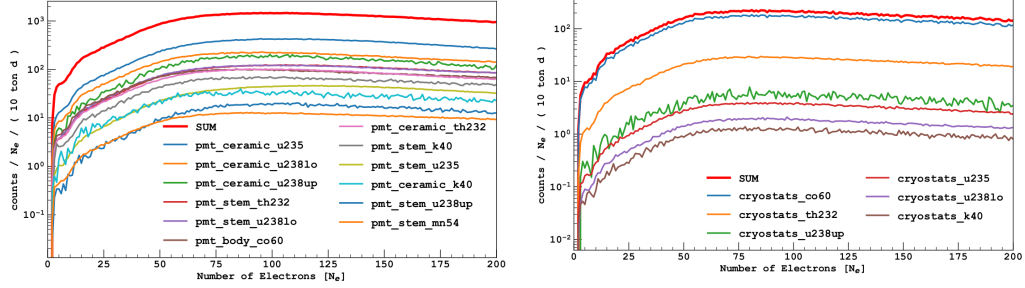

Figure 5.13. Individual components of the PMT (**left**) and the cryostat (**right**) background spectra, scaled by each measured activity. The red curves represent the summed spectra.

surements, the uncertainty associated with the partitioning of the different isotopes between stems and body must be considered. In Fig. 5.13 the list of the spectral contributions of all the single PMT and cryostat components is shown: the N_{e^-} spectrum is obtained by using the MC to properly generate events in TPC and applying, event by event, the detector response model. The predicted background event rate for the PMT component is $(4.72 \pm 0.60) \times 10^{-3}$ Hz and for the cryostat component is $7.29^{+0.48}_{-0.53} \times 10^{-4}$ Hz.

Chapter 6

DarkSide-50 analysis

In this chapter, we will present an innovative analysis of the DarkSide-50 ionization-only data aimed at searching for low mass dark matter candidates. For this analysis, we consider only low mass $(0.06\text{-}10\,\mathrm{GeV/c^2})$ DM candidates elastically scattering on atomic nuclei. We use the extended DarkSide-50 dataset of about $10\,\mathrm{ton}\times\mathrm{day}$, the latest calibration of the electron and nuclear recoil energy scales, the updated spectral correction for $^{39}\mathrm{Ar}$ and $^{85}\mathrm{Kr}$ beta decays, and the latest background model.

The innovative approach in our analysis consists in a comprehensive treatment of all systematic uncertainties. We exploit the new parametric detector response, retaining its analytical structure, to construct analytical or numerical expressions for each spectrum in the background and signal models. In such a way, we maintain a direct dependence of the spectra on each detector response parameter. A similar implementation is pursued for the parameters of the background and signal models. This approach allows writing a likelihood that explicitly depends on all such parameters and naturally retains the correlations between its different factors. The likelihood function is used to perform a Bayesian fit to the observed DarkSide-50 spectrum and compute the joint posterior *pdf*. The impact of the systematic uncertainties on the intensity parameter of the sought signal is obtained by marginalizing the joint *pdf* on all but the intensity parameter.

The information enclosed in the posterior *pdf*s can be directly associated with parameters which have a clear and meaningful interpretation, such as calibration and background parameters. We show that the inclusion of the systematic uncertainty degrades the statistical-only sensitivity by at the most a factor of ~ 1.6. The ability of the fit to constrain the detector response and background models substantially helps in reducing the impact of the systematic uncertainties on the sensitivity.

This analysis is implemented in a C++ and CUDA code [159] to construct the inputs to the likelihood function, and in BAT [116] for the Bayesian fit.

The structure of this chapter is the following: Sec. 6.1 describes the observed data, the background, and the signal input spectra; Sec. 6.2 presents the implementation of the detector response model; Sec. 6.3 describes the likelihood function; the technical description of the fitting procedure is given in Sec. 6.4; Sec. 6.5 presents the results of this work; finally, Sec. 6.6 summarizes and concludes the chapter.

Variable	Description
s1	Number of S1 photoelectrons
s2	Number of S2 photoelectrons
ne_corr	Number of S2 extracted electrons, defined as $S2/g_2$, where here g_2 is the mean number of pe/e^- measured by each PMT
s1_f90	f_{90} applied to S1 pulses
s2_f90	f_{90} applied to S2 pulses
s2_tba	S2 top/bottom asymmetry
npulses	Number of identified pulses
s2_max_frac	Maximum fraction of S2 light observed by one of the top PMTs
s2_max_chan	Top array PMT observing the maximum fraction of S2 light
deltaT	Time difference with respect to the previous event
sX_start_time	Start time of the pulse X, with X = $\{1, 2\}$
sX_end_time	End time of the pulse X, with X = $\{1, 2\}$
s2_peak_time	Time of the S2 pulse peak, identified with a moving average over a 64 ns gate, with respect to the pulse start time
s2_gate	Duration of the S2 pulse, defined as s2_end_time − s2_start_time
s2_fwhm	Full Width at Half Maximum (FWHM) of the S2 peak, computed on the moving average over a 64ns gate

Table 6.1. List of variables used for the data selection and related description.

6.1 Data, background, and signal spectra

Data selection

The data used for this analysis has been taken from December 12, 2015, to October 4, 2017, excluding the runs with trigger rate outside the [1.2, 1.8] Hz range, that are calibration runs. The total time exposure corresponds to 535.6 d of data taking. All the variables used in the data selection are listed in Tab. 6.1.

The events used for the low mass analysis are classified in two categories: the S2-only events and the S1+S2 events, depending on the absence/presence of an S1 signal. In order to select these events, three classes of selection cuts are applied: the global cuts, applied both to the S2-only and to the S1+S2 events, and additional dedicated cuts applied separately to the S2-only or the S1+S2 events. They are reported for completeness in Tab. 6.2, where the list of all the selection cuts and their description is shown. The efficiencies reported in the table are computed on the basis of MC simulated events reconstructed by means of the reconstruction algorithm or using the ^{83m}Kr calibration sample.

The resulting data spectrum is depicted in Fig. 5.5, compared to the dataset

Cut	Description	Efficiency
Global Cuts		
`s2_max_chan` is one between: $\{24, 25, 26, 29, 30, 31, 35\}$	Fiducial volume cut	0.412
`deltaT` > 20	Remove events if within $20\,\text{ms}$ from the preceding one, to avoid correlated events	0.97
`s2_gate` < 100	S2 pulse duration less than $100\,\mu s$, because long gates are associated to unresolved multiple pulses	1.00
$0.1 <$ `s2_tba` < 0.95	Top/bottom asymmetry in the $[0.1, 0.95]$ range, to exclude events with light concentrated entirely in the top or bottom PMT arrays	1.00
`s2_f90` < 0.1	Remove misidentified S1 pulses	1.00
`s2_max_frac` < 0.75	Remove with more than 75% of the light concentrated in a single top PMT	1.00
`s2_peak_time` > 0.2	Remove events with a peak time within $200\,\text{ns}$ from the start time, likely related to alpha particles from the lateral walls inducing S1 in coincidence with a spurious S2	1.00
`s2_peak_time` < 6	Remove events with a peak time greater than $6\,\mu s$ from the start time, likely corresponding to unresolved multiple S2 pulses. The efficiency has been estimated on the ^{83m}Kr sample.	1.00
`s2_fwhm` > 0.1	Remove events with S2 FHMW lower than $100\,\text{ns}$, compatible with S1 pulses	1.00
S2-only Cuts		
`npulses` $= 1$ or (`npulses` $= 2$ and S2 echo)	Either 1 pulse or two pulses where the second is an S2 delayed by more than $375\,\mu s$	1.00
`s2_start_time_cut`	Cut defined as a function of `ne_corr` using MC and the full reconstruction algorithm in order to remove unresolved multiple pulses. The main criterion is based on the pulse start time: if it is too early compared to the trigger time, it cannot be associated with a physical event. The cut is constructed to obtain a flat efficiency as a function of `ne_corr`.	0.99
S1+S2 Cuts		
`npulses` $= 2$ or (`npulses` $= 3$ and S1 echo) or (`npulses` $= 3$ and S2 echo)	Either 2 pulses or 2 pulses plus an S1 or S2 echo	1.00
$-6.1 <$ `s1_start_time` < -6.0	S1 start time close to the trigger time offset	1.00
`s1_f90` < 0.85	Remove Cherenkov events	1.00
rotated energy $< 25\,\text{keV}_{\text{er}}$	Remove high energy S1 events (mostly alphas)	1.00
`alpha_cut`	S2/S1 vs S1 cut, build on the AmBe calibration events, to remove alpha events close to the wall, with low S2, since most of the electrons are absorbed by the TPC surface	0.99
	Total efficiency	0.396

Table 6.2. Selection cuts applied for the low mass analysis. The Global Cuts are applied both to the S2-only and the S1+S2 events, while the S2-only Cuts and the S1+S2 Cuts are additional dedicated cuts applied separately to the S2-only or the S1+S2 events.
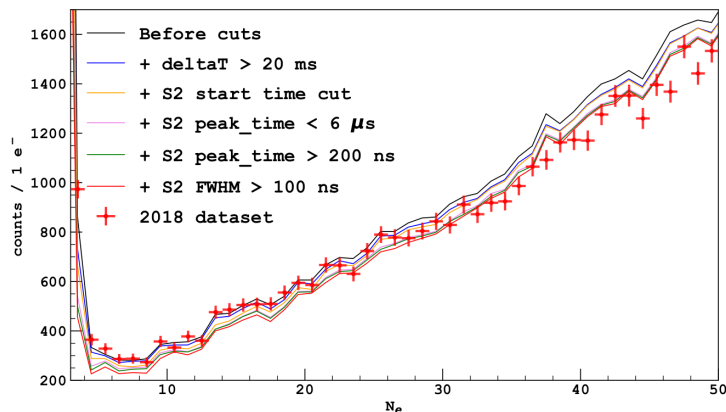
Figure 6.1. Comparison between the 2018 dataset (red points) with the current one, after applying a sequence of cuts that differs from the ones applied for the 2018 analysis [1]. The spectra are normalized to the current exposure of the new dataset.

considered in the 2018 analysis [1]. The difference in the $N_{e^-} \sim 50$ region is due to the non application of the $f_{90}^{S1} > 0.15$ cut. The depletion of events in the low N_{e^-} region is instead the result of the combination of a set of cuts, as depicted in Fig. 6.1.

Background and signal spectra

The background contributions considered in the analysis are those listed in Sec. 5.3.4 and shown in Fig. 5.12, 5.13. In addition to the systematic uncertainties on the calibration parameters, to the screening uncertainty and to the uncertainty on the β Q-value for the ^{39}Ar and ^{85}Kr, additional normalization systematic uncertainties has been considered: 14% on ^{39}Ar, 4.7% on ^{85}Kr, 12.6% on the PMT background and 6.6% on the cryostat background. The uncertainty on the ^{39}Ar normalization has been evaluated by means of a dedicated high energy background fit, while the uncertainty on the ^{85}Kr normalization has been estimated by means of the ^{85m}Rb fast coincidence analysis. The uncertainty on the cryostat background normalization stems from the uncertainty on the activity of the different isotopes, while the uncertainty on the PMT background normalization takes into account also the uncertainty on the location of the contaminants.

The DM scattering off nuclei (including also the Migdal effect) is simulated by assuming the Standard Halo Model with $v_{esc} = 544$ km/s, $v_0 = 220$ km/s, $v_{Earth} \simeq 233$ km/s and $\rho_{DM} = 0.3$ GeV cm^{-3} c^{-2}. In our simulation, based on the code reported in Ref. [160] and available on `GitHub`, we compute both the NR spectrum and the double differential Migdal effect spectrum, using the Migdal ionization probabilities given in Ref. [3]. This simulation has also been used in Ref. [2]. The signal spectra, which releases energy not only in the NR channel but also in the ER channel via the Migdal electron emission, assume that the detected nuclear and electronic recoil induced ionization are independent. Furthermore, we consider single ionization events and neglect double ionization. We consider however the fact that the ionized atom will return to the ground state, releasing further

electronic energy (equal to the binding energies that range from ~ 15 eV of the outer shells to ~ 3.2 keV of the inner one) in the form of photons or additional electrons. In principle, the computation of the ionization probability of the most external shell is affected by large theoretical uncertainties [3]. However, we do not consider any theoretical uncertainty on the spectra coming from the computation of the ionization probabilities, following the results in Ref. [93].

6.2 Detector response model

We will now briefly summarize the characterization of the detector response model, outlined in detail in Sec. 5.3.1, describing how it is possible to use the results of the calibration to keep the dependence on the calibration parameters up to the final N_{e^-} spectrum, where N_{e^-} is the number of reconstructed primary electrons. The detector response can be schematized in the following consecutive steps:

1. Conversion of the deposited energy to a certain number of detectable quanta (e.g. number of primary electrons) produced during the interaction.

2. Detection efficiency effects that depend on the position of the events inside the TPC.

3. Resolution effects of the photodetectors (e.g. PMTs).

All these steps contribute to distorting the original theoretical energy spectrum into a different observed one.

6.2.1 From energy deposit to N_{e^-}

The incoming particles interacting with the active volume inside the TPC can produce an electronic recoil (ER) or a nuclear recoil (NR). In terms of detector response, the difference in the deposited energy between the two cases lies in the production mechanisms of detectable quanta: in Sec. 5.3 a different yield of detectable quanta for the ER and for the NR interactions has been considered.

The average number of primary ionization electrons produced during, for instance, a NR is given by

$$\left\langle N_{e^-}^{NR} \right\rangle = E_{nr} Q_y^{NR} \left(E_{nr}, \boldsymbol{\theta}_{\mathrm{cal}} \right), \tag{6.1}$$

where E_{nr} is the energy deposited via NR, Q_y^{NR} is the yield per unit of energy and θ_{cal} is the complete list of calibration parameters that regulate the ER and NR yield functions. The production of an ion-electron pair surviving recombination is however a stochastic process, and in particular it can be represented by a binomial process. The maximum number of electrons that can be produced at a given energy E_{NR} can be estimated as

$$N_{e^-}^{\mathrm{max},\, NR} = \frac{E_{nr}}{w}. \tag{6.2}$$

The probability that a certain amount of energy is released in the TPC in the form of an ion-electron pair can be estimated, using Eq. (6.1), as:

$$\epsilon^{NR} = \frac{\left\langle N_{e^-}^{NR} \right\rangle}{N_{e^-}^{\mathrm{max},\, NR}} = w\, Q_y^{NR} \left(E_{nr}, \boldsymbol{\theta}_{\mathrm{cal}} \right). \tag{6.3}$$

Thus the probability of having produced a certain number N_{e^-} of ion-electron pairs is given by the binomial distribution

$$P(N_{e^-}^{NR}|E_{nr}, \boldsymbol{\theta}_{\text{cal}}) = \mathcal{B}(N_{e^-} \,|\, p = \epsilon^{NR}, n = N_{e^-}^{max}), \qquad (6.4)$$

where

$$\mathcal{B}(k \,|\, p, N) = \binom{N}{k} p^k (1-p)^{N-k}. \qquad (6.5)$$

For the ERs we have a different yield function Q_y^{ER} and the statistical fluctuations on the number of produced detectable quanta are implemented as normal fluctuations via the Fano factor. The probability of having a certain number N_{e^-} of primary ionization electrons can be represented as

$$P(N_{e^-}^{ER}|E_{er}, \boldsymbol{\theta}_{\text{cal}}) = \int_{N_{e^-}^{ER}}^{N_{e^-}^{ER}+1} \mathcal{N}\left(x|\mu = \left\langle N_{e^-}^{ER}\right\rangle, \sigma = \sqrt{F\left\langle N_{e^-}^{ER}\right\rangle}\right) dx \qquad (6.6)$$

where

$$\left\langle N_{e^-}^{ER}\right\rangle = E_{er}\, Q_y^{ER}\left(E_{er}, \boldsymbol{\theta}_{\text{cal}}\right), \qquad (6.7)$$

$$\mathcal{N}\left(x|\mu, \sigma\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (6.8)$$

and F is the Fano factor.

Once we have an expression for these probabilities that depends on the energy of the theoretical spectrum and the calibration parameters, we can compute the expected spectrum in N_{e^-} after recombination in a very straightforward way. Let's suppose that the theoretical spectrum, either ER or NR, is in the form of an histogram, and let's define x_j^{th} as the energy of the j-th bin and y_j^{th} the content of the j-th bin. We can then define the probability matrix $\mathcal{M}^1$ as

$$\begin{aligned} \mathcal{M}_{i,j}^{1,\,ER} &= P(N_{e^-}^{ER} = N_{e^-}^i|E_{er} = x_j^{th}, \boldsymbol{\theta}_{\text{cal}}) \\ \mathcal{M}_{i,j}^{1,\,NR} &= P(N_{e^-}^{NR} = N_{e^-}^i|E_{nr} = x_j^{th}, \boldsymbol{\theta}_{\text{cal}}) \end{aligned} \qquad (6.9)$$

where $N_{e^-}^i$ is the number of primary ionization electrons in the expected spectrum. Then the expected spectrum y_i^{exp} can be computed as the product of the $\mathcal{M}^1$ matrix and the theoretical spectrum, namely

$$y_i^{exp} = \sum_j \mathcal{M}_{i,j}^1 y_j^{th}, \qquad (6.10)$$

where we have omitted the ER, NR indices to simplify the notation.

The $\mathcal{M}^1$ probability matrix is also known as the smearing matrix, and has some important properties. First of all it is not a square matrix: in fact the number of bins in the theoretical and expected spectrum can be very different in principle. In addition, for the rule of probabilities, the sum of the columns of the matrix is equal to unity

$$\sum_{i=0}^{\infty} \mathcal{M}_{i,j}^1 = 1 \qquad \forall\, j. \qquad (6.11)$$

In the real applications obviously the number of bins in the theoretical spectrum is not infinite. The ending point of the theoretical spectrum must therefore be chosen in such a way to avoid finite size effects. In our particular case the observed spectrum is usually cut at $N_{e-} = 170$: we therefore need to check that, if N_{max}^{exp} is the number of bin of the expected spectrum after recombination,

$$\left| \sum_{i=0}^{N_{max}^{exp}=170} \mathcal{M}_{i,j}^{1} - 1 \right| < \delta \qquad \forall j, \tag{6.12}$$

with δ setting the precision of our approximation. However, the expected spectrum of detectable quanta obtained using Eq. 6.10 is not the final observed one because of other smearing effects, namely detection efficiency and resolution effects. It is therefore reasonable at this point to extend the length of the expected spectrum to a greater N_{max}^{exp} to avoid subsequent finite size effects. For a final spectrum cut at $N_{e-} = 170$, a good choice has been demonstrated to be $N_{max}^{exp} = 400$.

In order to visualize the structure of our statistical model up to the expected spectrum after recombination, we can draw the correspondent Bayesian networks. For the NR case, the Bayesian network can be drawn as
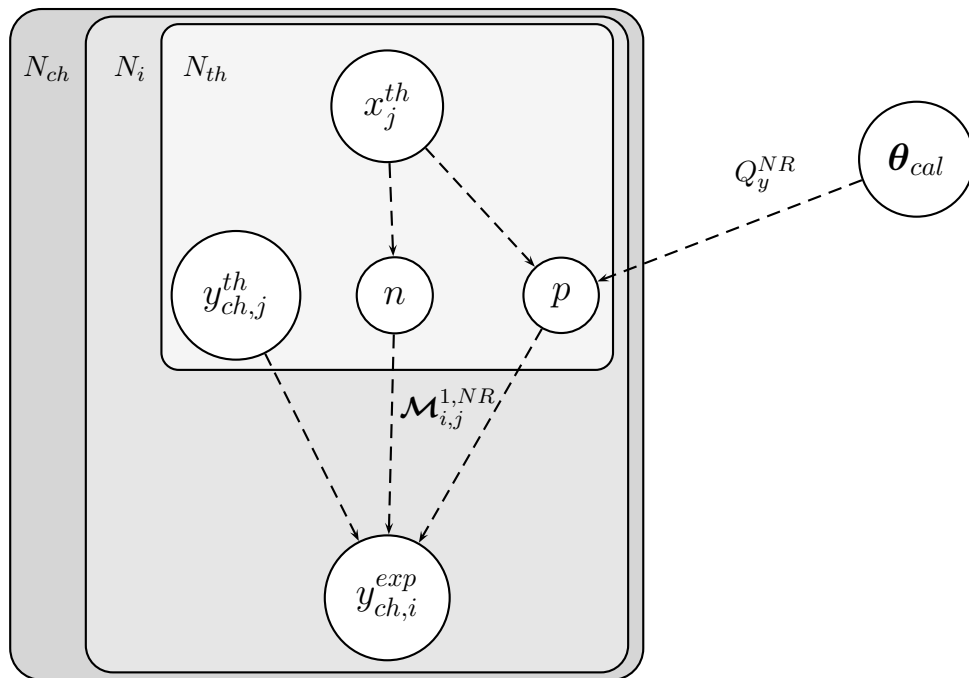
where $\boldsymbol{\theta}_{cal}$ are the calibration parameters, p and n are the parameters of the binomial *pdf* of Eq. 6.4, x_j^{th} is the j-th bin center of the theoretical energy spectrum in the ch-th PMT channel, and $y_{ch,j}^{th}$ is the correspondent bin content in the theoretical energy spectrum. N_{ch}, N_i, and N_{th} are the number of PMT channels, the number of bins of the resulting N_{e-} spectrum and the number of bins in the theoretical energy spectrum, respectively. For the ER case, the Bayesian network is in general quite similar, and can be drawn as
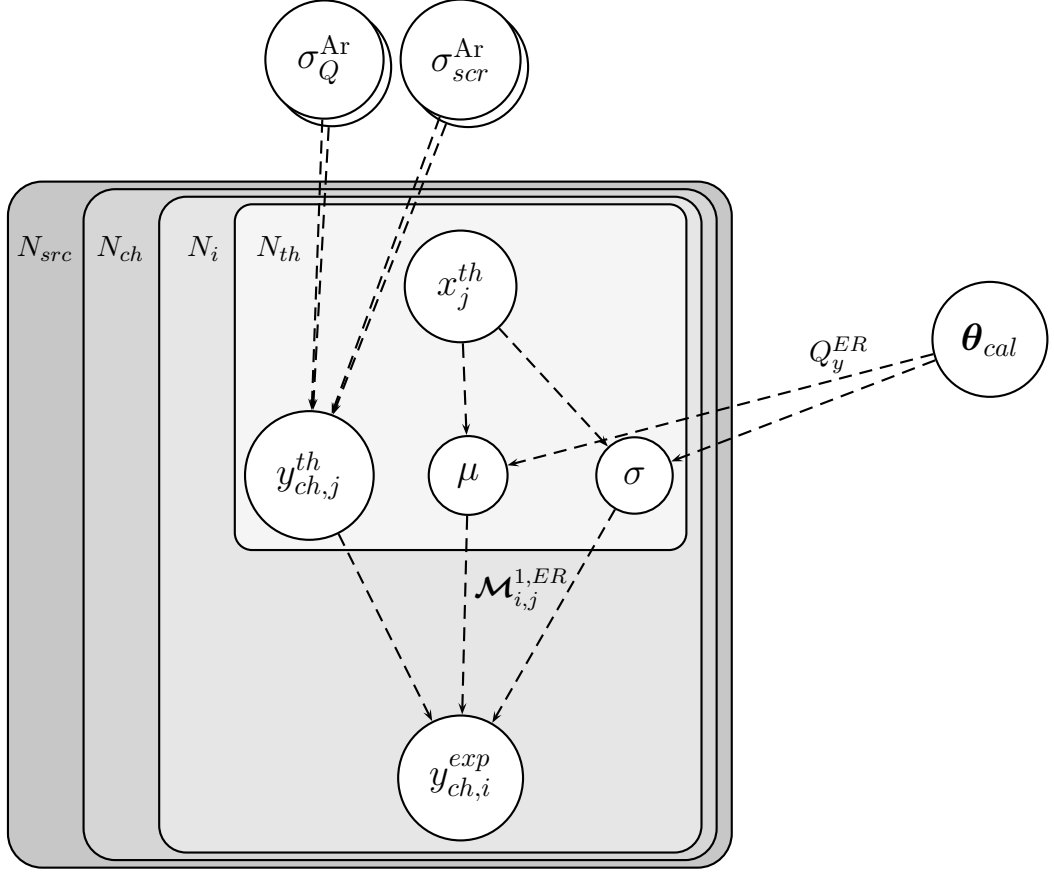
where now μ and σ are the parameters of the normal distribution of Eq. 6.6, and we have an additional gray box labeled with N_{src}, i.e. the number of the background sources, meaning that we have multiple contributions to take into account[1]. In addition, the nodes $\sigma_{scr}^{\mathrm{Ar,Kr}}$ and $\sigma_{Q}^{\mathrm{Ar,Kr}}$ represent the parameters regulating respectively the screening uncertainty and the Q-value uncertainty affecting the internal β backgrounds.

What is remarkable is that the $\mathcal{M}^1$ matrix as defined in Eq. 6.9 depends on an explicit way from the calibration parameters, and thus their uncertainties play the role of the systematic uncertainties of our experiment. It is then possible to take into account these parameters directly inside our likelihood, allowing us to treat the systematic uncertainties in a very clean and straightforward way, without using multi-templates methods to propagate the systematic uncertainties. These methods have in fact in general some drawbacks, as they often rely on some ad-hoc prescriptions (e.g. to take into account correlations, non-linear behavior, and so on) that derive from assumptions that are not always satisfied. This will be clearer and more explicit in the next sections, where the analysis will be described in detail.

---

[1]For the sake of notational simplicity, we omitted the *src* label in all the nodes of the network.

6.2.2 Implementation of the Migdal effect response model

A different way to proceed must be considered for the Migdal effect, since, during an event in which a Migdal electron is emitted, electrons both in the NR and ER channels are produced. Therefore, the total number of electrons produced in a single event is the sum of the electrons released by the NR and those released by the ER. The probability of releasing N_{e^-} primary ionization electrons, given a NR with energy E_{nr} and a Migdal emission depositing an energy E_{er}, is given by

$$
\begin{aligned}
p(N_{e^-}) &= \sum_{\substack{N_{e^-}^{NR} \\ N_{e^-}=N_{e^-}^{NR}+N_{e^-}^{ER}}} \sum_{N_{e^-}^{ER}} \sum_{E_{nr}} \sum_{E_{er}} p(N_{e^-}^{NR}, N_{e^-}^{ER}|E_{nr}, E_{er})\, p(E_{nr}, E_{er}) \\
&= \sum_{N_{e^-}^{NR}=0}^{N_{e^-}} \sum_{E_{nr}} \sum_{E_{er}} p(N_{e^-}^{NR}, N_{e^-} - N_{e^-}^{NR}|E_{nr}, E_{er})\, p(E_{nr}, E_{er}) \qquad (6.13)\\
&= \sum_{N_{e^-}^{NR}=0}^{N_{e^-}} \sum_{E_{nr}} \sum_{E_{er}} p(N_{e^-}^{NR}|E_{nr}) p(N_{e^-} - N_{e^-}^{NR}|E_{er})\, p(E_{nr}, E_{er}),
\end{aligned}
$$

where we didn't report the dependence on $\boldsymbol{\theta}_{cal}$ for shortness of notation. In the above equation, we can identify the three probability factors in the following way:

- $p(N_{e^-}^{NR} = N_{e^-}^k | E_{nr} = x_j^{th})$ is equal to $\mathcal{M}_{k,j}^{1,NR}$ - see Eq. 6.9;

- $p(N_{e^-}^{ER} = N_{e^-}^i - N_{e^-}^k | E_{er} = x_l^{th})$ is equal to $\mathcal{M}_{i-k,l}^{1,ER}$ - see Eq. 6.9;

- $p(E_{nr} = x_j^{th}, E_{er} = x_l^{th})$ is equal to the double differential Migdal rate normalized to unity.

If we explicate the indices, we obtain

$$
\begin{aligned}
p(N_{e^-} = N_{e^-}^i) &= \sum_{k=0}^{i} \sum_{j} \sum_{l} \mathcal{M}_{k,j}^{1,NR} \mathcal{M}_{i-k,l}^{1,ER} \left.\frac{d^2\tilde{R}_M}{dE_{er}dE_{nr}}\right|_{j,l} \\
&= \sum_{k=0}^{i} \sum_{j} \sum_{l} \mathcal{M}_{k,j}^{1,NR} \left.\frac{d^2\tilde{R}_M}{dE_{er}dE_{nr}}\right|_{j,l} \left(\mathcal{M}^{1,ER}\right)_{l,i-k}^{T} \qquad (6.14)\\
&= \sum_{k=0}^{i} \left(\mathcal{M}^{1,NR} \cdot \frac{d^2\tilde{R}_M}{dE_{er}dE_{nr}} \cdot \left(\mathcal{M}^{1,ER}\right)^T\right)_{k,i-k},
\end{aligned}
$$

where $\tilde{R}$ indicates the normalized double differential Migdal rate, namely it satisfies the condition

$$
\sum_{j} \sum_{l} \left.\frac{d^2\tilde{R}_M}{dE_{er}dE_{nr}}\right|_{j,l} = 1. \qquad (6.15)
$$

The Migdal expected spectrum after recombination is simply obtained by substituting the normalized double differential Migdal rate with the original one, namely

$$
y_i^{exp,Mig} = \sum_{k=0}^{i} \left(\mathcal{M}^{1,NR} \cdot \frac{d^2 R_M}{dE_{er}dE_{er}} \cdot \left(\mathcal{M}^{1,ER}\right)^T\right)_{k,i-k}. \qquad (6.16)
$$

This operation is a matrix-by-matrix product, and can be efficiently implemented on a GPU, as we did for our analysis.

6.2.3 Implementation of the screening systematic uncertainty and the Q-value systematic uncertainties

The low energy systematic σ_{scr} for the β spectra of ^{39}Ar and ^{85}Kr is implemented as a Gaussian uncertainty on the theoretical spectra of ^{39}Ar and ^{85}Kr that ranges linearly from 0% at $E_{er} = 200$ eV to 10% at $E_{er} = 0$ eV[2]. In other words, it is implemented, for example for ^{39}Ar, as

$$y_{th}^{\mathrm{Ar}}(E_{er}) = \begin{cases} y_{th}^{\mathrm{Ar},(0)}(E_{er}) & \text{if} \quad E_{er} \geq 200 \text{ eV} \\ \left[1 + \left(\frac{10}{100} - \frac{1}{200}\frac{E_{er}}{10\,\mathrm{eV}}\right)\sigma_{scr}^{\mathrm{Ar}}\right] y_{th}^{\mathrm{Ar},(0)}(E_{er}) & \text{if} \quad E_{er} < 200 \text{ eV} \end{cases}$$

$$(6.17)$$

where $y_{th}^{\mathrm{Ar},(0)}(E_{er})$ is the "central" theoretical spectrum and σ_{scr} is distributed according to a standard normal *pdf*.

The Q-value uncertainties are implemented as two Gaussian parameters regulating the ^{39}Ar and ^{85}Kr spectra as in the previous case, with an intensity dependent on the energy. For example, for example for ^{39}Ar, it is implemented as

$$y_{th}^{\mathrm{Ar}}(E_{er}) = \left[1 + r(E_{er})\,\sigma_Q^{\mathrm{Ar}}\right] y_{th}^{\mathrm{Ar},(0)}(E_{er})$$

$$(6.18)$$

where $r(E_{er})$ is the relative uncertainty due to the uncertainty on the Q-value, that depends on the ER energy, and σ_Q^{Ar} is distributed according to a standard normal *pdf*.

6.2.4 Detector effects: efficiency, radial corrections, PMT response

As outlined in Sec. 5.3.1, the two main efficiency effects that play a sizable role in our measurement are the so-called radial efficiency and the electron lifetime efficiency. The former effect can be parameterized as an efficiency factor depending only on the radial position of the event. The latter is instead an efficiency effect due to the electron lifetime inside the TPC, namely the probability that an ionization electron is captured by electro-negative impurities in the gas chamber. This is a small effect that depends mainly on the distance between the event and the liquid-gas interface, and then depends on the event depth inside the TPC.

From the implementation point of view, since we are dealing with energy spectra, without the information on the position of the events, which has been integrated out, one can proceed in the following way. The original theoretical spectra with the energy and 3D position information are simulated in the Monte Carlo. Therefore, by means of the Monte Carlo, we have also access to the position and the related total efficiency factor of each event. We can therefore use this information, to numerically compute the *pdf* to have a certain efficiency ϵ_{src}^{ch} in a given event. This *pdf* will depend on the PMT channel and on the background/signal source spatial features. The result of this procedure will be a set of numerical functions (one per source and PMT channel) over which we will integrate to get the full convoluted

---

[2]Private communication with X. Mougeot, author of `BetaShape`

effect. These *pdf*s depend only on the theoretical spectra, and do not depend on the calibration parameters. They can be therefore computed once before the final MCMC integration, and used as an input for the analysis.

An additional effect due to the PMT channel corrections has to be taken into account. In fact, in the data spectrum, the contributions from the various PMT channels are not simply summed up: the PMT response has been in fact equalized to the central PMT, and therefore, when implementing the detector response model to be compared with data, we need to apply the same correction. This correction is a simple multiplicative factor r_{ch} in the PMT energy response that depends on the PMT channel.

Finally, the PMT resolution effect can be treated as a Gaussian smearing effect, similar to the ER Fano fluctuations of Eq. 6.6. In other terms, including also the efficiency effects, we can express the probability that an event from the source *src* in the PMT channel *ch* is detected to have $N_{e^-}^f = i$ extracted electrons having originally produced $N_{e^-}^{(0)} = j$ electrons surviving recombination as

$$
P(N_{e^-}^f = i | N_{e^-}^{(0)} = j, ch, src) =
$$
$$
\int_i^{i+b_w} dx \int d\epsilon_{src}^{ch} \mathcal{N}\left(x | \mu = \epsilon_{src}^{ch} r_{ch} j, \sigma = S\sqrt{\epsilon_{src}^{ch} r_{ch} j} \right) p(\epsilon_{src}^{ch}),
$$
$$(6.19)$$

where: b_w is the width of the bins of the final observed N_{e^-} histogram, ϵ_{src}^{ch} is the overall efficiency factor and $p(\epsilon_{src}^{ch})$ is its *pdf* as computed by means of the Monte Carlo; $\mathcal{N}(x|\mu, \sigma)$ is the normal *pdf* with mean μ and variance σ; r_{ch} is the channel correction factor; $S = 0.27$ is a width factor that has been determined during the calibration. As already done for the binomial/Fano fluctuation smearing $\mathcal{M}^1$ matrices, we can define a set of matrices

$$
\mathcal{M}_{i,j}^2(ch, src) = P(N_{e^-}^f = i | N_{e^-}^{(0)} = j, ch, src) \tag{6.20}
$$

such that we can compute the final expected N_{e^-} spectrum $S_{ch,src}^f$ in the PMT channel *ch* induced by the source *src* as

$$
S_{ch,src,i}^f(\boldsymbol{\theta}_{cal}) = \sum_{j=1}^{N_{max}^{exp}} \mathcal{M}_{i,j}^2(ch, src) S_{ch,j}^{N_{e^-}}(\boldsymbol{\theta}_{cal}). \tag{6.21}
$$

The most important feature of the $\mathcal{M}^2$ matrices is that they depend only on the channel corrections, the width factor S and, by means of the efficiency *pdf*, from the theoretical spectra: these quantities do not depend on the calibration parameters $\boldsymbol{\theta}_{cal}$, and therefore they can be computed once before the final MCMC integration, and used as an input for the analysis. This is a great property in terms of performance, because this means that, unlike the two $\mathcal{M}^1$ matrices, they do not need to be computed at each step of the MCMC algorithm, but they are just input numbers.

To conclude, once we have $S_{ch,src}^f$, we can obtain the final expected spectrum

$S^f_{ER,NR}$ by summing over all the PMT channels and over all the sources, namely

$$S^f_{NR,i}(\boldsymbol{\theta}_{cal}) = \sum_{N_{ch}} S^f_{ch,NR}(\boldsymbol{\theta}_{cal}),$$

$$S^f_{ER,i}(\boldsymbol{\theta}_{cal}) = \sum_{N_{ch}} \sum_{N_{src}} S^f_{ch,src,ER}(\boldsymbol{\theta}_{cal}). \tag{6.22}$$

In terms of Bayesian network, the above portion of the response model can be reduced to
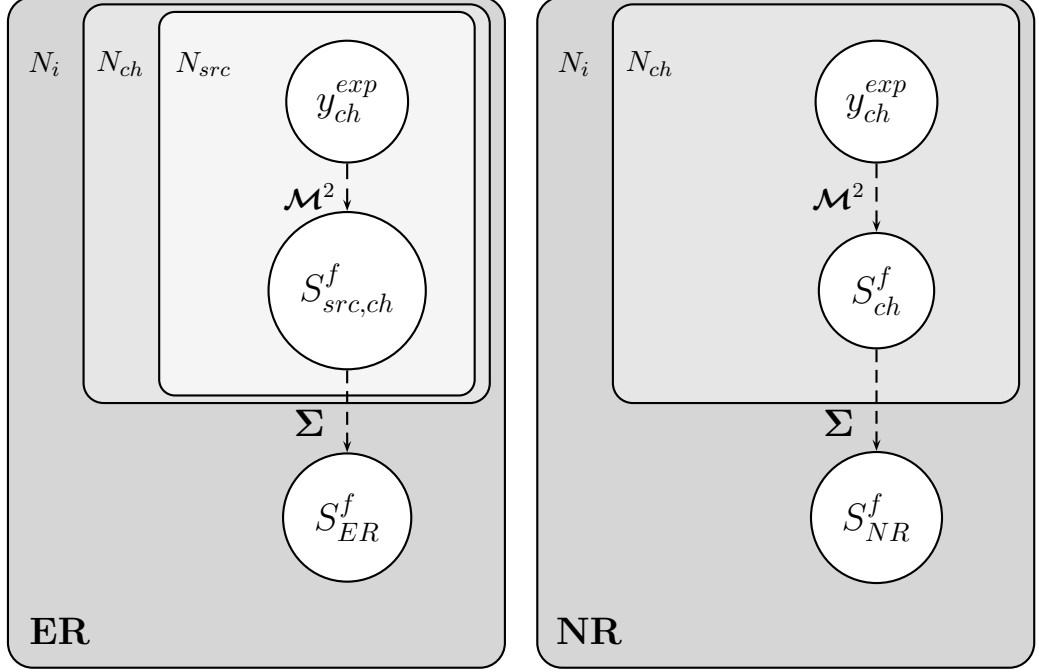
where $\boldsymbol{\Sigma}$ is a symbol representing the sum over all the channels and over all the background sources.

6.2.5 Validation of the response

We validated our response model on the implementation of the so-called `lowmass` code (used for the calibration and for the frequentist approach of the 2018 analysis [1], and based on a toy MC simulation of the detector response), in turn validated on the calibration measurements.

The results of the comparison are shown in Fig. 6.2,6.3. For the NR part, since the NR signal is concentrated in the low-N_{e^-} region, we also considered the AmC spectrum, which lives in a wider N_{e^-} region. For the ER side, we used, as initial input, the same PMT and cryostat background components of the `lowmass` code. The statistical fluctuations in the `lowmass` spectra are produced by the MC-like procedure intrinsic in the `lowmass` approach.

We also performed a crosscheck of the detector response to the Migdal effect, which depends at the same time on the NR and ER detector response. The results are perfectly consistent between the two approaches, as shown in Fig. 6.4 for a DM mass of $m_{DM} = 0.356 \, \mathrm{GeV}/c^2$.
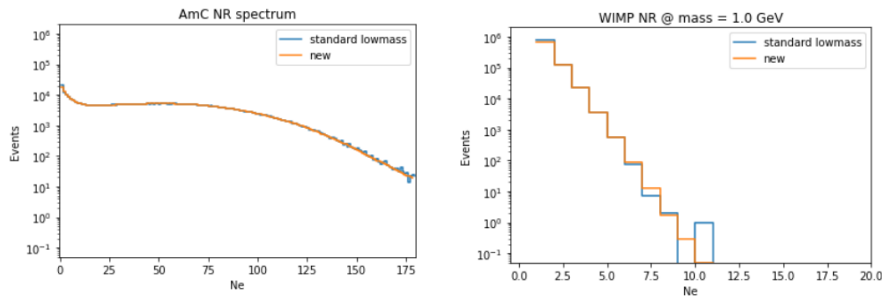
Figure 6.2. Validation of the NR response. The inputs of the two computations (the calibration parameter values and the theoretical spectra) are the same. The spectra are normalized to the total MC events used for the computation of the theoretical spectra.
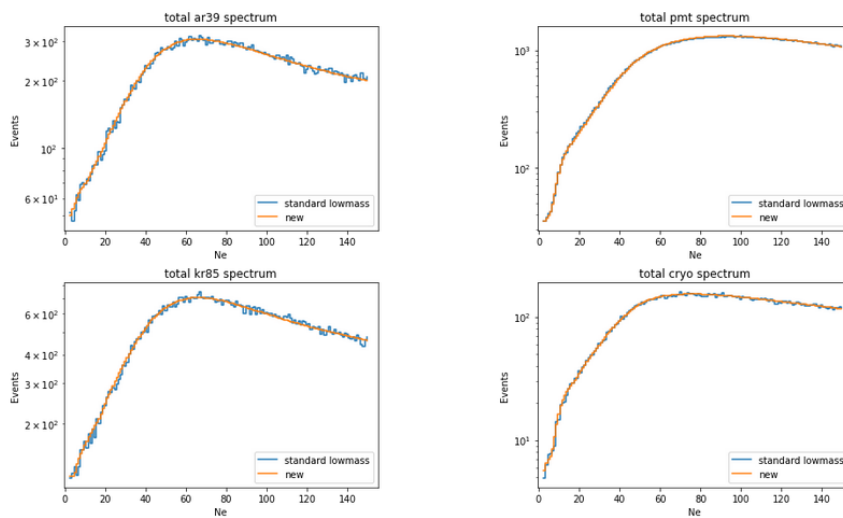

Figure 6.3. Validation of the ER response. The inputs of the two computations (the calibration parameter values and the theoretical spectra) are the same. The ^{39}Ar and ^{85}Kr spectra are normalized to the total MC events used for the computation of the theoretical spectra, while the PMT and the cryostat total backgrounds are instead normalized to their respective activities.
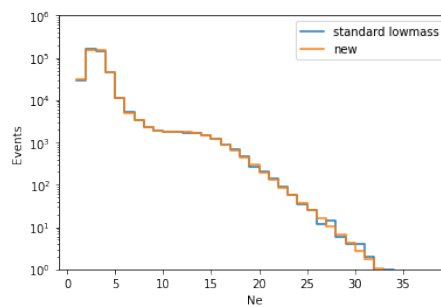

Figure 6.4. Validation of the response for the Migdal effect for a DM mass of $m_{DM} = 0.356\,\mathrm{GeV}/c^2$. The inputs of the two computations (the calibration parameter values and the theoretical spectra) are the same. The spectra are normalized to the total MC events used for the computation of the theoretical spectra.

6.2.6 Impact of the detector response uncertainties on the template spectra

In Sec. 6.2.4 we showed how the signal and background models can be expressed as explicit functions of the detector response parameters $\boldsymbol{\theta}_{cal}$. This is given in Eq. 6.22. Such a result is particularly relevant because it allows us to compute for any specific detector response configuration (identified by a choice of the parameters $\boldsymbol{\theta}_{cal}$) the associated background and signal spectra. In this section, we give some examples on how the spectra change as a function of $\boldsymbol{\theta}_{cal}$ and show how our approach is substantially different from constructing average spectra with the corresponding '1-sigma' variation. We limit ourselves to the background model, which depends only on the ER response model.

Figure 6.5 represents the joint *pdf* of the 4 parameters controlling the response to a ER energy deposit. Figure 6.6 shows the effect of sampling from the *pdf* shown in Fig. 6.5 onto the four relevant background spectra; the left column gives the total effect, while the right column the remaining effect after normalizing the spectra to the same area. The intensity of the gray shaded area gives an idea about how often that specific region of the spectrum is generated by different detector response models. The green spectrum is obtained with the expected value of $\boldsymbol{\theta}_{cal}$, i.e. the average detector response. At any given value of N_e we can construct the *pdf* of the expected number of events and compute the expected value and the ± 1 sigma intervals. From these values we can construct the average spectrum, given in orange in the figure, and the ± 1 sigma. We should stress here that, given the non-linearity of the problem, in general the envelope spectra are different from those computed with the corresponding choice of the $\boldsymbol{\theta}_{cal}$ parameters. This is illustrated in Fig. 6.6 for the average envelope spectra (orange) and the spectra computed with the average $\boldsymbol{\theta}_{cal}$ parameters (green). Some differences, especially in the low N_e region of the spectra, are clearly visible. In addition, the envelope spectrum completely loses the bin-bin correlations and produces a spectrum that does not correspond to any of the spectra predicted by the detector response model. Finally, we stress that the method presented here naturally takes into account the correlations between the $\boldsymbol{\theta}_{cal}$ parameters (no assumption on $\boldsymbol{\theta}_{cal}$ is needed on how to produce the '1-sigma' spectrum as their joint *pdf* is included in the method), moreover the correlation between different spectra induced by the $\boldsymbol{\theta}_{cal}$ parameters is also properly taken into account.

However, to have an idea of the impact of the different systematic uncertainties used as input to the fit on the final expected spectra, we constructed the envelope of the total background spectra and the two dominant background contributions - i.e. the ^{85}Kr and the PMT spectra. The result of this study is depicted in Fig. 6.7. On the right-hand side of the figure, we report the relative uncertainties for the different systematic contributions, as well as the statistical uncertainty.

The effect of the uncertainties of the $\boldsymbol{\theta}_{cal}$ parameters on the signal spectra is shown in Fig. 6.8 for DM masses of 0.356, 1.08, and 4.52 GeV/c^2.

The implementation in the signal and background models of the systematic variation of the detector response described in this section is the key element for the correct uncertainty propagation in the fit results. As it is described in the next sections, the fit samples from the space of possible detector response models,
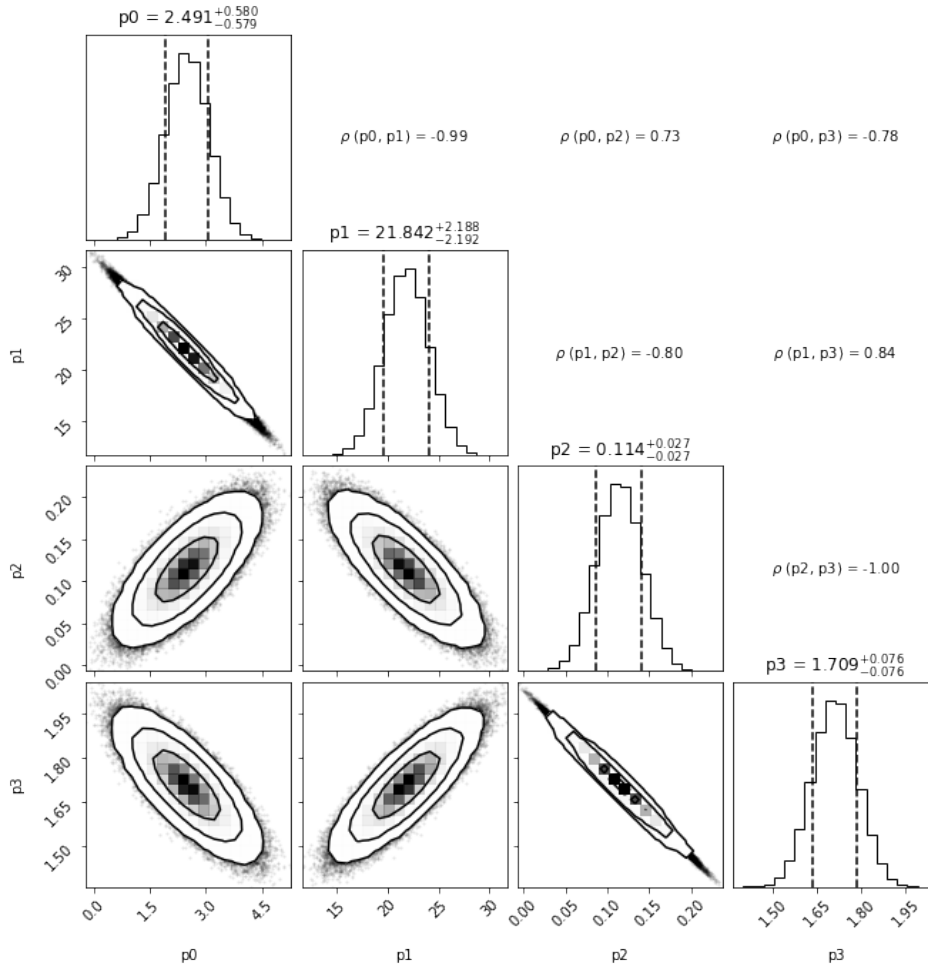
Figure 6.5. Graphical representation of the joint *pdf* of the ER response model $\boldsymbol{\theta}_{cal}$. With respect to the parameterization given in Sec. 5.3.2, the calibration parameters of this plot are defined as: $p_0 = (200 \text{ V cm}^{-1})\rho/C_{box}$, $p_1 = (200 \text{ V cm}^{-1})/C_{box}$, $p_2 = p^0$ and $p_3 = p^1$. The plots on the diagonal of the figure are the uni-dimensional *pdf* of each single parameter obtained by marginalizing on all the others. The bi-dimensional *pdf*s in the bottom-left corner of the figure give the joint *pdf*s of each pair of parameters obtained by marginalizing on the others. The plots show also the credible regions at 68%, 95%, 99.7% probability as solid contour lines. The correlation coefficients are given in the upper-right corner of the figure.

computes the corresponding spectra both for the signal and backgrounds, tries to fit the observed spectrum, and finally weights this result by the infinitesimal probability of the specific response model sampled. In such a way, one constructs the *pdf* for the model that fits the observed spectra, and from such a *pdf* can compute the average model and the corresponding uncertainty.

6.3 Likelihood and parameters

Let x_i denote the number of events in the data spectrum in the bin $N_{e^-} = i$. We assumed a bin-by-bin Poisson likelihood defined as

$$p(\{x_i\}|\boldsymbol{\theta}) = \prod_i \mathcal{P}\left(\{x_i\}|\lambda_i(\boldsymbol{\theta})\right) \equiv \prod_i \frac{\lambda_i(\boldsymbol{\theta})^{x_i}}{x_i!} e^{-\lambda_i(\boldsymbol{\theta})}, \tag{6.23}$$

where $\boldsymbol{\theta}$ indicates generically all the parameters of the fit - i.e. systematic uncertainties, signal rate, background rate, and so on. In the particular case of a fit with a DM signal, for instance, $\lambda_i(\boldsymbol{\theta})$ is given by

$$\begin{aligned}
\lambda_i = E \Big[r_{B,\text{Ar}} \quad & S_i^{\text{Ar}}(\boldsymbol{\theta}_{cal}, \sigma_{scr}^{\text{Ar}}, \sigma_Q^{\text{Ar}}) + r_{B,\text{Kr}} \quad S_i^{\text{Kr}}(\boldsymbol{\theta}_{cal}, \sigma_{scr}^{\text{Kr}}, \sigma_Q^{\text{Kr}}) + \\
& + r_{B,\text{PMT}} \quad S_i^{\text{PMT}}(\boldsymbol{\theta}_{cal}) + r_{B,\text{cryo}} \quad S_i^{\text{cryo}}(\boldsymbol{\theta}_{cal}) + \\
& + r_S \quad \left(S_i^{\text{NR}}(\boldsymbol{\theta}_{cal}) + + S_i^{\text{Mig}}(\boldsymbol{\theta}_{cal}) \right) \Big],
\end{aligned} \tag{6.24}$$

where:

- S_i^{src} represent the expected background and signal spectra as a result of the detector response - see Eq. 6.22.

- $r_{B,src}$ are proportional to the rate of the internal and external background components. Since, for simplicity, the background spectra are normalized to the DarkSide-50 exposure, they are normalized such that $r_{B,src} = 1$ corresponds to the case in which the exposure is equal to the DarkSide-50 exposure.

- r_S is proportional to the signal rate. Also in this case, the signal spectra are normalized to the DarkSide-50 exposure and computed for $\sigma_{SI}^{DM} = 10^{-38}\,\text{cm}^2$. Therefore, $r_S = 1$ corresponds to $\sigma_{SI}^{DM} = 10^{-38}\,\text{cm}^2$.

- E is the total exposure. Due to the choice of normalization of the expected spectra, $E = 1$ corresponds to the case in which the exposure is equal to the DarkSide-50 exposure.

- $\sigma_{scr}^{\text{Ar,Kr}}$ are the low energy β spectra systematic parameters, affecting the ^{39}Ar and the ^{85}Kr background spectra.

- $\sigma_Q^{\text{Ar,Kr}}$ are the low energy β spectra Q-value systematic parameters, affecting the ^{39}Ar and the ^{85}Kr background spectra.

- $\boldsymbol{\theta}_{cal}$ represents the set of calibration parameters. Just to fix the notation with respect to Sec. 5.3.2, 5.3.3, they are
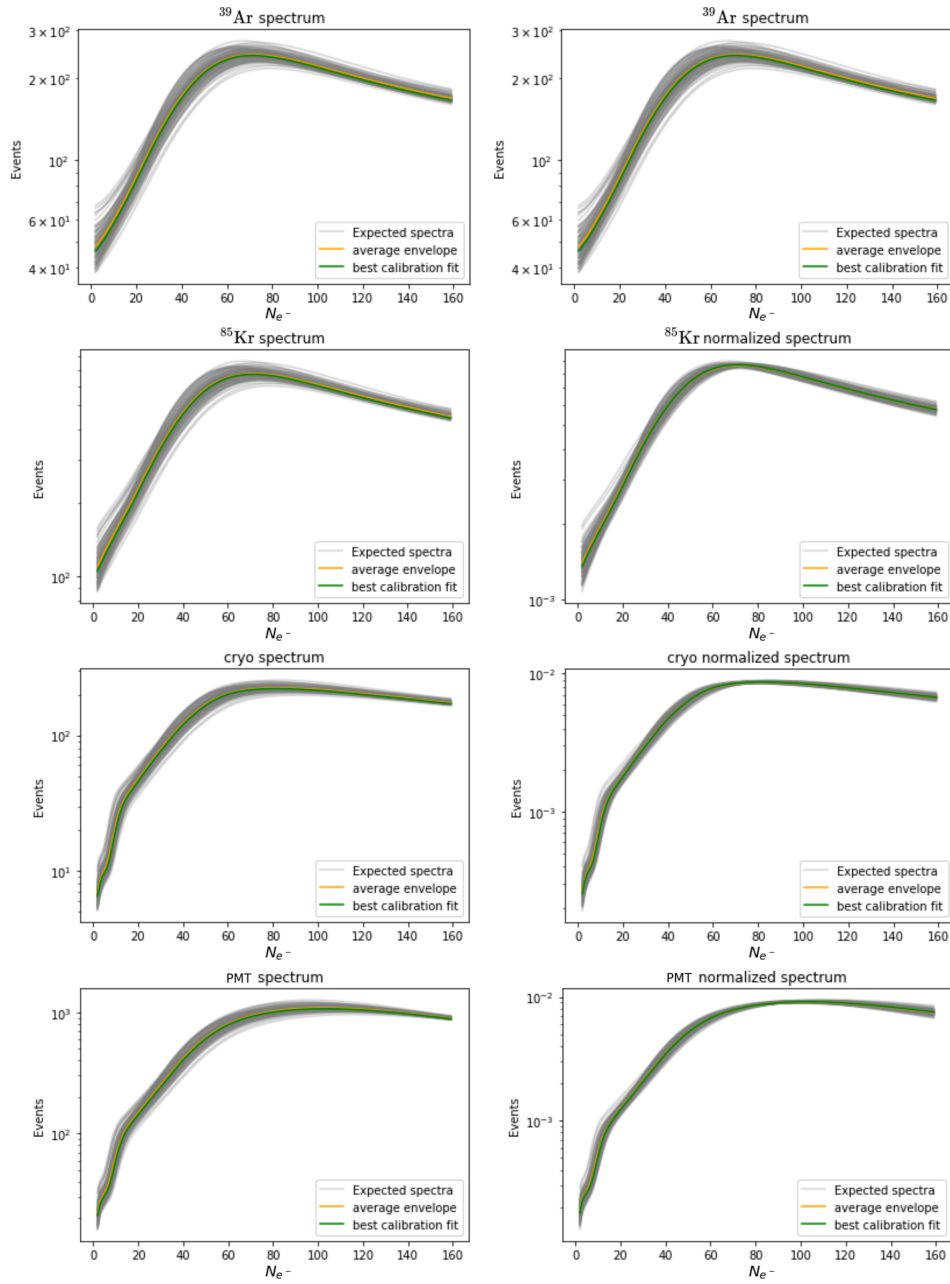
Figure 6.6. Background spectra generated by sampling the joint p.d.f of the $\boldsymbol{\theta}_{cal}$ parameters. The spectra on the right panels represent the same spectra given on the left panels but are normalized to the same area.
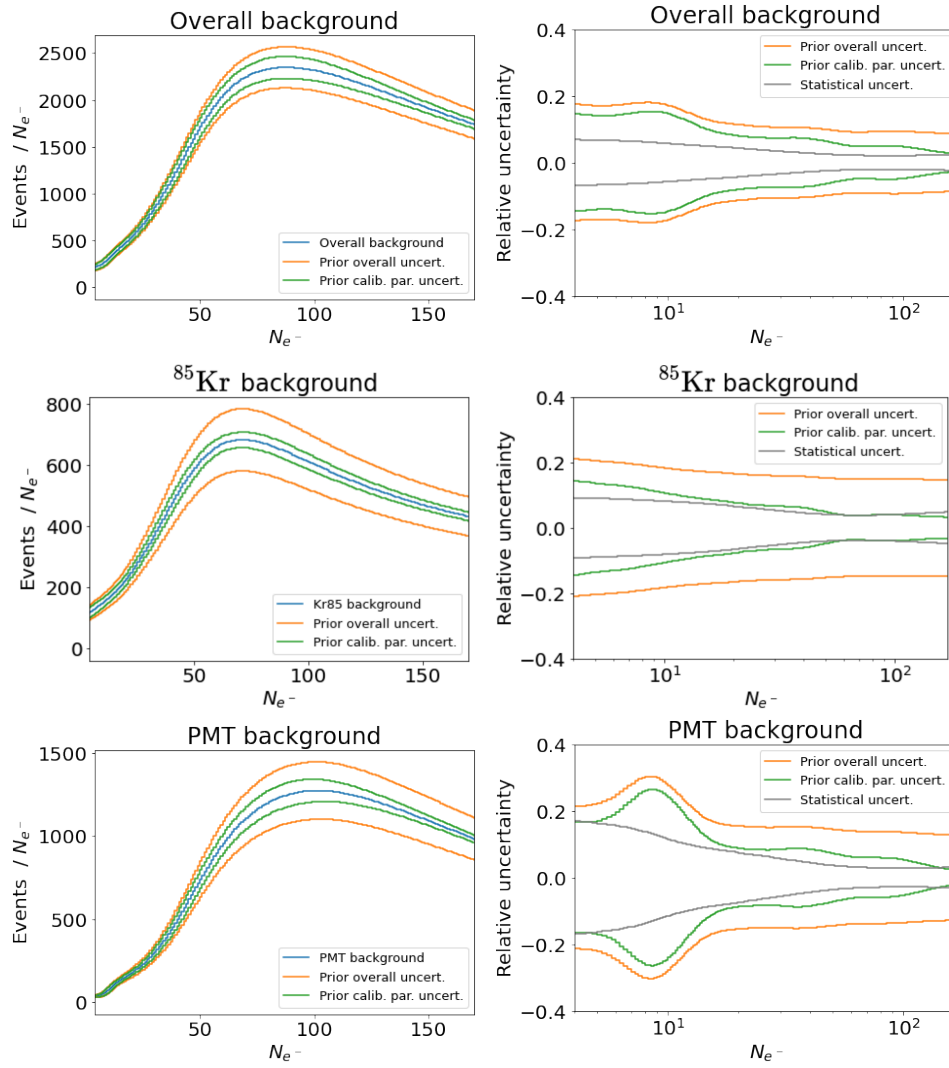
Figure 6.7. Impact of the different systematic uncertainties used as input to the fit on the final expected spectra. The green lines correspond to the propagation of the prior uncertainty on the calibration parameters to these spectra. The orange lines are instead the propagation of the prior uncertainty from all the parameters of the likelihood. The gray lines are the Poisson statistical uncertainties associated, bin by bin, to the correspondent exposure of the final expected background spectra.
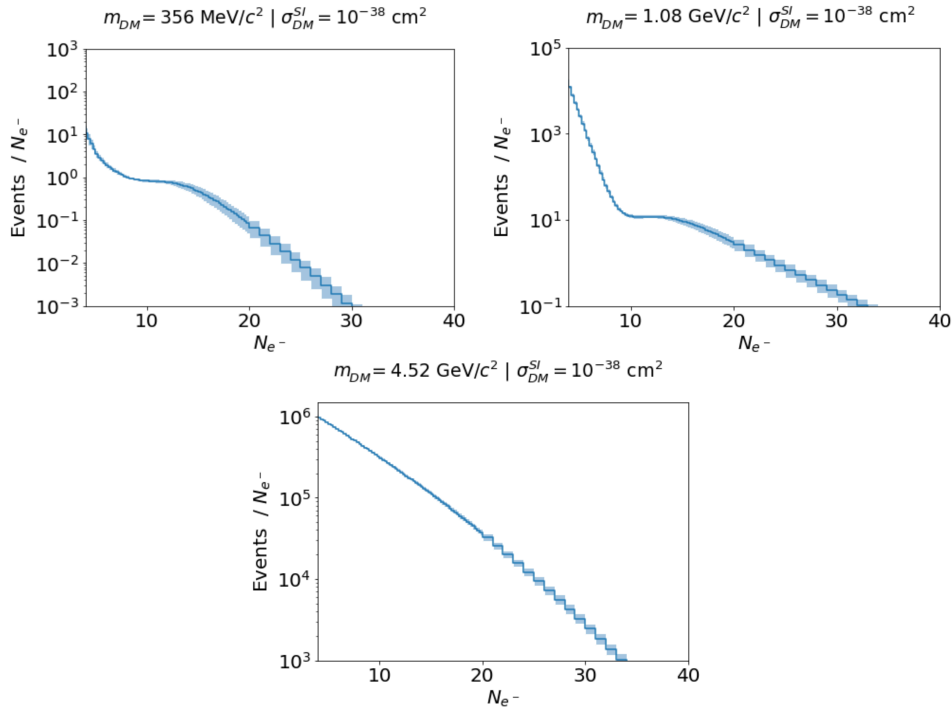
Figure 6.8. Signal spectra for DM masses of respectively 0.356, 1.08, 4.52 GeV/c^2 generated by sampling the joint p.d.f of the $\boldsymbol{\theta}_{cal}$ parameters.

- for the ER response: $p_0 = (200 \, \text{V cm}^{-1})\rho/C_{box}$, $p_1 = (200 \, \text{V cm}^{-1})/C_{box}$, $p_2 = p^0$ and $p_3 = p^1$;
- for the NR response: C_{box}^{NR}, $f_B = \beta \times 10^{-4}$.

The corresponding summarizing network can be drawn as:
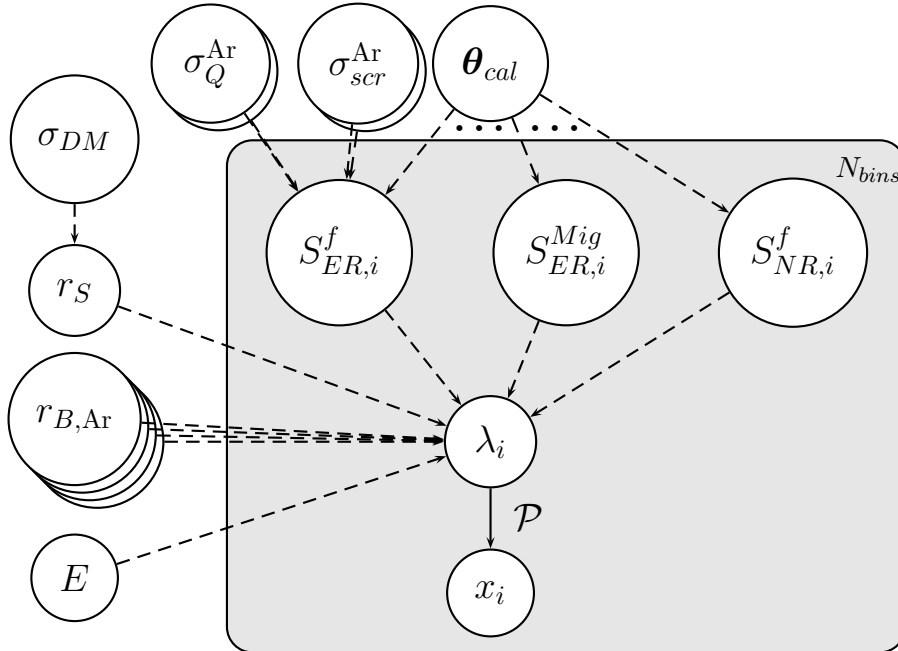
where the dots represent the detector response network and σ_{DM} is the DM-nucleon cross-section that can be computed given the signal rate r_S.

The result of the calibration measurements of [144], is implemented as prior information on the calibration parameters. To be more specific, for the following results we assumed the following priors on the parameters

- $r_{B,src}$ have an uncorrelated normal prior centered in 1 and with a 14% (^{39}Ar), 4.7% (^{39}Kr), 12.6% (PMT) and 6.6% (cryostat) uncertainty;

- r_S has a flat prior[3];

- E has a normal prior centered in 1 and with a 1.5% uncertainty[4];

- $\sigma_{scr}^{\text{Ar,Kr}}$ and $\sigma_Q^{\text{Ar,Kr}}$ have a standard normal prior;

- for the ER response, as a result of the calibration measurements, see Fig. 5.7, 5.11, the parameters are distributed according to a multivariate normal prior centered in

$$(p_0, p_1, p_2, p_3) = (2.49,\ 21.8,\ 0.114,\ 1.709), \tag{6.25}$$

  with a covariance matrix equal to

$$\Sigma_{ER} = \begin{pmatrix} 0.339 & -1.27 & 0.0116 & -0.0350 \\ -1.27 & 4.85 & -0.0478 & 0.142 \\ 0.0116 & -0.0478 & 0.000737 & -0.00208 \\ -0.0350 & 0.142 & -0.00208 & 0.00591 \end{pmatrix} \tag{6.26}$$

- for the NR response, as a result of the calibration measurements, the parameter are distributed according to a multivariate normal prior centered in $C_{box}^{NR} = 8.05\ \text{V cm}^{-1}$ and $f_B = 0.67$, with a covariance matrix equal to

$$\Sigma_{NR} = \begin{pmatrix} 0.0225 & 0.00217 \\ 0.00217 & 0.000400 \end{pmatrix} \tag{6.27}$$

The limits on the DM cross-section have been computed following the prescription of Sec. 4.1.1. Fig. 6.9 shows the posterior on r_S as a result of a fit on a pseudo-dataset generated from the background template, assuming a DM mass $m_{DM} = 4.52\ \text{GeV}/c^2$.

6.4 Technical description of the fitting procedure

The posterior of all the parameters is sampled by means of the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm, as implemented in `BAT` [116]. The `BAT` package is a set of `C++` libraries implementing statistical tools for Bayesian analyses, and it has been largely used in the HEP fields by many collaborations,

---

[3]For a detailed discussion on the importance of this choice see Sec. 4.1.2, where we also argued that a flat prior is a good reference for comparison purposes.

[4]This uncertainty, determined by means of a dedicated MC simulation, is related to the thermal contraction of the PTFE defining the cylindrical volume of the TPC.
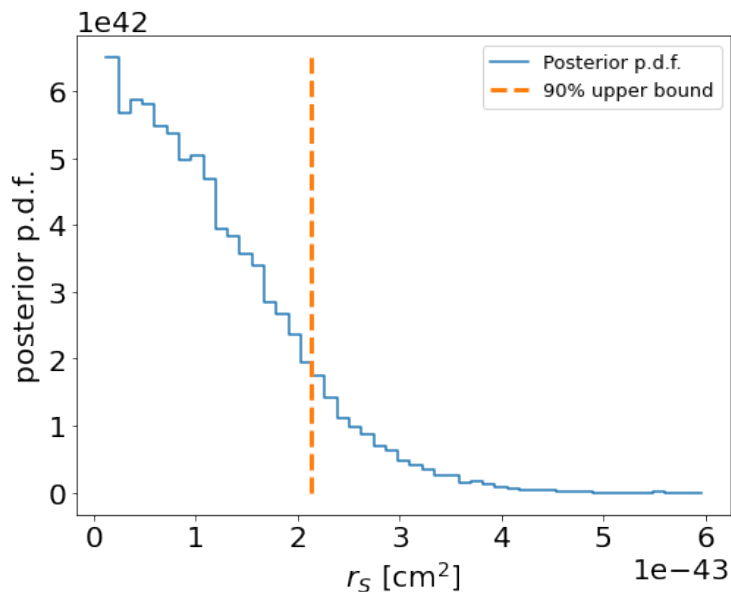
Figure 6.9. Posterior *pdf* on r_S (blue line) of a fit on the pseudo-dataset generated from the background template assuming a DM mass $m_{DM} = 4.52\,\mathrm{GeV}/c^2$. The orange vertical line represents the 90% C.I. as defined in Sec. 4.1.1.

such as UTfit [161] or HEPfit [162]. Since the Mathematics behind each single step of the chain - i.e. how the calibration parameters are connected to the final spectra - is mostly linear algebra, the code is also almost completely GPU integrated by means of the `CUDA` libraries [159].

In all the following analyses, the sampling is performed by means of 12 parallel MCMC chains, for a total number of steps equal to 1.2×10^6. This number of MCMC steps is large enough to sample the posteriors in a complete way and will allow us to obtain the results discussed in the next section. In Fig. 6.10 we show the typical behavior of the chains using as an example the parameters r_B^{Kr} and r_B^{PMT}. As we just mentioned, the convergence is obtained by means of a pre-run phase, which is already implemented in `BAT` and which, by tuning the Metropolis-Hastings MCMC parameters, assures that all the parallel chains converge to the same region of the parameters' phase space with an optimal MCMC rejection rate.

From the implementation point of view, the background model or background-plus-signal model are implemented as an extension of the `BAT BCModel` class. The likelihood and the priors are customized following the prescriptions of the previous sections, and the linear algebra implemented in CUDA. The inputs of the fit, i.e. the theoretical spectra and the $\mathcal{M}^2$ smearing matrices, are read at the constructor level. The GPU preliminary operations are also performed once inside the constructor, to optimize the sampling procedure.

6.5 Results

In order to avoid possible biases coming from using data, we performed a blind analysis based on a pseudo-dataset. This dataset has been generated from the
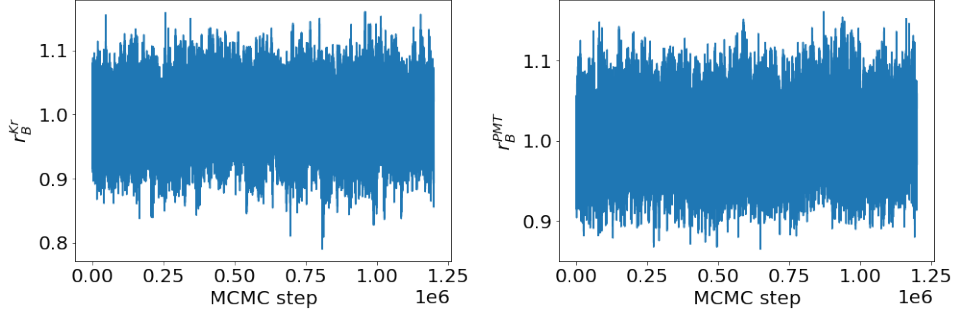
Figure 6.10. Evolution of the MCMC chains for the parameters r_B^{Kr} (**left**) and r_B^{PMT} (**right**) as a function of the MCMC step. The behavior is regular throughout the evolution. The two plots have been obtained running 12 parallel chains, for a total number of steps equal to 1.2×10^6.

expected background template, the so-called Asimov dataset, which is plotted in Fig. 6.11. If not specified, all the fits are performed in the full $N_{e^-} = [4, 170]$ range. Below $N_{e^-} = 4$, the data are indeed dominated by correlated events, mostly due to the contamination of spurious electrons. On the other hand, the upper threshold is chosen as the maximum point at which the detector response calibrations have been validated, namely $N_{e^-} = 170$.

The binning of the observed spectrum is denser (0.25 N_{e^-} binning) in the $N_{e^-} < 20$ region with respect to the higher N_{e^-} region (1 N_{e^-} binning): since the DM signal, which is expected to contribute in the low N_{e^-} region, is exponentially falling, having a denser binning in the signal region could improve the sensitivity. This has been verified to have a beneficial 5-10% impact on the sensitivity in the whole mass region explored by our analysis.

6.5.1 Impact of the systematic uncertainties on the limit

As a first study, we investigated the impact of the systematic uncertainties on the final limit. For this study, we focused on a single DM mass value $m_{DM} = 4.52\mathrm{GeV}/c^2$, in which the Migdal effect contribution to the signal spectrum is negligible. We arranged the nuisance parameters in the following 3 groups

- group CAL: p_0, p_1, p_2, p_3, C_{box}^{NR}, f_B, which represent the calibration parameters;

- group TH: $\sigma_{scr}^{\mathrm{Ar,Kr}}$, $\sigma_Q^{\mathrm{Ar,Kr}}$, which represent systematic uncertainties on the theoretical background spectra;

- group NORM: E, $r_{B,src}$, which represent the systematic uncertainties on the normalization of the background spectra.

We therefore performed 6 fits, with the following assumptions:

1. leaving free all the systematic parameters;

2. leaving free only the group CAL and fixing all the other parameters to their expected values;
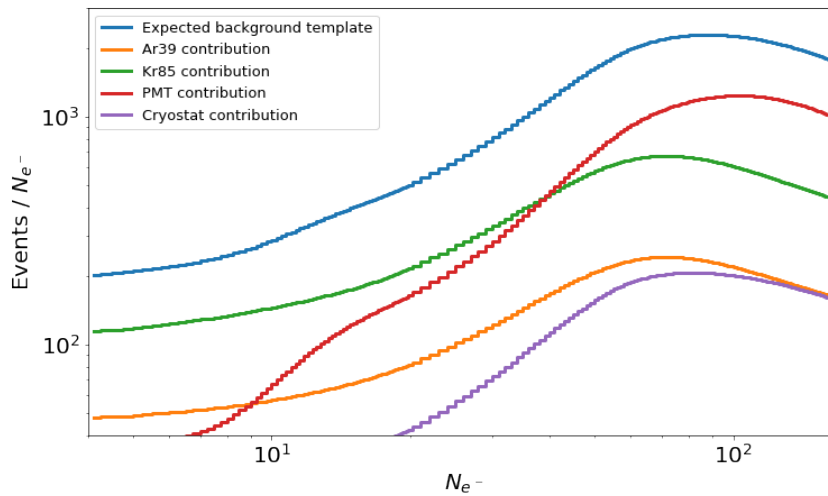
Figure 6.11. The so-called Asimov dataset, namely the expected background (blue line) in the $N_{e^-} = [4, 170]$ range. The different colored lines represent the expected contributions from the different background sources. The spectra have a 0.25 N_{e^-} binning for $N_{e^-} < 20$ and a 1 N_{e^-} elsewhere, see text.

3. leaving free only the group NORM+TH and fixing all the other parameters to their expected values;

4. leaving free only the group TH and fixing all the other parameters to their expected values;

5. leaving free only the group NORM and fixing all the other parameters to their expected values;

6. fixing all the systematic parameters to their expected values. This corresponds to a statistic-only fit, since the only active parameter is r_S.

The results in terms of the sensitivity to the DM cross-section are given in Tab. 6.3. The table shows also the comparison with the standard `lowmass` approach, used also for the 2018 analysis. The most relevant differences between our innovative approach and the standard one are two. First of all the standard approach computes the sensitivity as the frequentist 90% C.L., while here we perform the analysis in the Bayesian approach, and we quote the 90% quantile of the posterior *pdf* of the signal strength parameter r_S [27]. Even if the two quantities aim at expressing the experimental sensitivity, they are conceptually different, and are defined in totally different ways; however, when operating with the same inputs they should give comparable results. In addition, when we look at the posterior of r_S, we are integrating over all the nuisance parameter space, while the `lowmass` approached is based on the profiling of the likelihood: if the likelihood is Gaussian the marginalization and the profiling give the same result, but in the general case in which the Gaussian assumption is not valid, the profiling procedure typically returns underestimated propagated uncertainties [163]. On the other hand, in the standard approach the systematic uncertainties on the calibration parameters regulating the

Active syst.	σ_{SI}^{DM} [90% C.I.] [cm^2]	σ_{SI}^{DM} [90% C.L.] [cm^2] (`lowmass`)
All	2.11×10^{-43}	2.22×10^{-43}
CAL	1.80×10^{-43}	1.61×10^{-43}
NORM + TH	1.71×10^{-43}	2.16×10^{-43}
TH	1.61×10^{-43}	1.81×10^{-43}
NORM	1.47×10^{-43}	1.65×10^{-43}
None	1.29×10^{-43}	1.53×10^{-43}

Table 6.3. Upper bound results on σ_{SI}^{DM} for $m_{DM} = 4.52 \,\text{GeV}/c^2$ in our new approach and in the standard `lowmass` one using the expected pseudo-dataset of the two procedures fixing different groups of parameters, see text. We keep the third significant digit for the internal comparison.

detector response are treated in a very different way: we showed that in our approach we are able to apply the detector response to the theoretical spectra in an analytical way, allowing us to propagate in a solid way the uncertainties on the calibration parameters to the final result; the `lowmass` approach, instead, relies on a re-parameterization of the detector response using a single parameter responsible for changing the shape of the spectra in a rigid way, and we already elaborated on the possible biases that this approach could induce in Sec. 6.2.6.

The statistic-only fit, namely the last row of Tab. 6.3, shows a $\sim 20\%$ difference between the two approaches. This difference can be attributed to the combination of two effects: the difference between the frequentist C.L. and the Bayesian 90% quantile on the posterior of r_S, and small differences between the Asimov datasets of the two approaches[5].

If we look at the systematic parameters which are not in the CAL group, their impact in the new vs `lowmass` approach with respect to the statistic-only fit is comparable: 14% vs 10% for the NORM group, 24% vs 21% for the TH group, and 33% vs 44% for the NORM+TH group. What is noticeably different is the impact of the calibration parameters, which worsen the sensitivity by 40% according to our new approach, and by just 7% in the standard approach. As a result, when we put together all the systematic parameters, as reported in the first line of Tab. 6.3, we obtain comparable results in terms of expected sensitivity, even if in our case the degradation of the expected sensitivity due to the systematic uncertainties is larger (60%) with respect to what is found by the `lowmass` approach (48%).

As a final remark, we found that the impact of the systematic uncertainties on the calibration parameters has the same size as the impact of all the other systematic parameters, including the parameters regulating the normalization of the background and the theoretical uncertainties on the input spectra, and therefore the correct propagation of their uncertainty is relevant in the final result.

For completeness, we report in Fig. 6.12 the full multidimensional posterior *pdf*s of the complete fit, in which all the systematic parameters are not fixed.

---

[5]We showed how in our approach the final expected spectra are computed in a semi-analytical way, while the `lowmass` approach relies on a toy MC procedure to generate the spectra that induces small fluctuations of $\sim 1\%$ in the Asimov.
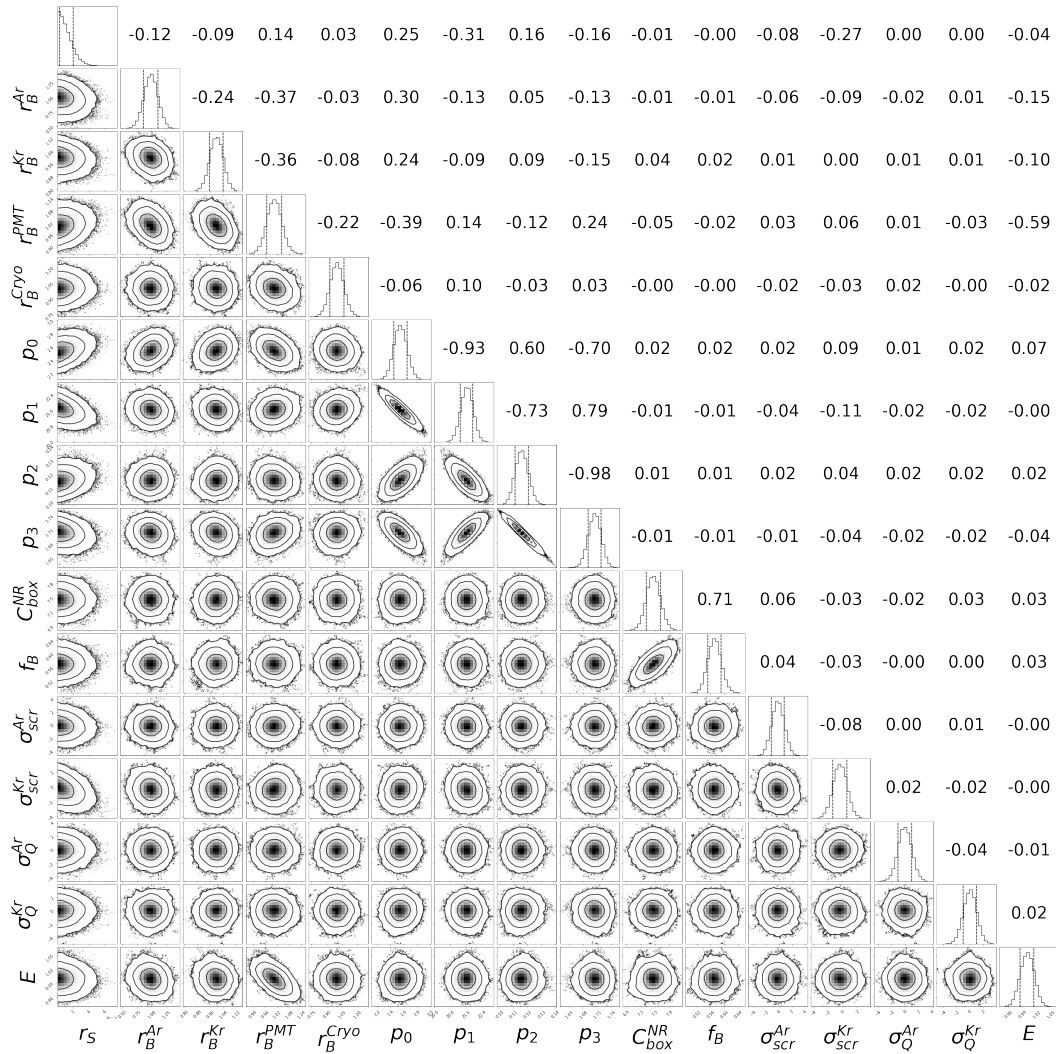
Figure 6.12. Graphical representation of the joint posterior *pdf* of the fit on the expected pseudo-dataset. The plots on the diagonal of the figure are the uni-dimensional *pdf* of each single parameter obtained by marginalizing on all the others. The bi-dimensional *pdf*s in the bottom-left corner of the figure give the joint *pdf*s of each pair of parameters obtained by marginalizing on the others. The plots show also the credible regions at 68%, 95%, 99.7% probability as solid contour lines. The correlation coefficients are given in the upper-right corner of the figure.

6.5.2 Impact of the high N_{e^-} data points on the limit and the calibration parameters

In this subsection we show the impact of choosing different N_{e^-} windows on the calibration parameters' posterior *pdf*s, and, as a consequence, on the upper limit on the DM cross-section. We want to remind that the priors on the calibration parameters correspond to the constraints from the calibration: therefore, the differences between priors and posteriors tell us that the data give additional information with respect to the calibration. In general, we expect that having more knowledge on the calibration parameters allows obtaining a limit less affected by the systematic uncertainties - see the "None" result of Tab. 6.3. In order to investigate this effect, we decided to perform a fit to the expected pseudo-dataset in the regions $N_{e^-} = [4, 30]$ and $N_{e^-} = [4, 170]$. The $N_{e^-} = [4, 170]$ region has a long high N_{e^-} tail and can provide better constraints on the calibration parameters.

In Fig. 6.13 we show the results of two background plus signal fits on the pseudo-dataset performed in the regions $N_{e^-} = [4, 30]$ and $N_{e^-} = [4, 170]$ (orange and blue shaded histograms) with respect to the prior *pdf* (gray dashed line), assuming a DM mass $m_{DM} = 4.52\,\mathrm{GeV}/c^2$.

The figure shows how, in general, the width of the calibration parameters' posteriors are smaller in the $N_{e^-} = [4, 170]$ case rather than both in the prior and the $N_{e^-} = [4, 30]$ cases. It is for this reason that in the $N_{e^-} = [4, 170]$ case we obtain the best sensitivity - i.e. σ_{SI}^{DM} [90% C.I.] $= 2.1 \times 10^{-43}$ cm^2 - while in the $N_{e^-} = [4, 30]$ case we obtain a milder bound - i.e. σ_{SI}^{DM} [90% C.I.] $= 2.5 \times 10^{-43}$ cm^2.

As an additional crosscheck, we verified that in the $N_{e^-} = [4, 30]$ case and in the $N_{e^-} = [4, 170]$ case the posterior *pdf* of $r_{B,\mathrm{PMT}}$ is different, being narrower in the $N_{e^-} = [4, 170]$ case since in this N_{e^-} window the dominant background is the PMT one - see Fig. 6.11. On the other hand, the posterior *pdf*s of $r_{B,\mathrm{Ar}}$, $r_{B,\mathrm{Kr}}$, $r_{B,\mathrm{cryo}}$, and E are pretty similar in the two cases. In other words, the $N_{e^-} = [30, 170]$ region gives additional information mostly on $r_{B,\mathrm{PMT}}$, rather than on $r_{B,\mathrm{Ar}}$, $r_{B,\mathrm{Kr}}$, $r_{B,\mathrm{cryo}}$, and E. Finally, as expected, we do not observe differences between the posterior *pdf* of $\sigma_{scr}^{\mathrm{Ar,Kr}}$ and $\sigma_{Q}^{\mathrm{Ar,Kr}}$: they regulate the background spectra of Ar and Kr, which give sensible contribution to the overall background only in the first bins of the spectrum - see Fig. 6.11.

For the calibration parameters, we notice that:

- the posterior *pdf*s of the ER group are more constrained than the prior *pdf*s, roughly by a factor of 2-3. This behavior does not depend on the data structure, since this study is performed on Asimov pseudo-dataset, representing the expected background pseudo-dataset.

- the posterior *pdf*s of the NR group are basically equal to their priors. This is the opposite of the ER case: since in the pseudo-dataset there is no injected signal and since the NR signal is relevant in the first few bins of the data spectrum, the data do not give additional information on the calibration parameters with respect to the calibration itself.

In summary, in this approach, we complement the calibration results by constraining some of its parameters and thus reducing their uncertainties by a factor

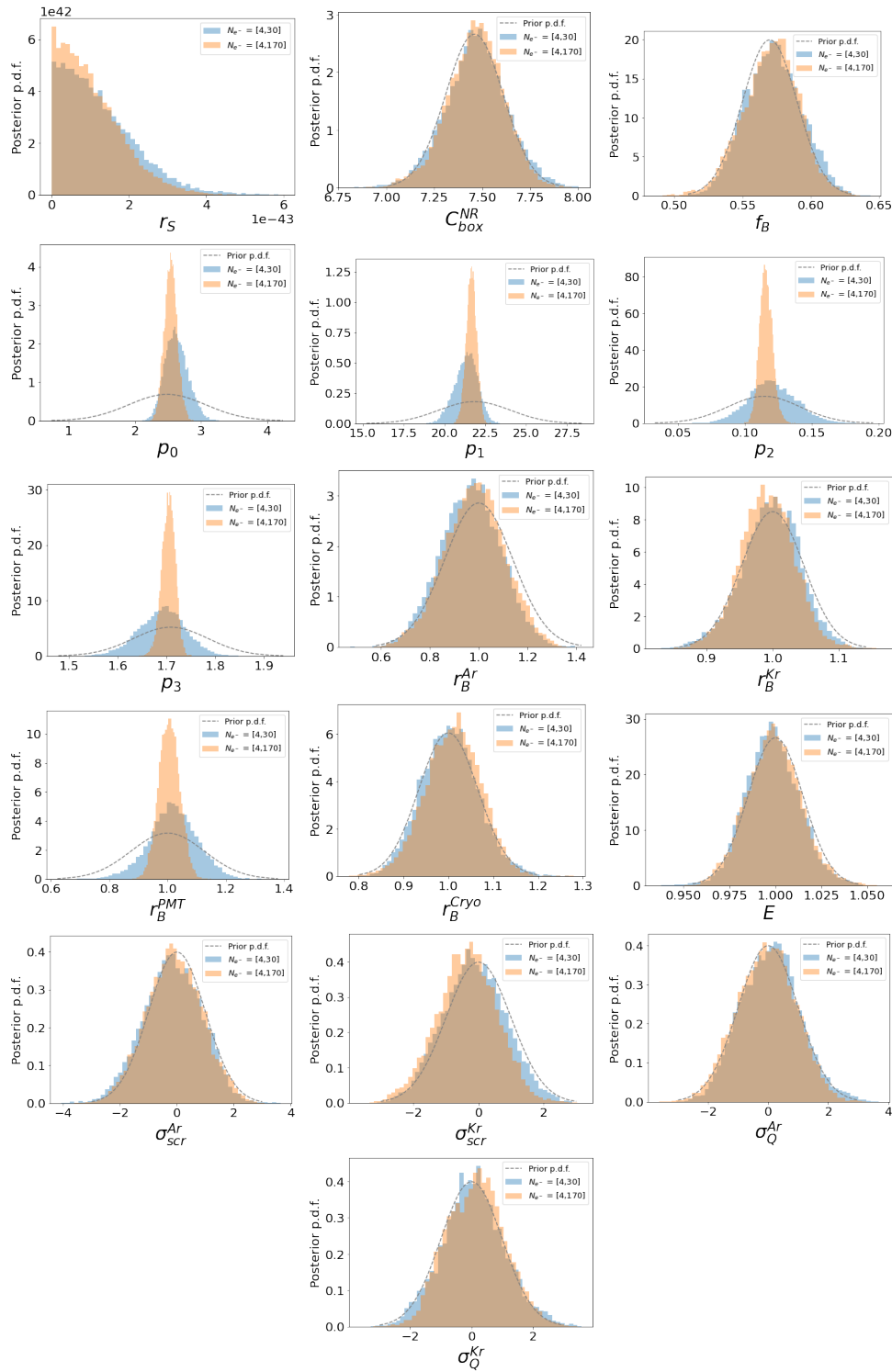**Figure 6.13.** Posterior *pdf*s obtained by fitting in different energy ranges a pseudo-dataset with the full background plus signal model assuming a DM mass $m_{DM} = 4.52 \, \text{GeV}/c^2$. The blue shaded histograms correspond to the $N_{e^-} = [4, 30]$ window, the orange shaded histograms correspond to the $N_{e^-} = [4, 170]$ window, and the gray dashed line are the prior *pdf* of each parameter.

of ~ 2. We can directly benefit from this improvement by having a stronger bound on the signal and being able to justify it in terms of a better knowledge of the calibration parameters.

6.5.3 Fit parameters *pdf*s as a function of the DM mass

Figure 6.14 shows the dependence of the parameters' mean and standard uncertainty of the posterior *pdf*s as a function of the DM mass m_{DM} (blue points); for reference, we also report the mean and standard uncertainty of the correspondent prior *pdf*s (gray solid and dashed lines, respectively). None of the parameters shows a strong behavior with the DM mass, demonstrating the stability of the fitting procedure.

6.5.4 Expected sensitivity of DarkSide-50

Figure 6.15 shows the 90% C.I. expected sensitivity on the DM cross-section σ_{SI}^{DM} as a function of the DM mass (orange solid line). This line is compared with the two most competitive exclusion limits, defined as frequentist 90% C.L., in the GeV and sub-GeV DM mass region, i.e. the XENON1T Migdal search [90] (black line), the XENON1T NR search [164] (black dashed-dot line), and the CRESST-III result [64] (gray line).

We would like to stress here the fact we are neglecting the theoretical systematic uncertainties on the Migdal signal contribution from the argon $n = 3$ valence shell. As we mentioned above, the theoretical uncertainty on the computation of the ionization probabilities on isolated atoms, as pointed out in Ref. [93], are negligible. However, we are not taking into account the uncertainty due to the fact the Ar atoms in the TPC are in a liquid state, and they are not isolated. We demonstrated in Ref. [2] that a possible $\mathcal{O}(100\%)$ uncertainty on the valence shell Migdal contribution would affect the sensitivity of the experiment in a noticeable way only for DM masses $\lesssim 200\,\mathrm{MeV}/c^2$: therefore, since this uncertainty is expected to affect mainly the $n = 3$ valence shell, we believe that our result is solid against possible theoretical uncertainties on the Migdal signal contribution in the $\gtrsim 200\,\mathrm{MeV}/c^2$ DM mass region. In the $\lesssim 200\,\mathrm{MeV}/c^2$ region the sensitivity would be weakened by such theoretical systematic uncertainties, as we showed in Ref. [2].

6.5.5 Fit on a pseudo-dataset with signal injection

In order to study the response of the fit in the case in which a signal is present inside the data, we performed a fit on a pseudo-dataset containing a signal injection. As usual, we performed this analysis at the $m_{DM} = 4.52\,\mathrm{GeV}/c^2$ DM mass point. We injected signal events in the Asimov pseudo-dataset corresponding to a DM interaction cross-section three times larger than the expected sensitivity, namely we injected a signal with $\sigma_{DM}^{SI} = 6.3 \times 10^{-43}\mathrm{cm}^2$. The result in terms of posterior *pdf* on r_S is shown in Fig. 6.16: the injected signal cross-section is perfectly reconstructed as $(6.2 \pm 1.5) \times 10^{-43}\,\mathrm{cm}^2$, which is ~ 4 standard deviations away from $r_S = 0$. The uncertainty on the reconstructed value is, as expected, of the same order of magnitude as the expected sensitivity.

Figure 6.14. Evolution of the posterior *pdf*s as a function of the DM mass. The blue points represent the mean of the posterior *pdf* and their uncertainty bar is the correspondent standard deviation. In each plot, the gray solid line is the mean of the prior *pdf*, while the two gray dashed lines correspond to the standard deviation of the prior *pdf*.
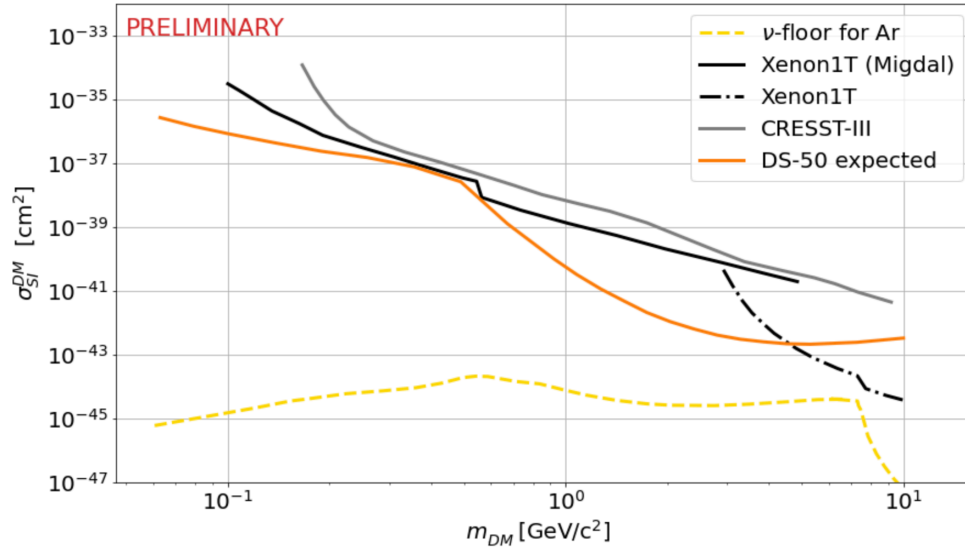
Figure 6.15. The 90% C.I. expected sensitivity on the DM cross-section σ_{SI}^{DM} as a function of the DM mass (orange solid line) with the inclusion of the Migdal effect. The black line is the frequentist 90% C.L. from the XENON1T Migdal search [90], the black dashed-dot line is the is the one from the XENON1T NR search [164], and the gray line is the one by the CRESST-III experiment [64]. The yellow dashed line represents the neutrino floor for an argon target.
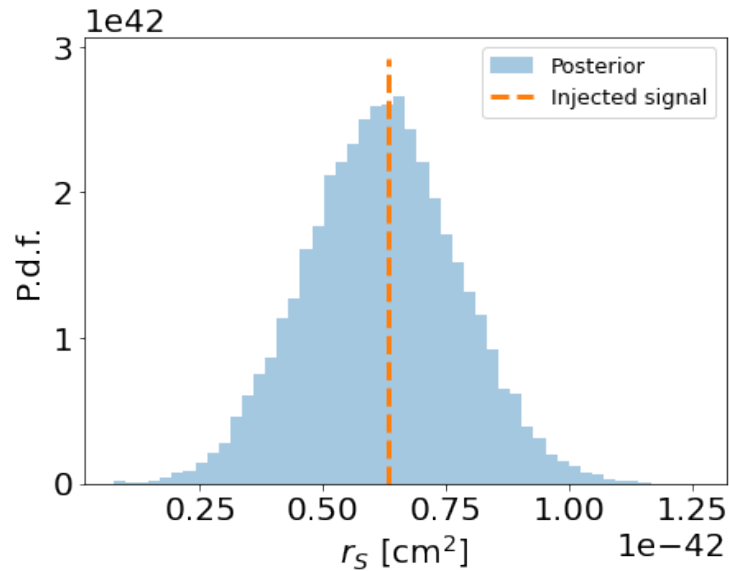

Figure 6.16. Posterior *pdf* of r_S (blue shaded histogram) obtained from a fit on a signal-injected Asimov pseudo-dataset. The injected value of the signal is represented by the orange dashed vertical line, and is totally in agreement with the result of the fit.

6.5.6   Impact of the nuclear recoil quenching fluctuations

In Sec. 6.2.1 we decided to model the intrinsic fluctuations of the number of ion-electron pairs produced as a consequence of a NR interaction by means of a binomial distribution. However, there has been no possibility to constraint this model via calibration at such low energies [144]. Therefore, the choice of using binomial fluctuation has been performed because we believe that this model is a reasonable model to describe this statistical effect, among all the other possibilities we investigated.

In order to better justify this assumption, let's describe in more details this model. The binomial model encapsulates three combined statistical processes:

1. the fluctuations in the number of visible quanta, namely the amount of energy released in the visible channel;

2. the fluctuations in the amount of ion-electron pairs given the number of produced visible quanta;

3. the fluctuations in the amount of non-recombined ion-electron pairs, given the number of the originally produced ion-electron pairs.

The last two steps of this list can be reasonably represented by a binomial process, in which the probability parameters are given by the probability p_i for a visible quanta to be an ion-electron pair (step n. 2) and the probability $p_R = 1 - f$ of not recombining (step n.3). The quantity p_R is given by the Thomas-Imel model (see Eq. 5.21), while the quantity p_i can be expressed as

$$p_i = \frac{N_i}{N_V}, \tag{6.28}$$

where N_i is the average number of ion-electron pairs produced after the nuclear recoil, and N_V is the average number of visible quanta produced after the nuclear recoil. In turn, N_V can be expressed as

$$N_V = \frac{E}{w}\mathcal{L}(E_{nr}), \tag{6.29}$$

where $\mathcal{L}(E_{NR})$ is the Lindhard factor [132] which models the average fraction of energy released in the visible channels in LAr. Finally, the fluctuations in the number of visible quanta (step n. 1), the so-called "quenching fluctuations", have been modeled as a binomial process, where the probability of producing a visible quanta is given by the Lindhard factor $p_V = \mathcal{L}(E_{nr})$. The Bayesian network describing this statistical model is the following:
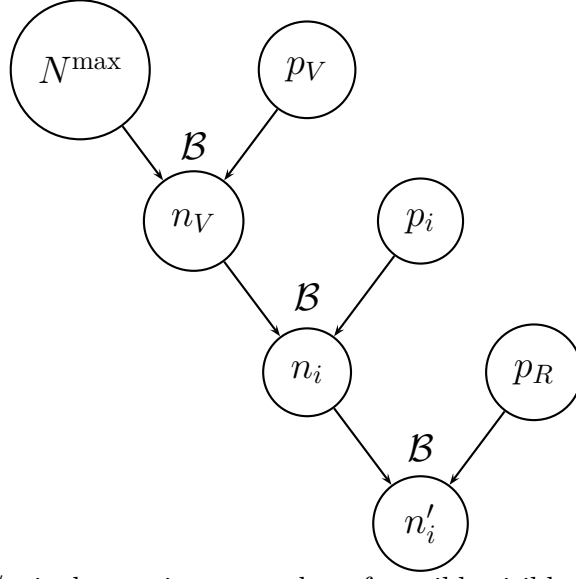
where: $N^{\max} = E_{nr}/w$ is the maximum number of possible visible quanta; n_V is the number of produced visible quanta; n_i is the number of produced ion-electron pairs suddenly after the nuclear recoil; n_i' is the number of ion-electron pairs surviving recombination. The combination of the three binomial processes result in an overall binomial with the parameters described in Sec. 6.2.1.

The above model, and in particular the binomial quenching fluctuation part, i.e. the top part of the above Bayesian network, can not be confirmed in the low energy regime, below ~ 2 keV, by any available calibration measurements. To convince ourselves of the validity of this model, we developed very simplistic MC simulations of the quenching fluctuation process at low energies. We modeled the energy lost into heat during the scattering of the recoiling atoms both assuming a continuous energy loss and a discrete energy loss (e.g. to simulate phonons excitation), see Fig. 6.17 for more details on the two models. We found in both cases that size of the fluctuations are comparable with that of the binomial distribution, see Fig. 6.18. These results are also confirmed by more detailed simulations of the micro-physics of the atomic scattering in LAr, based on SRIM [165], which show quite large fluctuations in the energy quenching, as shown in Fig. 4 of Ref. [166].

In any case, to have an idea of the mass region in which the impact of this choice is relevant, we performed an analysis - always on the Asimov expected pseudo-dataset - in which we completely switched off the NR quenching fluctuations. In our formalism, this is equivalent to impose

$$\mathcal{M}_{i,j}^{1,NR} = P(N_{e^-}^{NR} = N_{e^-}^i | E_{nr} = x_j^{th}, \boldsymbol{\theta}_{\text{cal}}) = \begin{cases} 1 & \text{if } \lfloor E_{nr} Q_y^{NR}(E_{nr}, \boldsymbol{\theta}_{cal}) \rfloor = N_{e^-}^i \\ 0 & \text{elsewhere.} \end{cases}$$
$$(6.30)$$

The results are reported in Fig. 6.19. The plot shows that, as expected, in the no-quenching (NQ) configuration the sensitivity is degraded in the $[0.4, 2]$ GeV/c^2 DM mass region: indeed, this is the region where the pure NR signal component goes, in the NQ case, under the $N_{e^-} = 4$ threshold. With the NQ hypothesis, in the $[0.4, 1.4]$ GeV/c^2 the sensitivity is weaker than the XENON1T Migdal [90] result
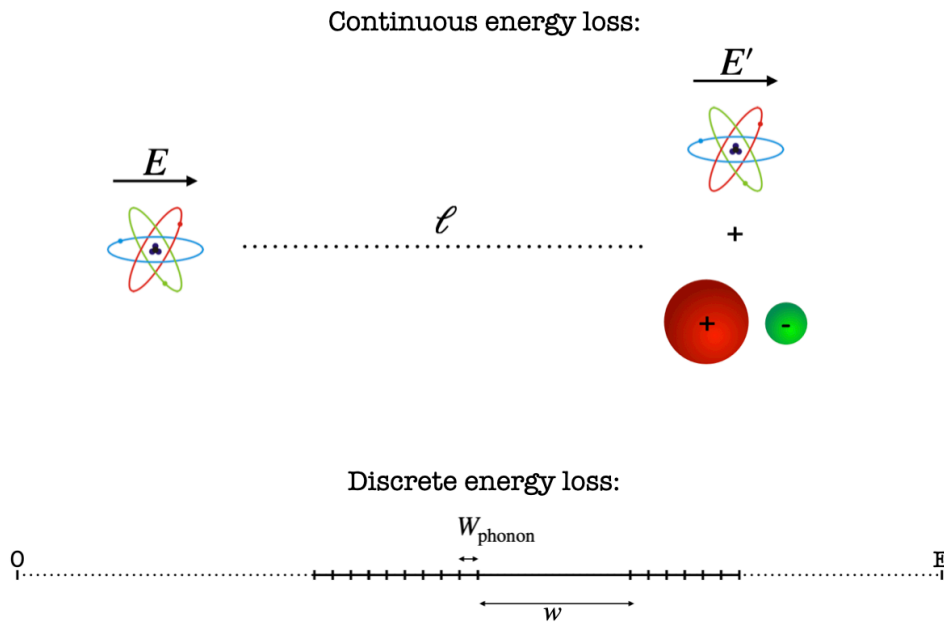
Figure 6.17. Top: drawing representing the model of the MC simulation for a continuous energy loss. The recoiling Ar atom with an initial kinetic energy E travels for a length ℓ in the gas before inducing the creation of a visible quantum, losing an amount of energy $e_{\text{heat}} \propto \ell$ via heat. After the production of a visible quantum the atom has an energy $E' = E - e_{\text{heat}} - w$. In the MC, the quantity ℓ is sampled at each step from an exponential distribution. **Bottom:** drawing representing the model of the MC simulation for a discrete energy loss. In this scheme, the energy release process is implemented as the combination of a high number of interactions depositing either a small amount of energy W_{phonon} into heat or an amount w of energy for the creation of a visible quantum. For the results reported in this work we assumed $W_{\text{phonon}} \sim w/100$, but we explored also different possibilities with greater or smaller values for W_{phonon} observing no remarkable differences in the results. The stochastic process related to the probability of releasing energy either in the "phonon" or in the visible channel is implemented as a Bernoulli process with a probability $p = N_{\text{Vis}}/(N_{\text{Vis}} + N_{\text{phon}})$, where N_{Vis} is the average number of visibile quanta and N_{phon} is the average number of "heat quanta".
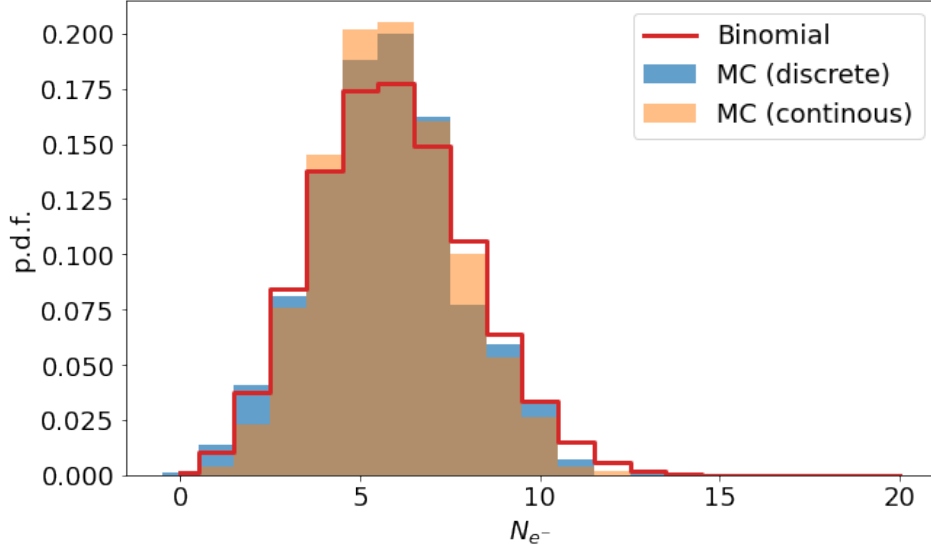
Figure 6.18. Quenching fluctuations assuming a NR with energy $E_{nr} = 640\mathrm{eV}$ (equivalent
in average to $N_{e^-} = 4$) for two MC-simulated models (coloured histograms) and the
binomial fluctuations model (red line). The two models assume either a continuous
energy loss (orange histogram) or a discrete energy loss (blue histogram).
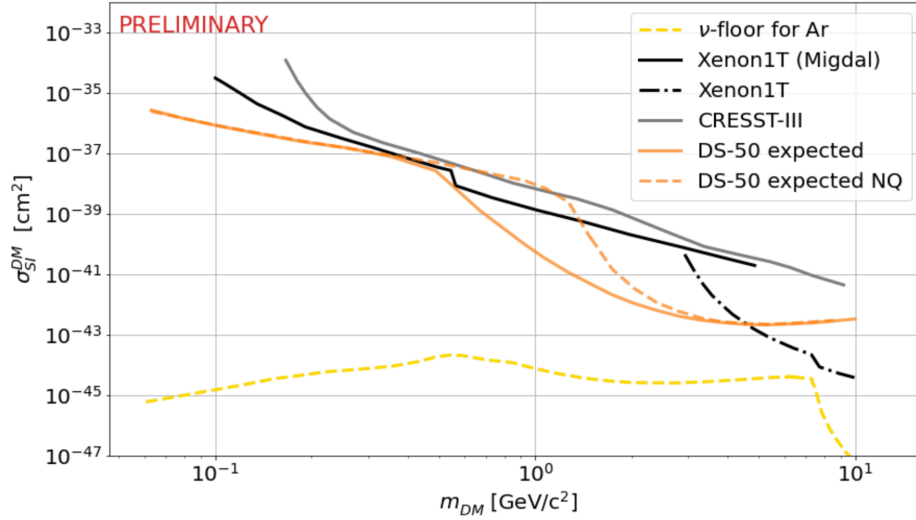
Figure 6.19. The 90%C.I. expected sensitivity on the DM cross-section σ_{SI}^{DM} as a function
of the DM mass with the inclusion of the Migdal effect, with (orange solid line) and
without (orange dashed lines) the NR quenching fluctuations. The black line is the
frequentist 90% C.L. from the XENON1T Migdal search [90], the black dashed-dot line
is the is the one from the XENON1T NR search [164], and the gray line is the one by
the CRESST-III experiment [64]. The yellow dashed line represents the neutrino floor
for an argon target.

(which includes the NR quenching fluctuations). Below $0.4\,\mathrm{GeV}/c^2$ the sensitivity with the binomial model is again the same as obtained with the NQ model, because, both with and without the quenching fluctuations, the Migdal effect contribution is always over the $N_{e^-} = 4$ threshold.

We would like to stress that the NQ model is completely non-physical: the production of ion-electron pairs after a NR in LAr is indeed a statistical process, and we believe that some model for these fluctuations should be anyway adopted.

6.5.7 Background-only fit on data

Now that we have under control the fitting procedure, we can look at the real DarkSide-50 dataset. In this sense, the first step was performing a background-only fit on the data: the best fit (top), the correspondent normalized residuals (middle), and their distribution (bottom) are reported in Fig. 6.20. As we can see from the figure, the background model describes the observed data without noticeable deviations or excesses. The behavior of the residual is satisfactory and their distribution over all the data points is consistent with the expectation, namely a standard normal *pdf*.

In terms of posterior *pdf* on the fit parameters, no significant deviation from the prior distribution - namely the results from the calibration measurements - is observed. For the sake of completeness, in Fig. 6.21 we report the full multidimensional posterior of the fit.

6.5.8 Observed sensitivity of the DarkSide-50 experiment

The last step is now to compute the observed sensitivity of the DarkSide-50 experiment. Similarly to what has been done in Sec. 6.5.4, we computed the 90% C.I. bound on the DM cross-section σ_{SI}^{DM} as a function of the DM mass using the real DarkSide-50 dataset, see the orange solid line in Fig. 6.22. Also in this case, this line is compared with the two most competitive exclusion limits, defined as frequentist 90% C.L.s, in the GeV and sub-GeV DM mass region, i.e. the XENON1T Migdal search [90] (black line), the XENON1T NR search [164] (black dashed-dot line), and the CRESST-III result [64] (gray line), as well as with the expected sensitivity (green dashed line) of Fig. 6.15.

No relevant deviations with respect to the expected sensitivity are observed. In addition, we performed the same stability test as we showed in Sec. 6.5.3 (namely the study of the behavior of the posterior *pdf* of the nuisance parameters as a function of the mass of the DM candidate), obtaining similar results.

The best sensitivity is reached for a DM mass of $5.3\,\mathrm{GeV}/c^2$, with a 90% C.I. upper bound on the DM SI cross-section $\sigma_{DM}^{SI}[90\%\ \mathrm{C.I.}] = 1.7 \times 10^{-43}\,\mathrm{cm}^2$. The improvement with respect to the previous analysis published in 2018 [1] is mainly due to the better comprehension of the detector response model, the background model, and an optimized data selection. The conclusion of this analysis is that the DarkSide-50 experiments reaches the world's best SI limits for DM candidates with masses in the range $[0.06, 4.7]\,\mathrm{GeV}/c^2$, and improves the previous one by a factor of ~ 10 at $3\,\mathrm{GeV}/c^2$.
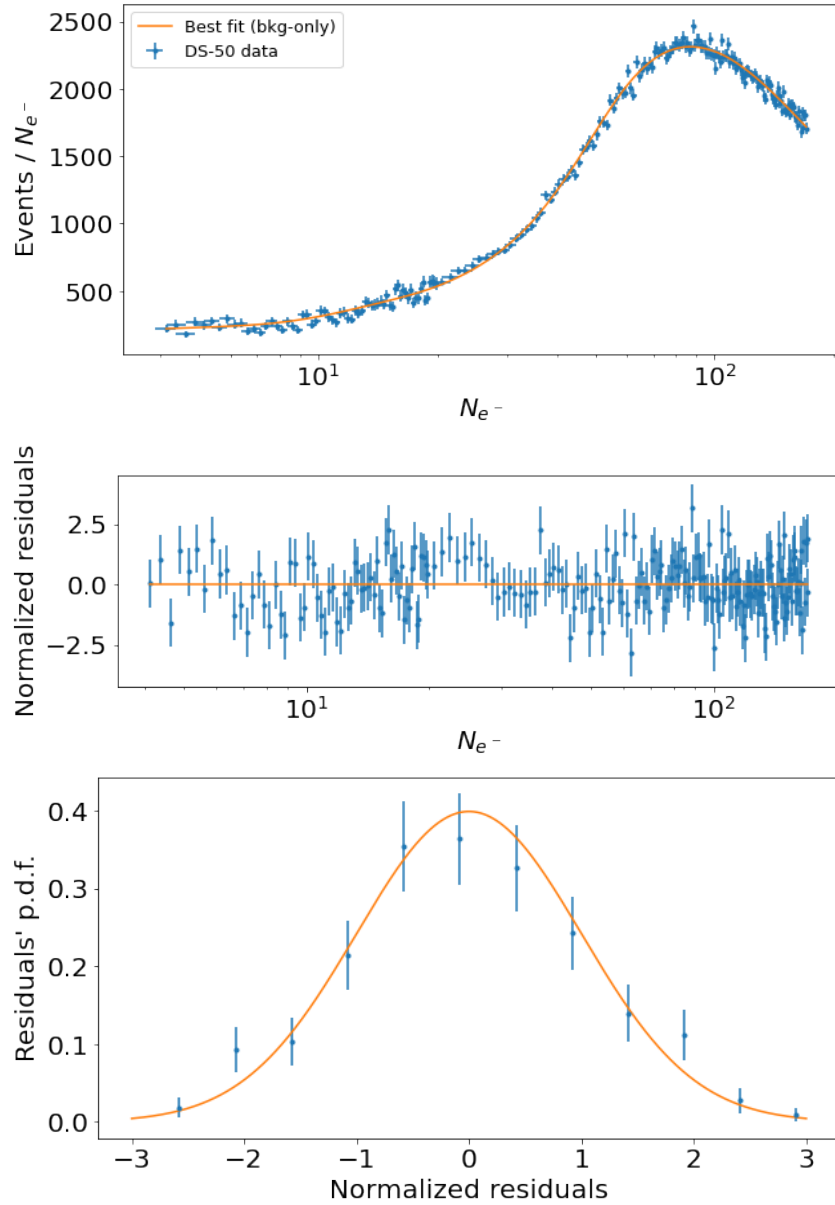
Figure 6.20. **Top**: Best fit (orange line) to the observed dataset (blue points) of the DarkSide-50 experiment. **Middle:** Normalized residual (blue points) with respect to the best fit. **Bottom:** Distribution of the normalized residuals (blue points). The orange line represents a standard normal *pdf*, which is the expected distribution for the normalized residuals.
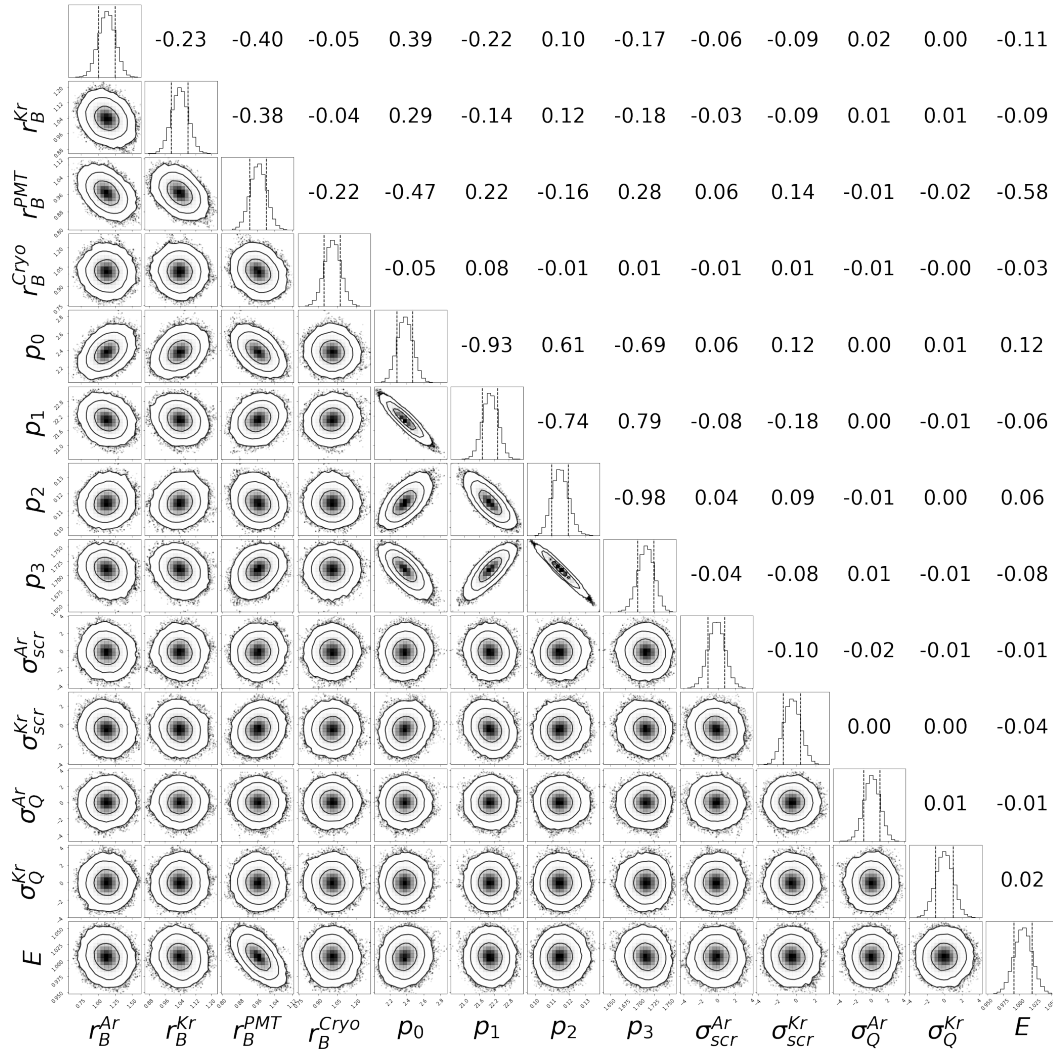
Figure 6.21. Graphical representation of the joint posterior *pdf* of the background only fit on the DarkSide-50 observed dataset. The plots on the diagonal of the figure are the uni-dimensional *pdf* of each single parameter obtained by marginalizing on all the others. The bi-dimensional *pdf*s in the bottom-left corner of the figure give the joint *pdf*s of each pair of parameters obtained by marginalizing on the others. The plots show also the credible regions at 68%, 95%, 99.7% probability as solid contour lines. The correlation coefficients are given in the upper-right corner of the figure.

Figure 6.22. The 90% C.I. observed sensitivity on the DM cross-section σ_{SI}^{DM} as a function of the DM mass (orange solid line) with the inclusion of the Migdal effect. The black line is the frequentist 90% C.L. from the XENON1T Migdal search [90], the black dashed-dot line is the is the one from the XENON1T NR search [164], the gray line is the one by the CRESST-III experiment [64], and the purple line is the one from the 2018 DarkSide-50 pure NR analysis [1]. The yellow dashed line represents the neutrino floor for an argon target, while the green dashed line is the expected sensitivity shown in Fig. 6.15.
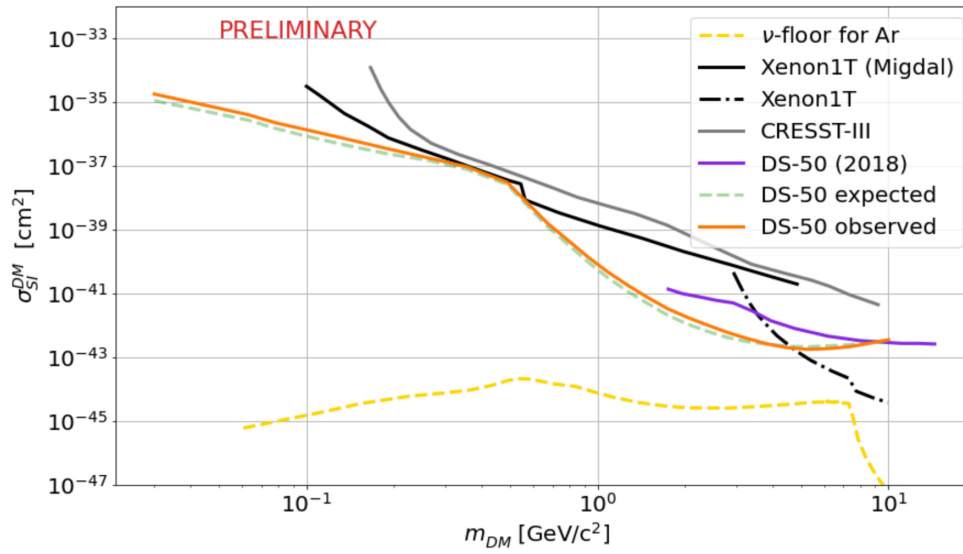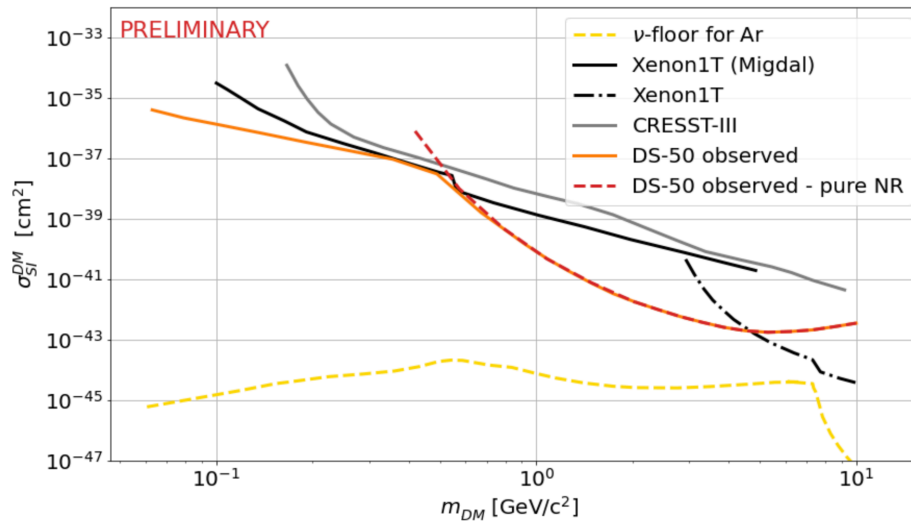
Figure 6.23. The 90% C.I. observed sensitivity on the DM cross-section σ_{SI}^{DM} as a function of the DM mass with (orange solid line) and without (red dashed line) the inclusion of the Migdal effect. The black line is the frequentist 90% C.L. from the XENON1T Migdal search [90], the black dashed-dot line is the one from the XENON1T NR search [164], and the gray line is the one by the CRESST-III experiment [64]. The yellow dashed line represents the neutrino floor for an argon target, while the green dashed line is the expected sensitivity shown in Fig. 6.15.

Finally, the portion of the sensitivity curve corresponding to DM masses $\lesssim$ 0.7 GeV/c^2, where the NR contribution starts to drop below the $N_{e^-} = 4$ threshold, is dominated by the Migdal effect. This is particularly visible in Fig. 6.23, where the red dashed line is the observed sensitivity if we neglect the Migdal effect. As expected, this has the power of extending the sensitivity of the experiment to sub-GeV masses. Following the same discussion outlined in Sec. 6.5.4, we believe that for DM masses $\gtrsim$ 200 MeV/c^2 this result is solid against possible theoretical uncertainties on the contribution of the liquid argon valence shell to the Migdal rate.

6.6 Conclusions

In this chapter we illustrated an innovative method to include the detector response, the background, and the signal models in the likelihood function.

With our approach, the likelihood function is explicitly a function of the calibration parameters and additional parameters describing the signal and background models. We exploit the new parametric detector response, retaining its analytical structure, to construct analytical or numerical expressions for each spectrum. In such a way we maintain a direct dependence of the spectra on each detector response background and signal parameters with their original *pdf*s without further assumptions (e.g. additional penalty terms or possible correlations). Therefore, there is no need to parameterize the systematic variations of the normalization or shape of the background and signal spectra, nor to define how such variations depend on the fit nuisance parameters. We perform a Bayesian fit to obtain the

joint posterior *pdf* from which we extract the limit on the parameter of interest, namely the signal strength r_S, and quantify the impact of the systematic uncertainties. The information enclosed in the posterior *pdf*s can be directly associated with parameters which have a clear and meaningful interpretation, such as calibration and background parameters. We showed that the fit to the whole N_{e^-} spectrum brings additional information to the detector response, resulting in smaller uncertainties on the calibration parameters. Finally, having implemented this method in a (semi)analytical way with the possibility to run on GPUs, drastically improves the performances, reducing the computational time without suffering from possible sampling fluctuations due to a Monte Carlo implementation of the detection model.

We applied this method to fit the DarkSide-50 S2 spectrum in the range [4, 170] N_{e^-} using as inputs the latest background model, the recent ionization response model and calibrations [144], and the NR and Migdal signals for DM masses in the range [0.06, 10] GeV/c^2 [2]. The results, in terms of performances, can be summarized as follows:

- the detector response and the background models represent a good description of the DarkSide-50 S2 data: the background-only fit does not exhibit significant tension between prior and posterior *pdf*s of the parameters;

- the DarkSide-50 S2 data provide significant information on the calibration and background parameters reducing some of their standard uncertainty by roughly a factor of 2;

- no significant change in the parameters' posterior *pdf*s is observed as a function of m_{DM};

- the systematic uncertainties on the calibration parameters determining the detector response model contribute significantly to the degradation of the sensitivity. Their impact has been evaluated to be a 40% degradation at 4.52 GeV/c^2;

- therefore, the ability of the fit to constrain the detector response and background models substantially helps in reducing the impact of the systematic uncertainties on the sensitivity.

Thanks to the improvement in the comprehension of the background and detector response model, and to the optimized data selection, the DarkSide-50 experiment results to set the world's strongest bound on SI DM interactions in the [0.06, 4.7] GeV/c^2 mass range. These results are confirmed by a parallel analysis performed using the frequentist 90% C.L. to assess the sensitivity, showing a quantitative agreement between the Bayesian 90% C.I. and the frequentist 90%C.L. estimation, as already indicated in Tab. 6.3.

We were able to determine the sensitivity in the sub-GeV range thanks to the inclusion of the Migdal effect. We elaborated on possible theoretical uncertainties affecting the Migdal contribution to the signal, concluding that our results are insensitive to these systematic effects for DM masses $\gtrsim 200\text{MeV}/c^2$. Below this mass region, the signal event yield is dominated by the contribution from the outermost

shell, which is affected by large theoretical uncertainties, and thus the impact on the sensitivity become more important.

Finally, this approach will also be employed for the assessment of the sensitivity of the DarkSide-50 experiment to other kinds of possible interactions, testing for example the solar axions hypothesis, the sterile neutrinos hypothesis, and the boosted dark matter hypothesis [167–171], along the lines of our preliminary analysis described in Ref. [172]. This work will contribute to a series of publications that are in the internal review phase and will be published in the next months.

Chapter 7

The CYGNO experiment

In the previous chapters we illustrated the importance of the Migdal effect in the field of DM direct detection, and we mentioned the possibility of measuring it using a CYGNO prototype. In order to better appreciate the details of this possibility, which will be treated in Sec. 8.3, this chapter will be dedicated to the description of the CYGNO experimental approach.

In particular, in Sec. 7.1 we will give a brief introduction to the main motivations for the CYGNO experiment. In Sec. 7.2, we will describe the general features of the CYGNO experiment. We will put particular emphasis on the performances and the results of its prototypes, shown in Sec. 7.3. Finally, in Sec. 7.4 we will illustrate the details of the CYGNO plan for the construction of a large TPC detector, whose preliminary sensitivity projections, based on simplistic assumptions on the possible backgrounds and angular resolution of the bigger detector, will be discussed in Sec. 8.1.

7.1 Motivations

Given the low NR interaction rate, the discrimination of NR induced by the DM from ER induced by other particles, which have much higher rate, is one of the main experimental challenge of direct DM searches in the GeV mass region. As already outlined in Sec. 2.1.2, since the experiments are located on Earth, they are exposed to an apparent DM wind pointing towards the Cygnus constellation, and two peculiar effects can be exploited to positively identify a DM signal. The first effect is the annual modulation of the DM-nucleus interaction rate, induced by the fact that, during its motion around the Sun, the Earth circular velocity is parallel to the DM wind during winter, and antiparallel during summer. This is the modulation claimed to be observed by the DAMA collaboration [31,32,121–124], see Sec. 4.4. However, the diurnal directional modulation of the DM signal would be an additional and more robust signature. Indeed, the maximum of the DM expected flux is realized in the direction of the apparent DM wind, i.e. towards the Cygnus constellation. Therefore, as a result of the Earth's rotation on its axis, an experiment on the Earth's surface would see the average DM direction changing of $\sim 96°$ every 12 sidereal hours, with an amplitude depending on the angle between the laboratory frame and the Earth's axis. The effect is minimum, with no modulation at all, along
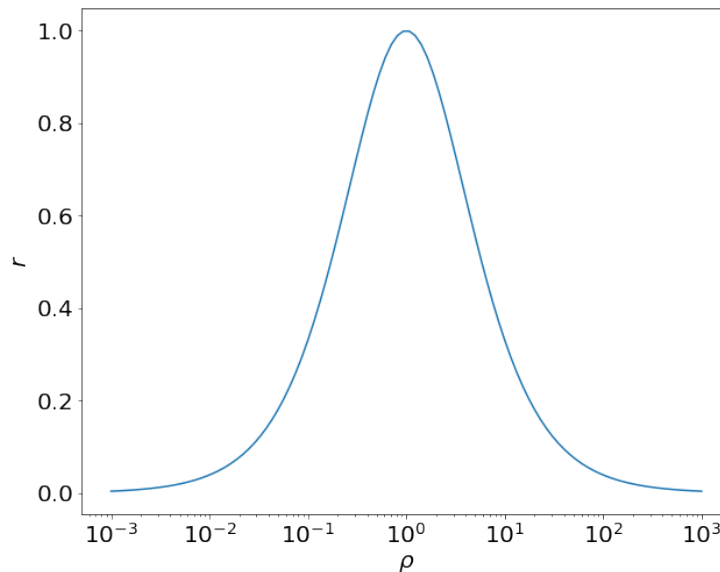
Figure 7.1. Graphical representation of Eq. 7.1. The horizontal axis is in logarithmic scale.

the directions parallel to the axis, while it is maximum at 45° inclination.

The incoming direction of the DM could therefore be used to infer correlations with astrophysical sources [173] that can not be mimicked by the existing background sources. In addition to that, since the pattern of the directional modulation depends on the properties of the DM halo, it would allow discriminating between several possible halo models, unlike conventional non-directional detectors [173].

Even if the > 10 GeV DM mass region has been intensively explored in the past years with improvements of many order of magnitudes in the DM cross-section sensitivity, the $\mathcal{O}(\text{GeV})$ mass range is still theoretically well motivated [174–176]. As we already outlined, in this region the sensitivity to the NR interactions is lost due to the lower energy threshold for the NR detection. Therefore, even considering the efforts to lower the thresholds, to include new signatures like the Migdal effect [2, 3, 85, 86, 177] and photon bremsstrahlung [178], or to include DM-electron scattering [106, 179, 180], the $\mathcal{O}(\text{GeV})$ mass range is still largely unexplored.

In Sec. 2.1.1 we illustrated the kinematics of a DM-nucleus elastic interaction. We also defined by Eq. 2.18 the parameter r as the maximum fraction of the energy that can be transferred to the target nucleus. If we now define $\rho = \frac{m_N}{m_{DM}}$, Eq. 2.18 can be rewritten as

$$r = \frac{4\rho}{(\rho + 1)^2}. \tag{7.1}$$

The above equation is reported in Fig. 7.1: this is suggesting that, due to kinematic reasons, the best sensitivity of a DM direct detection experiment is achieved when $\rho \sim 1$, i.e. when the DM mass is equal to the mass of the target nucleus. Therefore, if we want to focus on the direct detection of $\mathcal{O}(\text{GeV})$ DM candidates, the best choice as target nucleus is a light atom, such as hydrogen and helium.

The innovative approach of the CYGNO experiment is to develop a gaseous Time Projection Chamber (TPC) for the high 3D tracking of ERs or NRs. The

target will be a low mass gas mixture operated at atmospheric pressure, containing light elements such as helium and fluorine, with the aim of boosting the sensitivity to $\mathcal{O}(\text{GeV})$ DM masses, both for the SI and the SD coupling. Using light atoms has the additional advantage that the track lengths are expected to be longer, thus increasing the sensitivity to the direction of the observed recoils. The gas mixture which has been more extensively studied by the collaboration up to now is a He:CF_4 mixture, but new studies to add a small amount of an hydrogen based gas like isobutane (C_4H_{10}) are ongoing. The particle identification is also improved by studying the topological properties of the recoil event, allowing rejecting the background to the desired low energy threshold. Operating at atmospheric pressure, besides guaranteeing a reasonable active volume to mass ratio, has the advantage of reducing the engineering requirements for the vessel, and, as a consequence, the internal background. Thanks to these features, CYGNO will have the possibility to explore several physics cases, such as, for instance, solar neutrinos [181, 182] or the elastic scattering of sub-GeV DM [183].

The results obtained with current prototypes are the starting point in the view of designing a 1 m^3 demonstrator, expected to be installed and operated at LNGS in 2024/25. A competitive experiment with an even larger volume will be proposed by the collaboration, according to the performance of the 1 m^3 demonstrator. In order to reach this goal, however, a lot of work is needed to better assess the background characterization and rejection, as well as the performance in the reconstruction of the angular direction.

The project is part of the wider international CYGNUS, whose aim is developing an underground distributed Galactic Directional Recoil Observatory to detect DM even beyond the Neutrino Floor and measuring the coherent scattering of neutrinos from the Sun and Supernovae [184].

7.2 Description of the experiment

The CYGNO experiment will consist of an underground gaseous TPC with an high resolution optical readout employing Gas Electron Multipliers (GEMs), working at atmospheric pressure and using a helium/fluorine base gas mixture. This experimental setup will be built with the aim of studying rare processes with energy releases of hundreds of eV up to tens of keV.

The most important features that makes gaseous TPCs a promising approach for directional DM searches can be listed as follows:

- TPCs [185–187] allows acquiring large volumes with a small amount of readout channels, if compared to other 3D tracking systems;

- using a gaseous target allows having very low energy detection thresholds: an energy release of the order of few tens of eV is enough to produce a single electron cluster, which in turn has a quite good probability of reaching the multiplication stage and being detected;

- the measurement of the ionization allows accessing the energy released during the primary interaction, and the profile of the energy release, if the granular-

ity of the readout is enough to measure it, is in turn an important tool for signal/background discrimination;

- the 3D reconstruction of the track allows accessing the axis of the recoil, while the charge profile along the track allows accessing the track orientation (*head-tail*);

- the TPC can be filled with a variety of possible gas mixtures, including light nuclei with even or odd number of nucleons, sensitive to SI or SD DM interactions, respectively, also in the $\mathcal{O}(\text{GeV})$ mass region;

- the possibility of operating at room temperature and atmospheric pressure is a great economical and operational advantage, since there's no need for vacuum sealing or cooling: a simpler setup will also have a better scalability;

- even if a gaseous target is less dense with respect to a liquid or solid target, very large TPCs can be built: up to 100 m^3 of active volume TPCs have been already successfully operated [188,189], and even larger has been approved for construction [190].

7.2.1 The optical readout

In a similar way as we already explained it in Chap. 5 for LAr, charged particles traveling in the gas can ionize atoms and molecules but can also excite them. As a consequence of the de-excitation processes, photons are emitted. The energy spectrum and the intensity of this light will depend on the gas mixture, on its pressure and on the electric field [191–194].

In recent years, an increasing number of experimental collaborations have started using Micro-Pattern Gaseous Detectors (MPGDs). These detectors are characterized by a high granularity and rate capability, and the modern MPGDs are by now very high quality, stable and uniform devices. The GEMs [195], in particular, have been used for the readout of very large volumes [188] reaching high spatial and time resolution, and they have been proved to have good performances if coupled to pixelated light sensors [191,196–198].

The CYGNO experiment will use an sCMOS-based optical device for the GEM readout, as the collaboration has been proposing since 2015 [199]. This setup, in fact, as it is outlined in Sec. 7.3, has been proved to have an excellent performance at the energy needed for DM searches, but also for the detection of cosmic rays and high energy electrons [197,200–203]. Besides the fact that highly performing sensors can be easily procured, since they are being developed for commercial applications, one of the advantages of this choice for the optical readout is that these sensors can be installed outside the TPC volume, reducing possible issues related, for instance, to the gas contamination. In addition, by means of proper lenses, with this setup is possible to acquire images of a large area (up to $\mathcal{O}(1\,\text{m}^2)$) with a single sCMOS sensors and with a very good spatial resolution (with a $\mathcal{O}(100\,\mu\text{m})$ effective pixel size). However, the frame rate currently available for sCMOS is still too low if compared to the duration of the signals in a TPC. This is the reason why, in order to achieve a 3D track reconstruction capability, the CYGNO experiment will employ
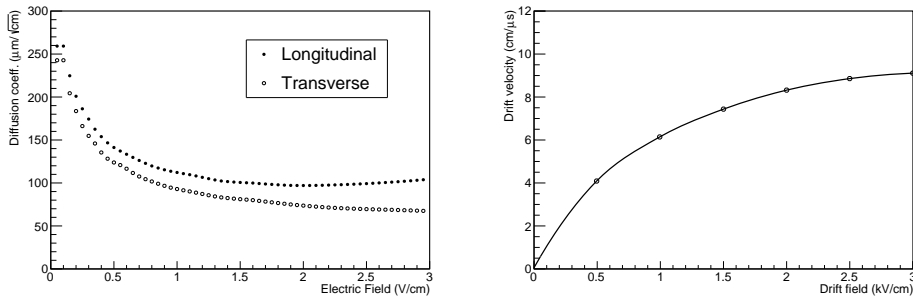
Figure 7.2. **Left**: longitudinal and transverse diffusion coefficients as a function of the drift field for a 60/40 He/CF$_4$ gas mixture. **Right**: drift velocity as a function of the drift field in the same gas mixture. Both the two plots were obtained by means of `Garfield` [206, 207]. Figure from Ref. [208].

additional fast photosensors (PMTs or SiPMs), which can provide the information on the track profile along the direction of the drift field.

7.2.2 The gas mixture

The features of the gas, namely its transport properties, such as the diffusion and drift velocity, the ionization statistics, the light production, and the electron multiplication, affect the overall detector performances, from the position and energy resolution to its photon yield. The CYGNO experiment employs, as a scintillator, the tetrafluoromethane (CF$_4$): besides its valuable scintillation properties, it is a particularly interesting gas for DM since, due to the quite large fluorine content, guarantees sensitivity to SD DM-nucleon interactions. In addition, the peak of the CF$_4$ light emission, as it has been found in preceding studies [191], is reached at approximately 600 nm, which is the wavelength in which the sCMOS Si-based sensors have the greater quantum efficiency.

Different He/CF$_4$ gas mixtures, with different relative proportions of the two components, has been tested and characterized by the CYGNO collaboration, finding the best results in terms of performance with a 60% helium and 40% CF$_4$ mixture [204, 205]. The two plots of Fig. 7.2 show the response of the mixture in terms of diffusion coefficients (left) and drift velocity (right) as a function of the drift field, as calculated with `Garfield` [206, 207]. These estimates indicate that the average energy loss per single ionization is ~ 42 keV. The left plot of Fig. 7.2 also shows that the use of CF$_4$ results in a small electron diffusion, which is a very valuable property since it means that the track's original shapes are not heavily deformed.

In order to better characterize the electronic and the He-nuclei recoils, MC simulations with the GEANT4 [209] and SRIM software [165], respectively, have been developed. In particular, Fig. 7.3 shows the 3D range, defined as the average distance between the starting point of the track and its end point, as a function of the recoil energy. As it is reported in the figure, He-nuclei recoils have a very small size, which is expected to be less than 1 mm up to 100 keV: this suggests that these events are detected as very bright spots with a size which is mainly due to the
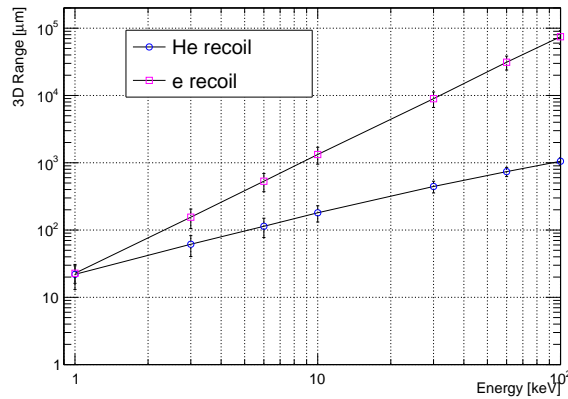
Figure 7.3. Average 3D range, defined as the distance between production and absorption point for electronic and He-nucleus recoils, as a function of their kinetic energy. The gas mixture considered for this study is the usual 60/40 He/CF$_4$ gas mixture. Figure from Ref. [208].

diffusion in the gas. On the other hand, electronic recoils with energy below 10 keV have a larger spatial size with respect to the He-nuclei nuclear recoils with the same energy, and are therefore expected to be detected as less bright spots. When the energy of the electronic recoil exceeds 10 keV, the size of the events becomes greater than 1 mm, and tracks of a few centimeters are expected for more energetic recoils.

Further studies with other gas mixtures are also ongoing. The possibility of adding a small amount of a hydrogen-rich gas, such as isobutane (C$_4$H$_{10}$) or methane (CH$_4$), is being investigated, with also the possibility of using alternative gas with a reduced Global Warming Power, such as HFO. Finally, in the view of a possible measurement of the Migdal effect induced by neutrons with a CYGNO prototype - see Sec. 8.3 for more details - an Ar/CF$_4$ gas mixture is also going to be considered and characterized.

7.3   Performance of the CYGNO approach: the LEMON prototype

Presently, the most comprehensive studies on the performance of the approach carried on by the CYGNO collaboration are the experimental results realized with the Long Elliptical MOdule (LEMON), a 7 liters prototype. Fig. 7.4 illustrates the experimental apparatus, which is characterized by a sensitive gas volume of 7 L confined in a TPC (labeled with A) of cylindrical shape with an elliptical base. The longitudinal length of the field cage is 20 cm, while the two axes of the ellipse constituting the TPC base have 24 cm and 20 cm length. The amplification stage is made of a 24×20 cm^2 stack of three GEMs, which are in turn optically coupled, through a black bellow (labeled with C), to the sCMOS camera (labeled with D). The bottom electrode of the last GEM is used as the anode, while the cathode, on the opposite side of the TPC, consists of a semi-transparent mesh. Behind the
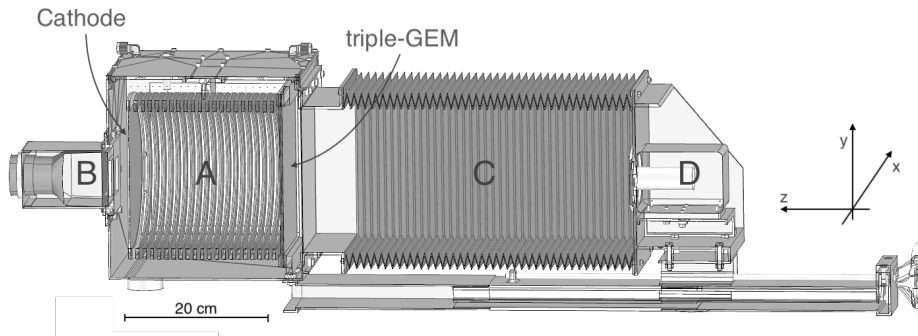
Figure 7.4. Drawing of the LEMON prototype [203] indicating the elliptical sensitive volume (A), the fast photo-multiplier (B), the optical bellow (C) and the sCMOS-based camera (D). Figure from Ref. [208].

mesh, a PMT (labeled with B) is used as a fast photodetector.

The LEMON prototype operated with the following standard conditions:

- a He/CF$_4$ (60/40) gas mixture flux of 200 cc/min;

- an electric drift field within the sensitive volume $E_D = 0.5$ kV/cm;

- an electric transfer field in the 2 mm gaps between the GEMs $E_{\text{Transf.}} = 2.5$ kV/cm;

- a voltage difference across the two sides of each GEM $V_{\text{GEM}} = 460$ V;

Following the results of Ref. [205], this configuration allows obtaining an expected electron multiplication of about 1.5×10^6.

In Sec. 7.2.1 we already mentioned the importance of having high quality camera sensors. It is for this reason that LEMON was instrumented with the ORCA Flash 4.0 camera by Hamamatsu. The camera is equipped with a 1.33×1.33 cm^2 sCMOS sensor of 2048×2048 pixels, each of which covers an active area of 6.5×6.5 μm^2. Its quantum efficiency at 600 nm is 70% and the expected readout noise is 1.4 electrons RMS. A calibrated light source [197] has been employed to test the response and the noise level of the sensor: the measurements resulted in a response of 0.9 counts/photon and in an RMS fluctuation of the pedestal of 1.3 photons/pixel.

A Schneider lens with 0.95 aperture and 25.6 mm focal length has been used in order to allow the camera to take pictures of the entire GEM surface. With these lens features, the camera sensor is placed at 52.6 cm distance from the GEMs, corresponding to a covered solid angle of 1.6×10^{-4}, which in turn regulates the geometrical acceptance of photons.

By itself, the camera sensor only provides the 2D projection of the track profile. In order to have an additional tool to reconstruct the 3D track profile, the solution chosen by the collaboration is to add to the setup a fast photo-sensor, a Photonics XP3392 PMT. A different solution was considered, since the time information on the detected signals could be extracted by reading the signal induced on the third GEM bottom electrode. However, as it has been shown in Ref. [200], this setup
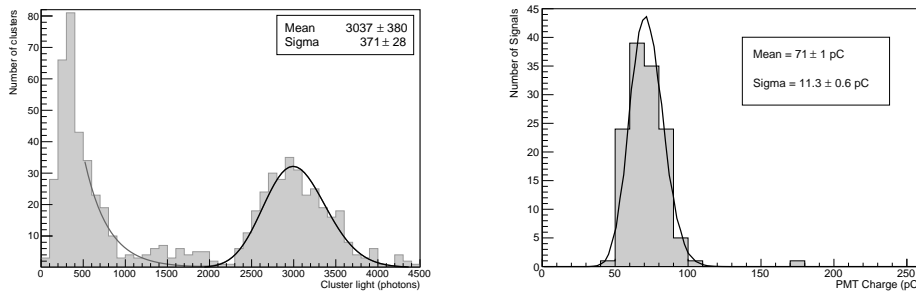
Figure 7.5. Left: light spectrum of the ^{55}Fe events reconstructed by the NNC algorithm from the sCMOS images. **Right**: distribution of the total charge signal collected by the PMT. Figure from Ref. [208].

would result in considerable additional noise, which could affect in a critical way the detection of low energy recoils. Therefore, the PMT, with its 5 ns rise-time, a maximum quantum efficiency of 12% at 420 nm, a 76 mm square-window, and single-photon sensitivity, offered the possibility to access the time profile of the events with significant reduce noise, if compared to the GEM electric signal solution.

In the following sections, we will illustrate in detail the performance of the LEMON prototype. The tests shown and discussed here have been performed at the overground site of INFN Laboratori Nazionali di Frascati (LNF) by means of ^{55}Fe and AmBe radioactive sources, 450 MeV high energy electrons from the so-called Beam Test Facility (BTF) [210, 211], and cosmic rays.

7.3.1 Light yield, energy resolution and detection efficiency

The detector response in terms of light emission and detection has been assessed with an ^{55}Fe source. This source is characterized by a 5.9 keV X-ray emission, by means of which it has been possible to study the light yield and the energy resolution of both the PMT and the sCMOS sensor. The tests on the sCMOS were performed acquiring 100 ms exposed images in a trigger-less mode. The pedestal noise of each pixel of the sensor has been determined averaging over 100 images in which no light signal has been detected, and it was subtracted by the images which have been employed for the analysis. In order to identify the energy deposits induced by the ^{55}Fe source, the images has been processed with a simple nearest-neighbor clustering (NNC) algorithm.

In the left plot of Fig. 7.5 the light spectrum of the ^{55}Fe events as reconstructed from the sCMOS images is reported. The right plot of the figure shows instead the distribution of the total charge signal measured by the PMT. A Polya fit [212] has been performed on the two spectra: the resulting average light yield for the sCMOS sensor is 514 ± 63 photons per keV, with a 12% RMS energy resolution, while the resulting average charge yield of the PMT is (12.0 ± 0.2) pC per keV together with an RMS energy resolution of 16%. The width of the two distributions can be attributed to the statistical fluctuations in the number of primary ionization electrons produced during the event (8%) and to the statistical fluctuations related to the gain of the first GEM (10%).
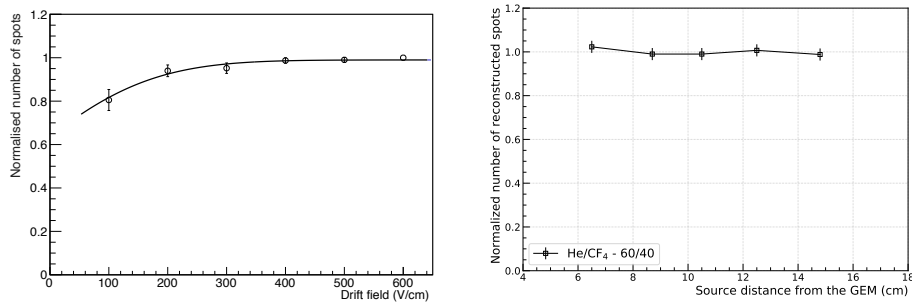
Figure 7.6. Left: distribution of the number of ^{55}Fe spots as a function of the drift electric field E_D, normalized to the value obtained for $E_D = 600$ V/cm. **Right**: distribution of the number of ^{55}Fe spots as a function of their distance from the GEMs, normalized to the average value. Figure from Ref. [208].

In order to evaluate the detection efficiency along the whole 20 cm drift length, a collimated ^{55}Fe source was placed at different positions in order to vary the distance of the original interaction events from the GEMs. Moreover, different configurations for the drift field were studied. The left plot in Fig. 7.6 shows the number of ^{55}Fe events in the TPC, reconstructed with the simple NNC algorithm, as a function of the drift electric field E_D, normalized to the value obtained for $E_D = 600$ V/cm: a plateau is reached for $E_D > 300$ V/cm, corresponding to the full detection efficiency. The right plot of Fig. 7.6 shows instead the dependence of the number of ^{55}Fe spots as a function of the distance of the source from the GEMs, with a drift field of $E_D = 600$ V/cm and normalized to the average value: the results are consistent with a constant value, independent of the distance of the interaction point from the multiplication stage.

7.3.2 Track absolute distance along the drift direction

In order to determine the ability of the LEMON setup to reconstruct the z position of the energy deposits along the drift direction, the experiment has been exposed to the 450 MeV electrons from the LNF-BTF facility [203]. The drift length can be extracted by studying the transverse diffusion during the drift, and this allows inferring the absolute z distance at which the track was originally generated. The performance for small energy releases has been evaluated on 7 mm long track segments. The estimated position resolution is 100 μm for events generated closely to the GEMs and 300 μm for events generate on the opposite side of the TPC, at a 20 cm distance from the multiplication stage.

The light profile transverse to the track direction has a Gaussian aspect [203]. The total amount of light L is proportional to the width σ and the amplitude A of the Gaussian. However, not all the electrons produced in the gas volume after the primary interaction can reach the multiplication stage, due to the attachment effect in the gas [213]: therefore the total amount of light of a track follows an exponentially decreasing behavior, and goes as $L = L_0 e^{-z/\lambda}$, with λ being the electron mean free path in the gas and z being the distance of the interaction point from the GEM stack. The quantity $\eta \equiv \sigma/A$, considering that $\sigma \propto \sqrt{z}$ due to
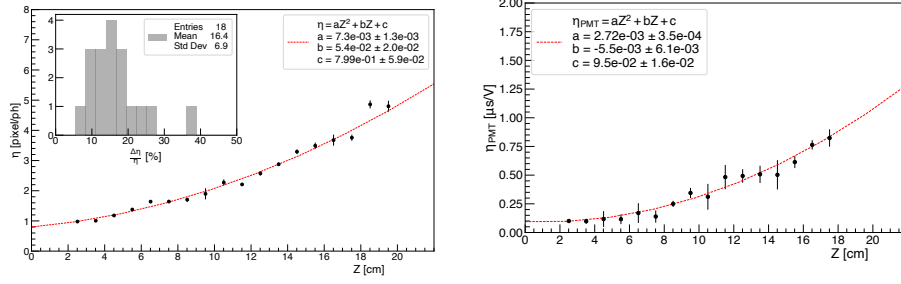
Figure 7.7. Behavior of η (**left**) and η_{PMT} (**right**) as a function of z, the track distance from the multiplication stage. The red lines are a fit of a quadratic polynomial to the experimental data. The inset in the left plot contains the distribution of the ratio between the RMS and the average values of the η spectra as a function of z. Figure from Ref. [208].

the diffusion in the gas, is expected to grow quadratically with z [203]. The time dependence of the signal detected by the PMT is influenced by the longitudinal diffusion, too, and thus, even for the PMT, the quantity $\eta_{PMT} \equiv \sigma_{PMT}/A_{PMT}$ is expected to increase with z. The left and the right plots in Fig.7.7 show the behavior of dependence of η and η_{PMT}, respectively, as a function of z. The red lines are a quadratic fit to the experimental data, while the distribution of the ratio between the RMS and the average values of the η spectra is shown in the inset in the left plot as a function of z. By using these observable, it is possible to infer the absolute z with a 15% precision over the 20 cm length field cage [203]. Having access to the z position of the event will provide the full 3D track reconstruction, allowing discriminating events in the fiducial signal volume to reject events produced by internal radioactivity of both the GEMs and the cathode.

7.3.3   Discrimination between nuclear and electronic recoils

The high granularity of the optical readout allows us to extract a variety of observables related to the tracks, such as their shape, size, and light density. This information is an important tool for the discrimination of NR from ER events.

The need of having access to these observables is the reason why an advanced clustering algorithm, more refined with respect to very simplistic NNC, was developed for the CYGNO experiment. This algorithm, called iDBSCAN [214], is a custom extension of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [215], and it is able to recognize and cluster the light signals in the images produced by the sCMOS camera. A subsequent super-clustering algorithm, based on the Geodesic Active Contours algorithm (GAC [216, 217]), taking as an input the clusters reconstructed by iDBSCAN, is able, according to the light intensity in each pixel of the image, to assemble together the clusters belonging to the same longer track.

The algorithm was tested analyzing the ER events produced by the 5.9 keV X-rays produced by the ^{55}Fe source and the NR events produced by an AmBe neutron source [218]. The AmBe source, besides neutrons, also produces 59 keV photons, which were almost completely removed by means of a lead shield surrounding the
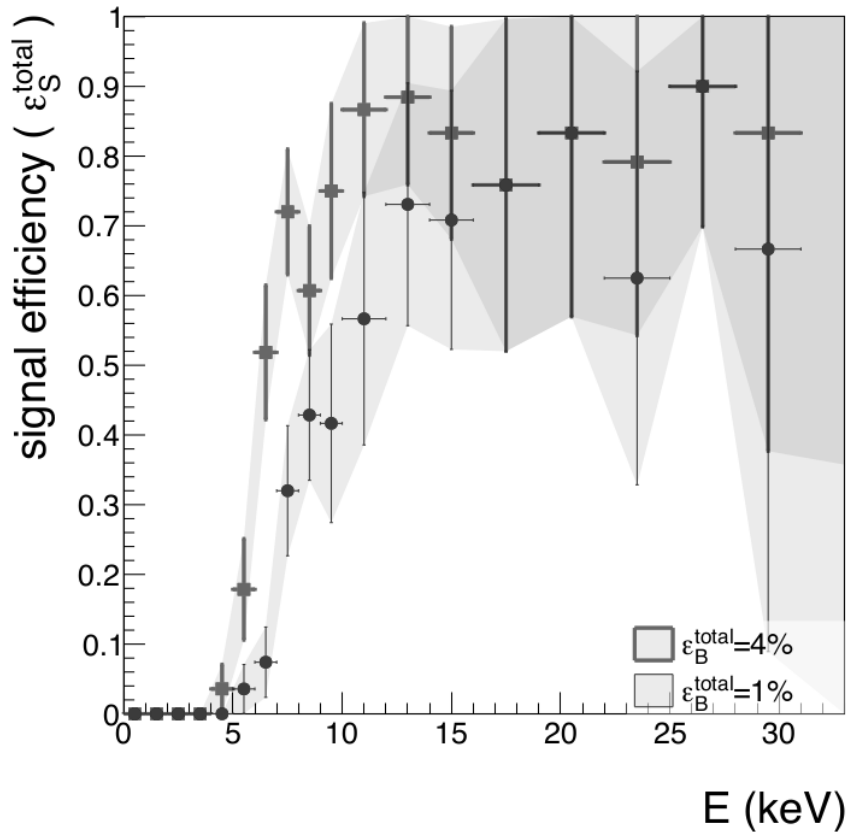
Figure 7.8. Signal efficiency for nuclear recoils as a function of their detected energy. The different points correspond to an electronic recoil rejection efficiency of 4% (squares) and 1% (circles). Figure from Ref. [208].

detector. The cosmic ray background was however still present since, as it was for the other measurements described in this chapter, the data has been taken overground, at LNF.

The selection of a sample of pure NR induced by the AmBe neutrons required the adoption of several cluster observables, such as the "slimness" ξ, i.e. the ratio of the width of the track in the transverse direction and the projected path length, and the light density δ, namely the ratio between the total number of photons of a cluster and the cluster size in terms of pixels. The slimness was employed to exclude the clusters belonging to cosmic ray events, while the light density was used for the NR/ER discrimination.

As depicted in Fig. 7.8, using a simple cut on the observable δ, a 50% (40%) NR discrimination efficiency is found in correspondence to an ER background rejection of 96.5% (99.2%) in the 5.9 keV$_{ee}$ energy region [218]. These results are of course obtained with a very simple cut, which therefore could be improved in the future including more sophisticated topological variables together with the information from the PMT data. However, this simplistic approach provides a promising rejection

factor of 10^2 for ER with an energy of 5.9 keV, and a high fraction of the NR events is not removed by the cut.

7.4 The CYGNO roadmap

In order to be able to fully characterize and optimize the detector, the CYGNO experiment will be developed and commissioned in different phases. Each phase will be characterized with a different size of the target mass, which will be given in terms of the volume. In order to get the target mass in each phase, we can consider that, for a He:CF$_4$ 60:40 gas mixture at room temperature and atmospheric pressure, 1 m^3 of gas corresponds to a mass of ~ 1.55 kg. In particular, the plan involves the following steps:

- PHASE_0: the installation of a 50 L prototype underground at LNGS, with the aim of studying its performance in a low background environment;

- PHASE_1: developing a demonstrator of $\mathcal{O}(1 \text{ m}^3)$ volume to test the scalability of the project, as well as better characterizing the internal background due to the radioactive isotopes in the materials constituting the apparatus;

- PHASE_2: propose a larger experiment of $\mathcal{O}(30 - 100 \text{ m}^3)$, depending on the outcomes of the previous phases, to test the 1-10 GeV DM mass region with a competitive sensitivity for both SI and SD DM interactions. With this kind of apparatus, we will also have the possibility of measuring the events induced by the solar neutrinos. Both for DM and neutrino searches, the directional capabilities of the CYGNO experimental approach will also provide useful information on the incoming particles from the astrophysical point of view.

In the following, we will summarize the main aspects characterizing the different stages of the project.

CYGNO PHASE_0: the LIME larger prototype

The larger prototype foreseen to conclude the R&D phase of the project is called Long Imaging ModulE (LIME), see Fig. 7.9. It has the same drift length as the future demonstrator, i.e. 50 cm, and the same readout apparatus, namely a 33 $\times$ 33 cm^2 triple GEM foil, stretched on a plexiglass frame and revealed by an sCMOS sensor and 4 small PMTs placed around the sensor in a symmetric way. Since the PHASE_1 detector will be made of LIME modules, this phase is crucial to validate the experimental approach towards the larger detectors.

The camera chosen to be used in LIME is the ORCA-Fusion because of its improved performance with respect to the Orca Flash: it is characterized by a lower noise (0.7 versus 1.4 electrons per pixel), a larger quantum efficiency (80% versus 70% at 600 nm), and a larger number of pixels (2304 $\times$ 2304 versus 2048 $\times$ 2048). The 4 PMTs are needed to have a better reconstructing ability for the track position and inclination.

The TPC gas mixture is enclosed in a 10 mm thick plexiglass box, and the field cage is composed of square copper rings, with rounded shape, at a distance of 16 mm from each other.
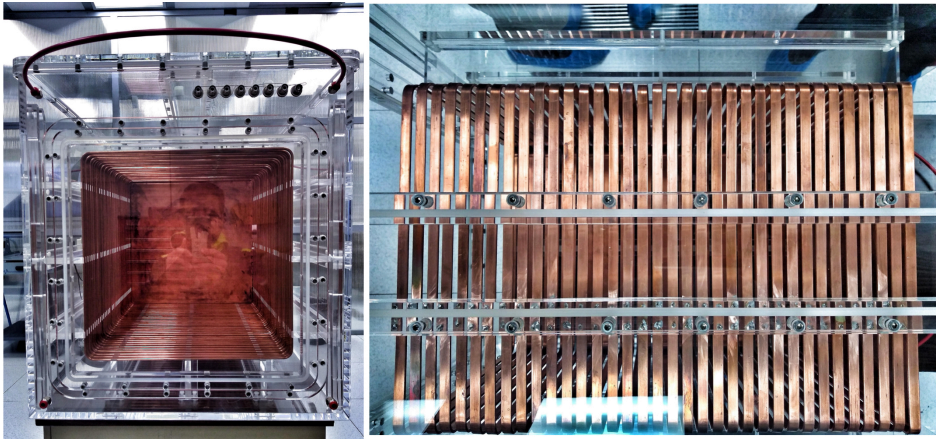
Figure 7.9. Pictures of the LIME detector. On the **left** there's a front view of the field cage, while on the **right** the field cage copper rings are shown. Figure from Ref. [208].

The DAQ system that will be employed for the underground campaign will be also used in the PHASE_1 detector, and it's currently under test.

With respect to the 514 ph/keV obtained with LEMON, see Sec. 7.3.1, thanks to the larger sensitivity of the Orca Fusion camera, it has been possible to achieve a measured response of about 650 ph/keV. It will be also possible to achieve a lower effective energy threshold of hundreds of eV thanks to the lower sensor noise, which is about 1 photon/pixel. The preliminary studies performed at LNF, where the LIME prototype has been assembled, showed an energy resolution on the ^{55}Fe peak of 14% in the whole 50 cm drift length.

To assess the stability of the prototype, it has been operated for one entire month, showing comparable stability to LEMON [208]. The installation at LNGS, completed with the PHASE_1 auxiliary systems, will allow testing the detector performance in a low radioactivity environment, and characterizing in more detail the external background present at the experimental location, validating the MC simulations.

CYGNO PHASE_1: the $\mathcal{O}(1 \text{ m}^3)$ demonstrator

Once the PHASE_0 will be concluded and the LIME apparatus will be optimized and characterized, the project will move to PHASE_1, with the aim of studying and minimizing the internal backgrounds on a bigger experiment.

The precise size of the detector will be defined depending on the available underground site, and it is still under discussion. However, in this section we will discuss a 1 m^3 volume demonstrator, schematically shown in Fig. 7.10. The layout has been designed in such a way that, as soon as the final detector size is defined, it can be directly and easily adapted to the definitive detector dimensions.

The gas volume of the detector will be contained in a gas vessel made of acrylic glass. Inside the vessel, there will be two field cages with a length of 500 mm, made of two back-to-back TPCs separated by a central aluminised mylar cathode. The two end-caps will have an active area of 1 m^2 surface each, and both of them will be readout by a matrix of 3×3 modules of 33×33 cm^2 area, each one equipped with
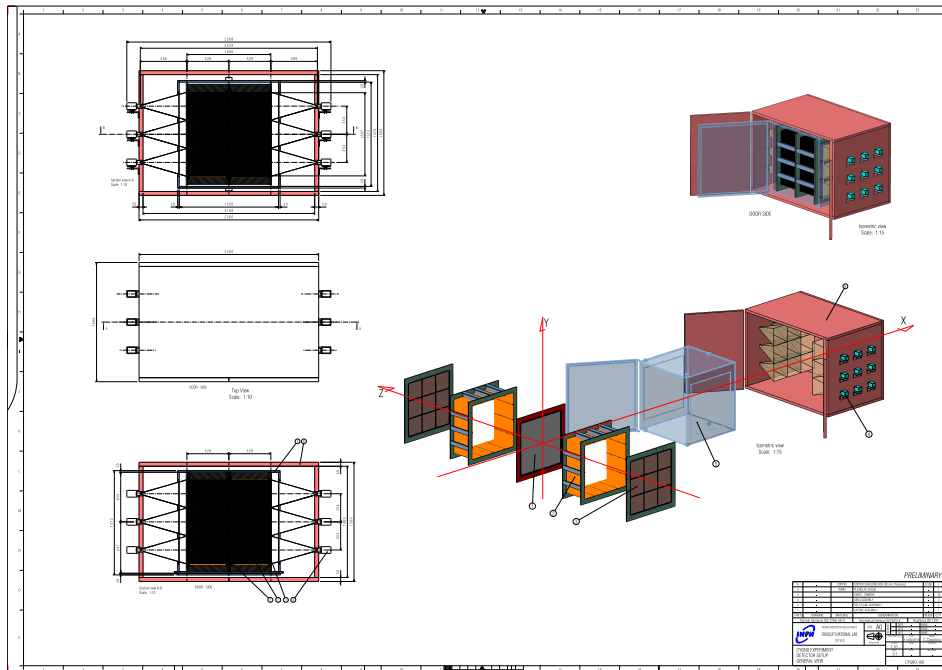
Figure 7.10. Rendering of the CYGNO PHASE_1 detector. Figure from Ref. [208].

the same readout of the LIME prototype: a triple GEM stack, a sCMOS sensor, and four PMTs.

The DAQ system, needed to be able to collect data from the cameras and the photodetectors, will have to handle a camera exposure from 0.2 to 1 second (1 to 5 Hz frame rate), approximately 10 MB of data per picture (5 MP, 16 bits/pixel), and the 12-bit digitization of photodetector waveforms at ~ 250 MS/s in $\lesssim 1$ μs windows. The fast signals from the PMTs could be used to trigger the acquisition of the sCMOS sensors. However, different possible trigger schemes are under assessment, even with the possibility of running in trigger-less mode.

However, in order to be able to deal with the fast signals of the PMTs, digitization boards are being considered. With this expected setup, in order to not be limited by the throughput to the disk, typically limited to $\mathcal{O}(200$ MB/s), some preselection of the images, processed for instance by a farm of CPUs, would be needed.

Different shielding options are also still under debate. For instance, the use of lead would reduce the experiment dimensions: however a GEANT4 based Monte Carlo simulation of the whole apparatus in Fig. 7.10 showed that this configuration would require archeological lead in order not to induce additional background from the shielding, therefore largely raising the cost of this layer. A compromise solution for the shielding has been found, namely a layer of 2 m of water and 5 cm of copper. This setup is able to reduce by a factor of 10^{-7} the external photons and by a factor of 5×10^{-5} the external neutrons: the number of expected ERs in the active volume is expected to be below 10^3 cpy (counts per year) with $\mathcal{O}(1)$ cpy NRs in the 1-20 keV range.

A study aiming at the background evaluation for the PHASE_1 detector showed

Figure 7.11. Rendering of the CYGNO PHASE_2 possible detector layout. Figure from Ref. [208].

that, in the 0-20 keV energy range, $\mathcal{O}(10^6)$ ERs per year and $\mathcal{O}(10^3)$ NRs per year are expected in the sensitive volume. The CYGNO experiment and its prototypes are the first detectors employing an sCMOS camera for DM searches. This is the reason why the intrinsic background of the sCMOS sensor has never been determined in this context, and, for this purpose, a collaboration with the sCMOS camera producer companies and with LNGS Services has been established.

CYGNO PHASE_2

A bigger CYGNO detector with an $\mathcal{O}(30-100\text{m}^3)$ volume would have the sensitivity needed to give a significant contribution to the DM searches in the $m_{DM} \lesssim 10\,\text{GeV}$ mass region, both for the SI and the SD interaction. In addition, in the case in which a signal incompatible with the ordinary matter interactions is detected, the directional capabilities of the detector would be fundamental to confirm the spatial origin of the detected signal and its properties.

In Fig. 7.11 a 30 m^3 sensitive volume possible apparatus that would fit, together with all its shields, into LNGS experimental Hall C is shown.

In order to build such a big detector with the CYGNO experimental approach, a properly scalable design, an improved readout system, and better materials are required. A list of possible improvements would include, for instance:

- the development of a custom sCMOS sensor, fulfilling the CYGNO requirements (low noise, high sensitivity, reduced intrinsic radioactivity, lower production cost);

- the reduction of the internal background due to the detector material radioactivity;

- the characterization of different gas mixtures to improve the tracking performances and the sensitivity to low energy releases.

Chapter 8

Future perspectives

In the previous chapters, we showed the state-of-the-art of the DM direct detection searches. We developed phenomenological studies stressing the importance of the Migdal effect in LAr experiments, and we carried out the analysis of the DarkSide-50 experiment with an innovative tool. Our conclusions, thanks also to the inclusion of the Migdal effect, are that DarkSide-50 has currently the world best sensitivity in the DM low mass region. Finally, we introduced the CYGNO experimental approach, with the aim of using it to get the first measurement of the Migdal effect in nuclear scattering.

The goal of this chapter is to close the loop of the work done so far: after a brief description of the available sensitivity projections for the future upgrades of the CYGNO and DarkSide projects, outlined in Sec. 8.1 and Sec 8.2, respectively, we will conclude the chapter addressing, in Sec. 8.3, the possibility of measuring the Migdal effect in NR interactions with the CYGNO experiment or its prototypes.

8.1 Preliminary study of the sensitivity of the CYGNO experiment

In this section, we will illustrate the preliminary study we developed to assess the expected sensitivity of CYGNO PHASE_2 to DM searches in the Bayesian approach and the tools developed to evaluate it. Since the definitive assessment of the actual size of the background levels, the detector response, and the angular resolution of the CYGNO approach is still under evaluation, the preliminary analysis shown in this section is based on simplistic assumptions yet to be confirmed by the R&D activities.

Even if the CYGNO approach allows us to measure the energy spectrum and the angular spectrum of the tracks simultaneously, and both will be used in the future analyses, this sensitivity study, for the sake of simplicity, will only consider the angular information. This is a first-order approximation, in the sense that the angular discriminating power is significantly stronger than the energy spectrum shape. The angular spectrum of the background has been assumed to be isotropic in the Galactic coordinates; on the other hand its energy spectrum, which depends on the materials used for the detector and the shielding materials, is instead difficult to predict.
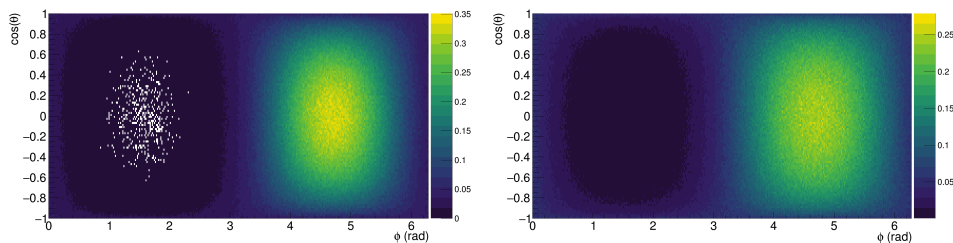
Figure 8.1. Angular distribution of DM induced events in the Galactic coordinates as a result of a MC simulation. The **left** plot shows a 10 GeV/c^2 DM signal on a helium target, while the **right** plot shows a 100 GeV/c^2 DM signal on a fluorine target.

The energy threshold is relevant for the determination of the signal angular distribution, because it is required for the events to be selected. In this sensitivity study, we considered two possible scenarios: a conservative 1 keV$_{ee}$ threshold, based on the published results [219], and a realistic 0.5 keV$_{ee}$ threshold, as suggested by the performances obtained with the new camera of the PHASE_0 LIME prototype, see Sec. 7.4. The quenching factor for the elements in our gas mixture - including hydrogen, since, as already mentioned, studies on hydrogen rich gas mixtures to improve the tracking performances are ongoing - has been evaluated by means of a SRIM simulation. The quenching factor is indeed needed to convert the electronic recoil energy into nuclear recoil energy. The quenching factors for H, He, C, and F in He/CF$_4$ 60/40 and at 1 atm resulted to be in the range 10%-30% for E[keV$_{ee}$]=100 eV and 60%-90% for E[keV$_{ee}$]=100 keV. These in turn resulted in an effective energy thresholds of 1.4 (0.8) keV$_{nr}$ for H, 2.1 (1.2) keV$_{nr}$ for He, 3.1 (1.8) keV$_{nr}$ for C and 3.8 (2.2) keV$_{nr}$ for F for 1 (0.5) keV$_{ee}$ energy deposits.

Taking into account the above results, we computed the signal angular distribution in the Galactic coordinates [55, 220, 221]. For this purpose, the effects induced by motion of the Earth has been neglected, having a marginal impact on the angular distribution. In Fig. 8.1, we show the DM NR signal distributions in the 2D Galactic coordinates. These distributions strongly depend on three quantities: the mass of the DM particle, the mass of the target element, and the energy threshold. In particular, they are more peaked at low masses and more spread at heavier masses, where there is no angular region forbidden by kinematics.

For this analysis, we decided to assess the 90% C.I. sensitivity on pseudo-datasets generated by extracting events according to the expected angular spectra and by adding the detector response effects. Since the CYGNO's angular resolution is still under evaluation, we assumed an angular resolution of 30° in the whole detectable range, as it is suggested in literature [222] and by the CYGNUS simulation [184].

A flat prior distribution of the number of expected signal events μ_s, which is one of the parameter of the likelihood, between 0 and 1000 is used: the number of events is a non-negative defined variable, and the upper limit of 1000 events is considered to be reasonable, since, due to the current results in the DM direct detection searches, it is hardly believable that more than 1000 events per year would be produced in the CYGNO detector.

The number of expected background events for CYGNO PHASE_2 can not easily be predicted at this stage of the project. In order to address this problem,

different possible background hypotheses, exploring different orders of magnitude, are simulated, corresponding to 100, 1000 and 10000 events. The sensitivity projections for a greater background event rate can be easily obtained by extrapolating the results of this analysis. For the number of background events μ_b, another parameter of the likelihood, a Poisson prior is adopted.

For each possible value, the actual number of events in the pseudo-data spectrum is randomly extracted from a Poisson distribution. For each of these events, a direction is also randomly sampled from the background angular isotropic distribution, and a resolution Gaussian smearing is applied. To avoid possible biases induced by the fluctuations of the extracted pseudo-dataset, 500 data samples has been analyzed, and the average 90%C.I. upper limit on the signal cross-section has taken as the final value.

The likelihood of the detected events, evaluated on each data sample, is defined as

$$p(\boldsymbol{D}|\mu_s,\mu_b) = (\mu_b+\mu_s)^{N_{evt}} e^{-(\mu_b+\mu_s)} \prod_{i=1}^{N_{\text{bins}}} \left[\left(\frac{\mu_b}{\mu_b+\mu_s}P_{i,b} + \frac{\mu_s}{\mu_b+\mu_s}P_{i,s} \right)^{n_i} \frac{1}{n_i!} \right]$$

(8.1)

where

- $D = \{n_i\}$ is the dataset;

- n_i is the number of events of the i-th bin;

- N_{evt} is the total number of events of the data sample;

- i is an index running over the bins of the histogram in the 2D angular Galactic coordinates;

- μ is the expected signal (μ_s) or background (μ_b) events;

- $P_{i,x}$ is the probability that a certain event is observed in the i-th bin, according to the model x (background or signal): this marginalized probability takes into account the theoretical angular distribution, the smearing caused by resolution effects, and which element is hit during the primary interaction, see 8.2.

The posterior probability is computed using `JAGS`.

The top part of Fig. 8.3 shows the expected limits on the SI DM cross-section for a 30 m^3 CYGNO PHASE_2 experiment with 3 year exposure in the different background scenarios defined above, assuming a 1 keV energy threshold. The lower part of Fig. 8.3 is instead obtained assuming a threshold of 0.5 keV; the plot also shows the results for an hydrogen-rich gas mixture with 2% isobutane (C_4H_{10}) content.

The different nuclear composition of the gas mixture is reflected by the shape of the limits in Fig. 8.3. The sensitivity is such that, as expected, the lower DM mass that can be detected is reached in the 0.5 keV$_{ee}$ energy threshold configuration.The kink on the curve at ~ 0.9 GeV/c^2 corresponds to the transition from hydrogen dominated to helium dominated recoils, and, at 3 GeV/c^2, from helium to fluorine dominated recoils. Since the Carbon amount (8%) is too small, it cannot produce an appreciable effect on the sensitivity curve.
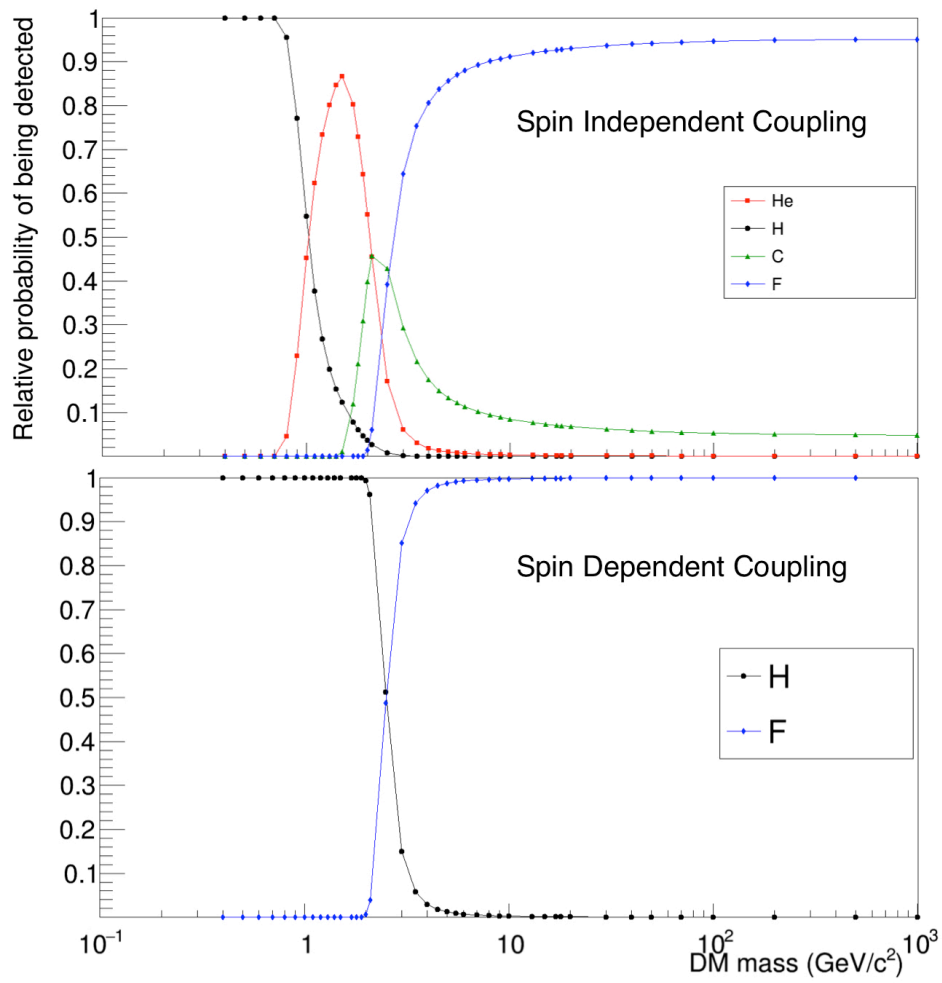
Figure 8.2. Relative NR detection probability as a function of the DM mass for different target elements and for both the SI (**top**) and SD (**bottom**) couplings. These results have been obtained including the quenching factors and with an energy threshold of 1 keV$_{ee}$. Figure from Ref. [208].
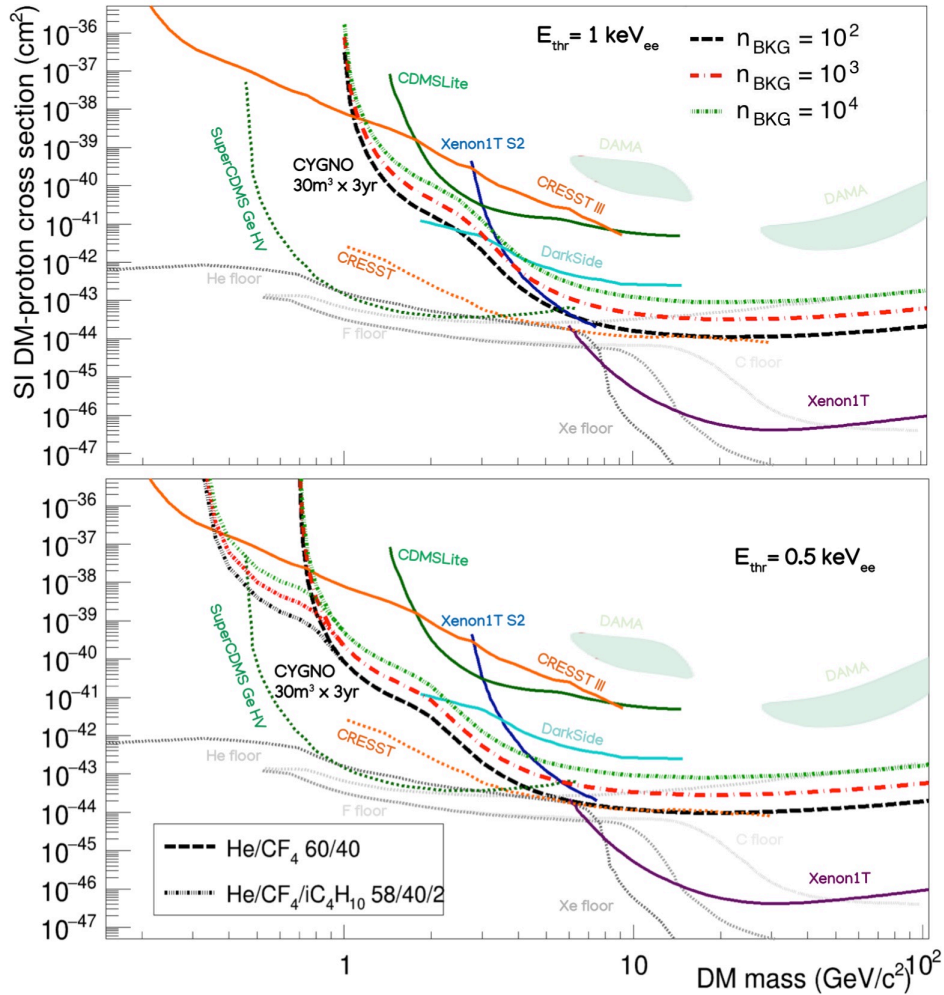
Figure 8.3. Sensitivity to SI DM-nucleon cross-section for 30 m^3 CYGNO detector with 3 years exposure for different background level assumptions. The **top** plot has been obtained with a 1 keV threshold, while the **bottom** plot has been obtained with a 0.5 keV threshold. The dashed curves correspond to a HeCF$_4$ 60/40 gas mixture with N$_{bkg}$ = 100 (black), 1000 (red) and 10000 (dark green), while the dotted curves correspond to a HeCF$_4$:isobutane 58/40/2 gas mixture. In the plot the current bounds from XENON1T (violet) [61], XENON1T S2 analysis (blue) [164], DarkSide (cyan) [1], CRESST-III (orange) [223] and CDMSLite (green) [224] are reported. The densely dotted curves show the future expected limits of SuperCDMS Ge (green) [225] and CRESST (orange) [226]. The light gray regions are the DAMA results [227], and the different gray curves are the neutrino floor for different target elements [228]. Figure from Ref. [208].

Figure 8.3 shows a promising sensitivity in every configuration considered here. Therefore, due to its directional capabilities, CYGNO could provide the galactic origin of the detected signal, providing a decisive evidence of a positive identification of a DM signal.

Due to the presence of fluorine in the gas mixture, CYGNO PHASE_2 is expected to be also sensitive to the SD interactions. The upper plot of Fig. 8.4 displays the expected limits on the SD DM cross-section, showing the potential to put constraints in a DM mass region which has not been yet excluded by the PICO experiment [229]. Finally, the possibility of having a 0.5 keV$_{ee}$ threshold and adding a small amount (2%) of a hydrogen-rich gas like isobutane would allow having access to a low DM mass region which will be inaccessible even by the future upgrade of PICO.

The preliminarily estimated sensitivities of CYGNO PHASE_2 suggest how this experiment could have a very important role for the dark matter direct detection searches in the low mass region, both for the SI and the SD interactions [208].

8.2 Sensitivity projections for the DarkSide-20k experiment

The DarkSide-20k experiment will be the next upgrade of the DarkSide project. It will be built and commissioned in Hall C at LNGS, it will operate for a minimum of 10 years, and much of its design parameters are based on the successful experience of constructing, commissioning and operating the DarkSide-50 experiment, in a background-free mode. The detector is designed to have a dual phase LAr TPC, characterized by a 51.1 t underground LAr mass, much larger than the (46.4 ± 0.7) kg constituting the DarkSide-50 TPC. Also in this case, the overall structure will be concentric, with the TPC surrounded by the neutron veto. However, as it is shown in Fig. 8.5, a novelty with respect to DarkSide-50 is that the neutron veto and the TPC are integrated into a single mechanical unit, which is depicted in detail in Fig. 8.6. This unit is in a common bath of low radioactivity argon, and is separated from the atmospheric argon of the main cryostat by a titanium vessel.

The material used for the neutron veto will be the Gadolinium-loaded polymethylmethacrylate (Gd-PMMA), and layers of the Gd-PMMA will be placed around the entire TPC. In this material, the neutron capture process is highly efficient, with a capture resulting in the emission of several photons with a total energy of 7.9 MeV. Both the thickness and the concentration of Gadolinium in the Gd-PMMA are designed to achieve a neutron capture inefficiency of $< 1\%$. The Gd-PMMA are positioned in an octagonal shape made up of 8 vertical panels of 15 cm thickness and two end caps of 15 cm thickness. The 8 vertical panels also have the role of lateral walls of the TPC.

Outside the Gd-PMMA layers, between them and the titanium vessel the space is filled with a 40 cm of UAr, to ensure the production of enough scintillation light from the neutron capture γ-rays in the case in which they do not reach the TPC. All the surfaces in contact with this outer UAr will be covered by a reflective foil and a wavelength shifter. In order to detect the scintillation light produced in this UAr volume, the surface of the Gd-PMMA is instrumented with an array of Silicon
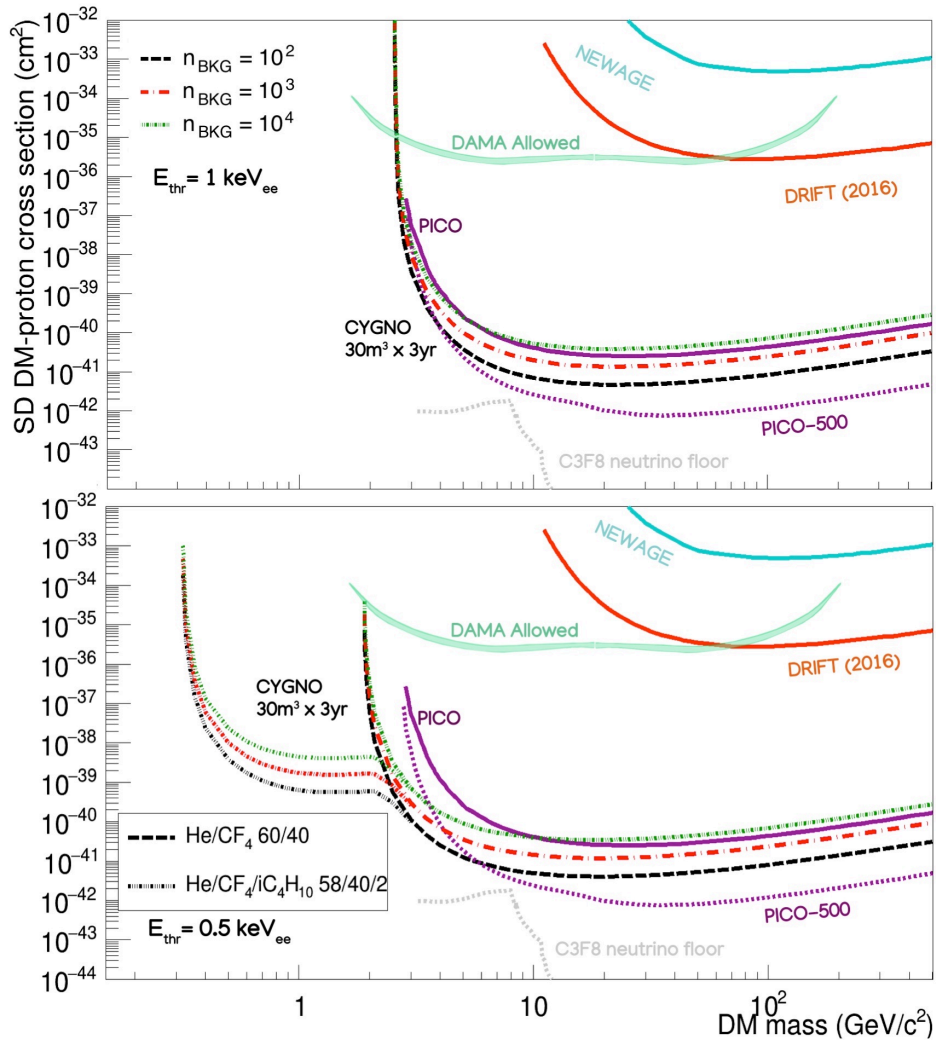
Figure 8.4. Sensitivity to SD DM-proton cross-section for 30 m^3 CYGNO detector with 3 years exposure for different background level assumptions. The **top** plot has been obtained with a 1 keV threshold, while the **bottom** plot has been obtained with a 0.5 keV threshold. The dashed curves correspond to a HeCF$_4$ 60/40 gas mixture with N$_{bkg}$ = 100 (black), 1000 (red) and 10000 (dark green), while the dotted curves correspond to a HeCF$_4$:isobutane 58/40/2 gas mixture. In the plot the current bounds from PICO (purple) [229], DRIFT (orange) [230], NEWAGE (cyan) [231] are reported. The light gray region is the DAMA result [227], and the dotted line is the neutrino floor for C$_3$F$_8$ [229]. Figure from Ref. [208].
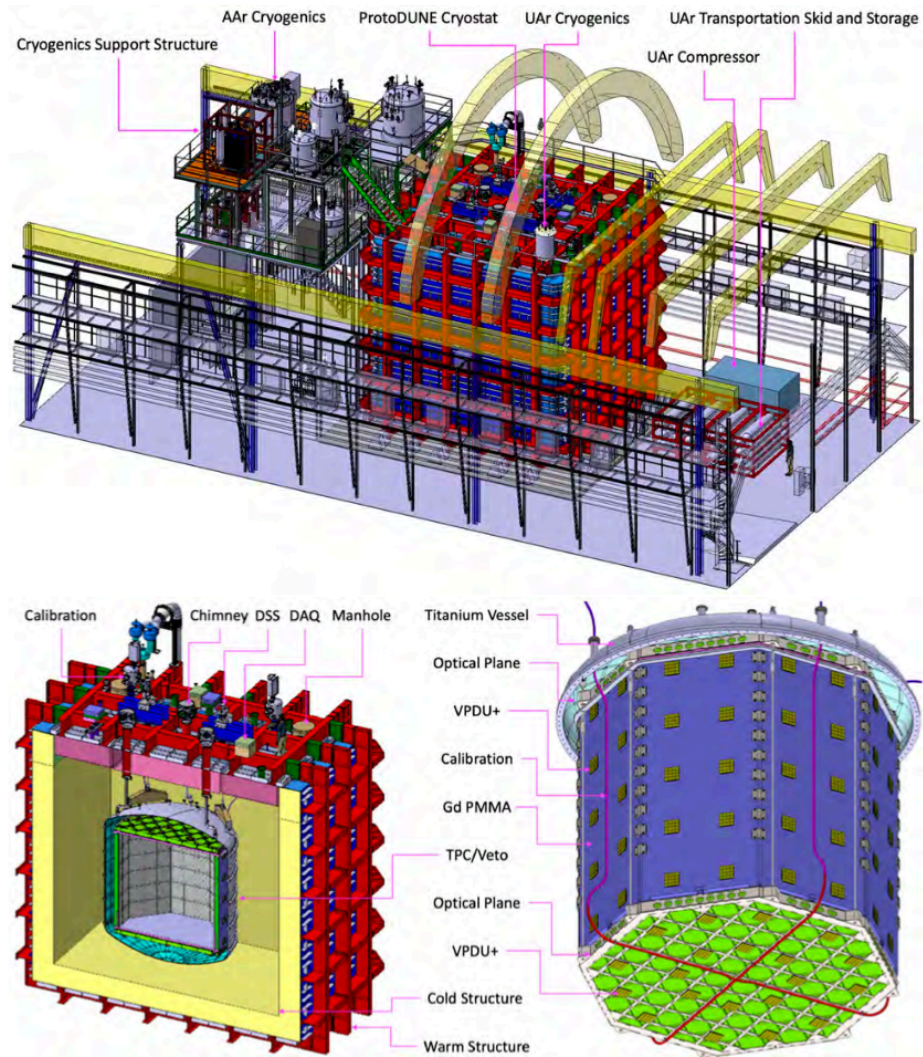
Figure 8.5. **Top**: Rendering of the DarkSide-20k experiment, which will be placed in Hall C at LNGS. The red ProtoDUNE-like cryostat will contain the detector and the cryogenic system. **Bottom left**: cross-section of the cryostat, with the Gd-PMMA in green and the titanium vessel in gray. **Bottom right**: external view of the TPC and veto with full assembly features. Figure from Ref. [232]
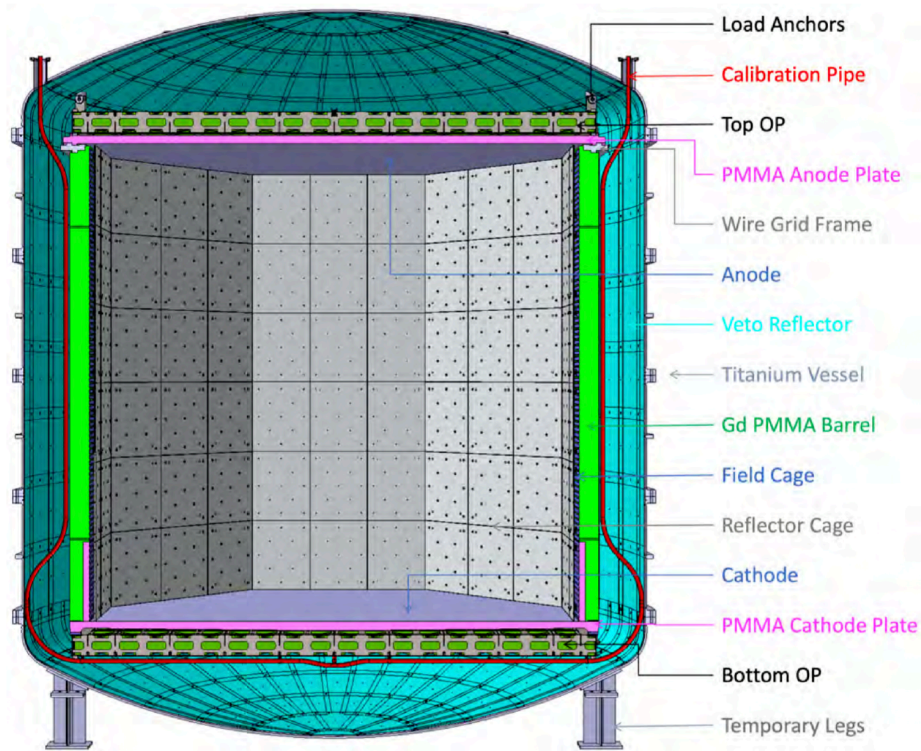
Figure 8.6. Drawing of the inner detector. The TPC and the Veto are contained in the titanium vessel hosting the UAr target. Figure from Ref. [232]

PhotoMultipliers (SiPMs), for a total SiPM geometrical coverage of 3.2 %.

The TPC volume is therefore a prism with a right octagonal base. The height of the prism is designed to be 348 cm, while the base is such that the circle inscribed in the octagon has a diameter of 350 cm. The top and the bottom caps of the TPC are made of pure, transparent, acrylic (PMMA). The electrodes and the drift field in the TPC will be obtained by coating the inside of the bottom cap (cathode), the inside of the top cap (anode) and the inside walls (field cage) with Clevios, a transparent conducting polymer. Just below the anode, a gas phase of 7 mm thickness will be maintained, and, below the liquid surface, a grid of wires will act as the extraction grid. The inner walls are then covered with high reflectivity reflectors, and all the inner surfaces are coated with TPB acting as the wavelength shifter. Finally, the scintillation light produced inside the TPC is detected by planes of SiPMs placed between the PMMA caps and the Gd-PMMA caps. This is one of the greatest differences with respect to DarkSide-50, which, instead, was instrumented with PMTs.

The sealed, vacuum capable, titanium vessel containing the detector will be immersed within a bath of 700 t of liquid AAr, which will act as shield and outer veto detector. The AAr is contained in a large ProtoDUNE-like membrane cryostat, see Ref. [233] for details.

DarkSide-20k is designed to have ultra-low background. Besides the external background coming from the cave and the residual cosmic rays, which will be vetoed and removed by the outer ProtoDUNE-like cryostat, other background sources will be the residual radioactive isotopes present in the materials used to build the detector. Whereas the β and α particles produced by the decay of a radioactive isotope can contribute to the background only if it is in contact with the LAr target, the gammas and the neutrons produced after a spontaneous fission or an (α, n) interaction can induce events inside the TPC from more distant sites. In addition to this internal contamination, additional background contributions could be induced by cosmogenic activation of the materials during manufacturing, storing or transportation.

The experiment is foreseen to reach a very high sensitivity to DM-nucleon cross-section. The most recent projection assumes a 10 yr run, corresponding to a fiducial volume exposure of 200 t yr and is reported in Fig. 8.7. The radiogenic NR background spectra in the fiducial volume used for the analysis, estimated from current data and dedicated MC simulations, are reported in Tab 8.1, and also the irreducible neutrino background from coherent scattering off nuclei has been included. This preliminary analysis, restricted to the nominal fiducial volume, shows that the projected sensitivity is 6.3×10^{-48} cm^2 for the 90% C.L. exclusion limit and 2.1×10^{-47} cm^2 for the 5σ discovery of a 1 TeV/c^2 DM particle. DarkSide-20k is foreseen to start its data taking in 2025.
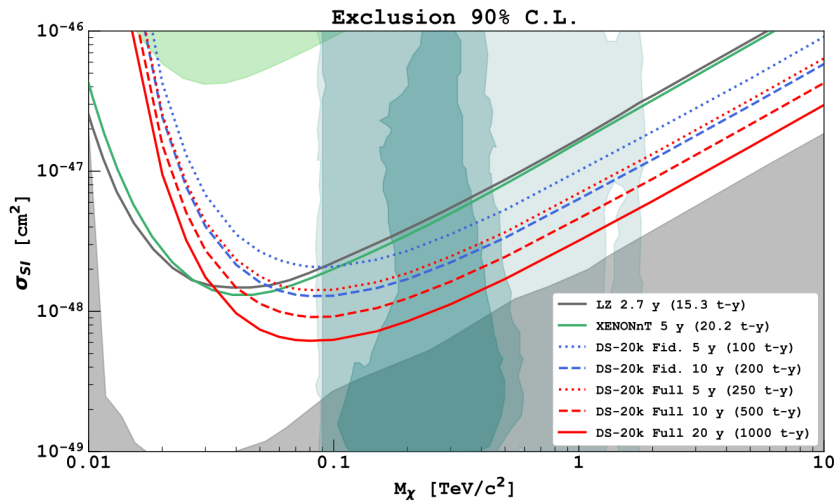
Figure 8.7. Sensitivity projections of DarkSide-20k to DM spin independent interactions for hypothetical different lengths of runs shown compared to the nominal sensitivity of currently funded experiments LZ (2.7 yr run, 15.3 t yr exposure [234]) and XENONnT (5 yr run, 20 t yr exposure [235]) that are expected to have the highest performances in the next few years. The run time of DS-20k will be 10 yr, corresponding to a fiducial volume exposure of 200 t yr. The green shaded area represents the XENON1T [61] excluded region, the gray shaded region is the neutrino floor for argon [236], and the turquoise filled contours are the 1, 2, and 3σ favored regions constrained by astrophysical measurements and the LHC data at 13 TeV [237]. Figure from Ref. [232].

Background	Bg events $[200 \, \text{t yr}]^{-1}$
(α,n) neutrons from U and Th	9.5×10^{-2}
Fission neutrons from ^{238}U	$< 2.3 \times 10^{-3}$
Neutrons from ^{222}Rn diffusion and surface plate-out	$< 1.4 \times 10^{-2}$
Cosmogenic neutrons	$< 6.0 \times 10^{-1}$
Neutrons from the lab rock	1.5×10^{-2}
Random surface α decay + S2 coincidence	$< 5.0 \times 10^{-2}$
Correlated ER + Cherenkov	$< 1.8 \times 10^{-2}$
Uncorrelated ER + Cherenkov	$< 3.0 \times 10^{-2}$
ER	$< 1.0 \times 10^{-1}$

Table 8.1. Radiogenic NR background levels expected in the fiducial volume during the full DarkSide-20k exposure, estimated from the current data and dedicated MC simulations. The results are given in terms of events surviving the veto cut, the fiducial volume cut and the Pulse Shape Discrimination (PSD).
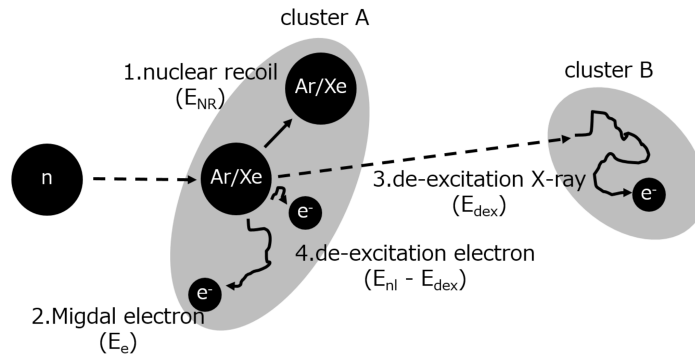
Figure 8.8. Graphical representation of the signature proposed in Ref. [239] to measure the Migdal effect. Figure from Ref. [239].

8.3   Measuring the Migdal effect with CYGNO

In Chap. 3 we illustrated the Migdal effect and its importance in the context of the direct detection of low mass DM candidates. However, even if the Migdal effect is expected in the Standard Model, it has not been measured yet in nuclear scattering: a compelling question is therefore if direct detection experiments are able to observe it. Indeed, several DM experiments [62, 66, 87–92] have recently started to exploit the Migdal effect including it in their analyses and extending their reach into the sub-GeV mass regime. Similar to DM searches, the exploitation of the Migdal effect in the coherent neutrino-nucleus scattering process can increase the sensitivity to Solar neutrinos detection [3]. For all these reasons we believe the experimental observation and measurement of the Migdal effect is of paramount importance and timely for current and future DM searches and solar neutrino spectroscopy. Parallel to other experimental proposals [238] to measure this effect, we believe that the experimental approach developed within the CYGNO experiment for directional DM searches to be nicely suited (with a dedicated optimisation) to perform this measurement.

   In this section, we will then investigate the possibility of measuring the Migdal effect with a detector based on the CYGNO approach exposed to a neutron source. In fact, since the measurement requires a detector with a good 3D position and energy resolution, we will focus on the possibility of performing the measurement at LNF or at CERN with the CYGNO PHASE_0 detector, i.e. the 50 L LIME prototype, which fulfills these requirements.

   The signatures that can be exploited for this purpose are mainly two. The first one, that was proposed in Ref. [239], is illustrated in Fig. 8.8 and can be explained as follows:

1. the elastic scattering of the neutron on an heavy nucleus like Ar or Xe can induce a nuclear recoil with energy E_{NR};

2. a Migdal electron with energy E_e can be emitted: the recoiling nucleus and the Migdal electron are detected as a single cluster A;

3. if the emitted Migdal electron comes from the K-shell, a monochromatic X-ray of energy E_{dex} can be emitted as a consequence of the de-excitation, with a probability given by the K-shell fluorescence yield of the target atom: the X-ray is absorbed in a distance of $\mathcal{O}(\text{cm})$ from cluster A, resulting in another detected cluster B;

4. additional multiple Auger electrons and X-ray photons, with a total energy of $E_{nl} - E_{\text{dex}}$, are emitted for energy conservation, and are indistinguishable from cluster A.

For the Ar atoms, the monochromatic X-ray has an energy $E_{\text{dex}} \sim 3\,\text{keV}$, with a fluorescence yield $f_{\text{Ar}} = 0.14$, while, following as usual Ref. [3], the Migdal branching ratio is expected to be 7.2×10^{-5} assuming $q_e = 511\,\text{eV}/c$. For what regards the characteristics of the incoming neutron, several possibilities are being considered as neutron sources. A good candidate is the Frascati Neutron Generator (FNG) Facility [240] at LNF, which can produce $14.1\,\text{MeV}$ neutrons with a rate of $10^{11}\,n/\text{s}$ from deuteron-deuteron reactions or $2.5\,\text{MeV}$ neutron with a rate of $10^9\,n/\text{s}$ from deuteron-triton reactions. Besides having the possibility of using different neutron energies, in order to optimize the measurement sensitivity it would be possible to regulate the distance of the target from the neutron source, which in turn is located in a $130\,\text{m}^2$ hall to minimize the neutron background due to the reflections. From a quantitative point of view, assuming the 2.5 MeV neutron source, we expect, for the X-ray signature, the following number of events,

$$N_{\text{events}} = N_T\,\Phi\,\sigma_{\text{Ar}}\,f_{\text{Ar}}\,q_e^2\,\text{BR}_{\text{ME}} \simeq 400\,\text{day}^{-1}, \tag{8.2}$$

where N_T is the number of target nuclei, Φ is the flux for a 2.5 MeV neutron source at a distance of around 3 m, σ_{Ar} is the n-Ar scattering cross-section for a 2.5 MeV neutron, f_{Ar} is the fluorescence yield, q_e is the electron momentum in the nucleus rest frame immediately after the neutron collision, and BR_{ME} is the Migdal branching ratio (considering only the most internal shell). With this expected rate, as highlighted in Ref. [239], the expected exponential distribution for absorption length of the X-ray should help to reject flat neutron multiple scattering or the distribution for accidental backgrounds. However, as recognized in Ref. [239], due to the very high spatial resolution that can be reached with LIME, if the Migdal electron is emitted with enough energy to produce an ER track distinguishable from the NR cluster, the directionality could be used as additional information to reduce both the gamma and the neutron backgrounds: the angular distribution of the Migdal electron from the K-shell indeed goes as $\cos^2\phi$, where ϕ is the polar angle with respect to the direction of the recoiling nucleus.

Another good candidate as a neutron source is neutron beam of the neutron Time-Of-Flight facility (n_TOF) [241] at CERN. This facility provides a neutron pulse every 1.2 s, or multiples of this intervals, emitted as a consequence of the spallation reactions induced by a 7×10^{12} protons $20\,\text{GeV}/c$ pulse hitting a fixed lead target. The charged particles produced during the collisions are deflected away by means of a magnet field, and the neutron beam is shaped thanks to two collimators along the beam line. In the $\mathcal{O}(0.1 - 10\,\text{MeV})$ energy range, the neutron flux of the facility, as determined by the n_TOF collaboration, is of the order of 10^6 neutrons
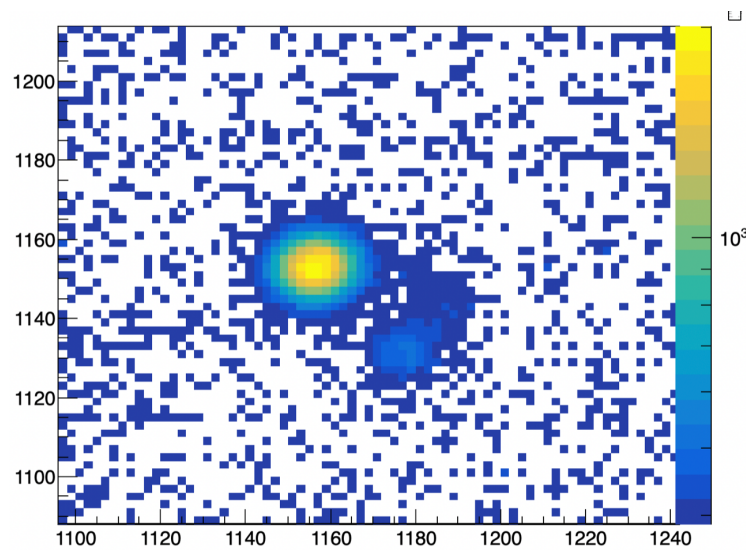
Figure 8.9. Simulated event with a 100 keV He nuclear recoil and a 30 keV electronic recoil starting from the same vertex.

per bunch. A prompt component of ionizing particles arrives at the experimental area, which is located at 182.3 m from the spallation target and has a length of 7.9 m, along with the neutron beam. These particles are mainly the photons and the residual charged particles produced at the spallation target, and the halo of photons and charged particles produced by the interactions of the neutrons and high energy photons with the second collimator. The interesting feature is that all these particles are observed as a prompt signal, with a time of flight below $1\,\mu$s. It is therefore in principle possible to exploit the bunched structure of the neutron beam for a proper DAQ and trigger scheme, in order to greatly reduce the background coming along with the neutrons. In particular, it would be possible to trigger the sCMOS camera acquisition synchronizing it with the beam clock. For the X-ray signature at n_TOF we expect $N_{\mathrm{events}} \sim 480\,\mathrm{day}^{-1}$.

The second possible signature is just the detection of the Migdal electron, exploiting all the atomic shells. This signature is detected as a NR track and an ER track starting from the same vertex. The Migdal emission probability in this case is larger, since now we are sensitive to the contribution of the external shells. In addition, since the X-ray emission is not required, a He target could be used. However, this needs good tracking capabilities to be able to resolve events like the one shown in Fig. 8.9, in which a MC simulated event with a 100 keV He nuclear recoil and a 30 keV electronic recoil starting from the same vertex is illustrated. In general, this should be possible for large nuclear recoil energies and for electron recoil energies down to $\sim 5-10$ keV. The right plot of Fig. 8.10 shows the number of events per day as a function of the nuclear and electron recoil energy for a mixture 60:40 of Ar:CF$_4$ and a neutron source of 2.5 MeV at 3 m. Integrating from a nuclear recoil energy of 100 keV to E_R^{max} and from E_e of 5 keV to E_e^{max}, we obtain ~ 600 events per day. The left plot of Fig. 8.10 shows similar results for a 60:40 mixture of He:CF$_4$, leading to ~ 350 events.

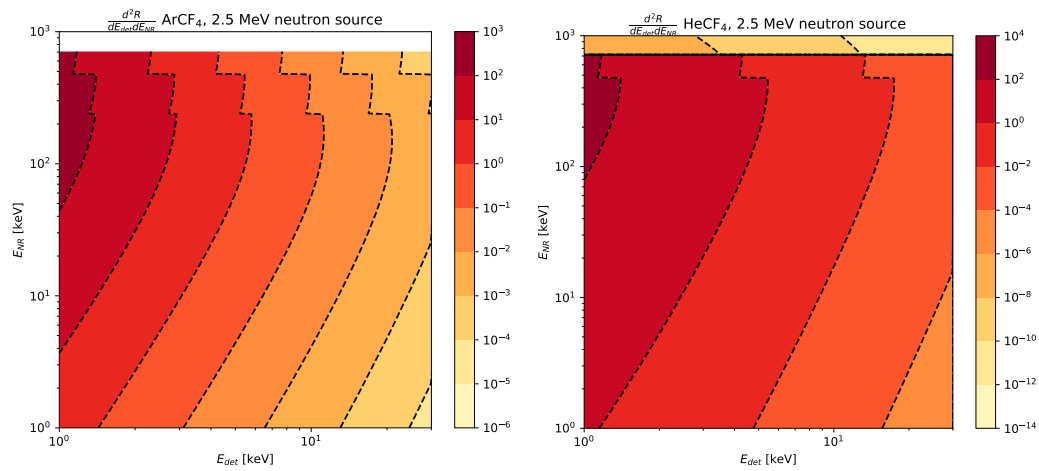In order to get the achievement of performing the measurement of the Migdal

Figure 8.10. Double differential rate for neutron scattering on ArCF$_4$ (**left**) and HeCF$_4$ (**right**).

effect, some work is therefore needed:

- **to assess the real feasibility of the measure by means of the detector simulations**: we are working on signal and background simulations in order to define the optimal detector configuration. We already developed a simplistic simulation of a possible experimental setup. Such a simulation will be further refined including a detailed description of the detector, of the experimental environment, and of the background sources.

- **to choose the most suitable gas mixture**: in fact, from first order computations, with respect to the He-based mixture which is currently employed for the tests by the collaboration, heavier noble gasses seems to be preferred. Tests with Ar:CF$_4$ mixture in LIME are currently going on in order to characterize the detector response with a different gas mixture.

- **to choose the best solution for the neutron source**: in fact, even if for our previous estimates we considered the 2.5 MeV FNG source and the n_TOF facility at CERN, an AmBe source is also going to be taken into account. For all these neutron sources, a photon background (which is the main background for the measurement), with an intensity depending on the specific source, is expected. Therefore, a more detailed study, with the help of some simulation, is needed to make the right choice.

- **to develop an appropriate trigger system and analysis strategy**: since the process we are looking at is very rare, a well suited trigger system is needed to select only the desired events and remove the expected high number of background events. In particular, a refined analysis strategy to tag the events is needed, and this would be even more challenging if the He:CF$_4$ gas mixture is chosen, since in that case the X-ray signature will be absent.

Conclusions

Establishing the nature and the properties of dark matter is one of the most open and discussed problems of contemporary physics. However, up to this day, no incontrovertible evidence of dark matter particles has been found yet. Several and very different experimental techniques have been developed in order to be able to detect dark matter on Earth. The direct detection experiments, studying the possible recoils that the dark matter particles should induce mainly in the atomic nuclei, produced very strong constraints on the dark matter interaction cross-section. In this context, many experimental approaches have been developed, each of them giving their best performances for different possible values of the dark matter mass. The region explored with the highest sensitivity by the noble liquids TPC experiments is the $m_{DM} \gtrsim 1 - 10 \, \text{GeV}/c^2$ region.

Within this context, we contributed to the conclusive analysis of the DarkSide-50 experiment. For this purpose, we developed an innovative analysis tool in the Bayesian approach that is able to reproduce the detector response in a semi analytical way. With respect to the previous 2018 analysis [1], the DarkSide-50 observed sensitivity benefits from the better understanding of the the detector response model, the background contaminating the experiment, and the optimized data selection. The result presented in this thesis represents the world's best bound on DM spin independent (SI) cross-section in the $m_{DM} = [0.7, 4.7] \, \text{GeV}/c^2$ mass region, being able to exclude at 90% Credible Interval a cross-section $\sigma_{DM}^{SI} = 2.2 \times 10^{-43} \, \text{cm}^2$ at $m_{DM} = 4.5 \, \text{GeV}/c^2$. In addition, introducing the Migdal effect in the analysis, we were able to extend the sensitivity of the experiment to lower dark matter masses, up to 60 MeV/c^2, where, again, DarkSide-50 revealed to currently set the world's strongest bound. Indeed, the scattering of dark matter particles in the low mass range is a challenging task: for a DM particle with a mass in the sub-GeV region, the typical energy transfer is below the experimental threshold. Nevertheless, due to the Migdal effect, after a nuclear recoil induced by dark matter, as a consequence of the perturbation of the electron cloud, an electron of the recoiling atom can be excited and ionized. We computed the Migdal effect rate for hydrogen, helium and argon atoms, demonstrating that, in the specific case of a liquid argon target, the Migdal effect is able to extend the experimental sensitivity to lower DM masses by more than an order of magnitude, up to tens of MeVs.

This analysis, which focused on WIMP-like DM induced interactions, will be extended to other possible kind of interactions, as we mentioned in Sec. 6.6. In parallel, the collaboration is working on the development and construction of the next upgrade of the experiment, called DarkSide-20k, which, with an expected exposure of 200 tyr assuming a 10 yr run, will allow reaching a $\sim 10^{-48} \, \text{cm}^2$ DM SI

sensitivity at $\mathcal{O}(100\,\mathrm{GeV}/c^2)$ DM masses.

Finally, the CYGNO experiment, based on a gaseous TPC detector with optical readout, will be a complementary experiment with respect to the DarkSide experiment: thanks to its lighter target atoms and high 3D position resolution, it will be able to access the direction of the nuclear scatterings and potentially distinguish the DM induced recoils from the irreducible neutrino coherent scattering background. Thanks to its good properties, we believe that the CYGNO approach is suitable for a measurement of the Migdal effect induced by neutrons. Indeed, up to this day, the Migdal effect has been measured only in α and β decays, and its verification in nuclear scattering would be a strong confirmation of the results obtained by the direct detection experiments. However, the feasibility of this measurement must be furthermore studied, and in particular, since the process we are looking for is predicted to be very rare, the following steps will be pursued: developing dedicated and detailed detector simulations; choosing the most suitable gas mixture; choosing the best solution for the neutron source to be used for the measurement; developing an appropriate trigger system and analysis strategy. The CYGNO project is currently finishing its R&D phase, and the collaboration is planning to build a large detector with an $\mathcal{O}(30-100\,\mathrm{m}^2)$ volume which is expected to have a competitive sensitivity in the $m_{DM} \lesssim 10\,\mathrm{GeV}/c^2$ mass region, as we showed with the analysis illustrated in Sec. 8.1.

We expect that, with the commissioning of better and better experiments, in the next decades dark matter direct detection searches will be able to exclude DM interaction cross-sections of $\mathcal{O}(10^{-48}\,\mathrm{cm}^2)$ in the $O(10-100\,\mathrm{GeV}/c^2)$ DM mass region. With such sensitivities, the new experiments would have to deal with the background from the coherent scatterings induced by the solar neutrinos, and thus the capability of measuring the directionality of the recoils, which is a significant feature of experiments like CYGNO, will be an important handle to discriminate this contribution from a DM signature.

Ringraziamenti

I ringraziamenti li scrivo in italiano, per essere sicuro di esprimere al meglio la mia gratitudine alle persone che mi sono state vicine durante questo percorso.

Prima di tutto vorrei ringraziare il mio relatore, il prof. Andrea Messina, per tutto quello che ha fatto per me, e per avermi dato l'opportunità di fare ricerca su qualcosa che mi appassioni veramente. Grazie al suo aiuto ho imparato molto sia dal punto di vista professionale che umano. Mi ha introdotto con infinita pazienza e dedizione al mondo della fisica sperimentale, insegnandomi tutto ciò che so sulla materia oscura, l'analisi statistica dei dati, i sistemi di trigger e data acquisition, introducendomi alle collaborazioni DarkSide e Cygno e permettendomi di fare esperienza nel mondo della didattica grazie alle attività del corso di Laboratorio di Meccanica. È stato fonte di supporto nei momenti più difficili, sempre con una parola gentile o con una battuta divertente per stemperare la situazione. Inoltre, mi ha insegnato a sciare, in quel caso, per colpa dell'allievo, non con molto successo.

Un sentito ringraziamento va a Giovanni Grilli di Cortona, che mi ha introdotto per la prima volta all'effetto Migdal, e con il quale ho avuto il piacere di collaborare in questi tre anni. Ringrazio anche lui per la pazienza che ha avuto nel rispondere ai miei dubbi sulla fisica e per l'aiuto che mi ha dato nello sviluppo delle analisi. Auguro il meglio a lui e alla sua famiglia.

Vorrei ringraziare profondamente il prof. Marco Nardecchia, per le utilissime chiacchierate sui problemi di Meccanica Quantistica e non, e per essere sempre coinvolgente e divertente in tutto quello che fa. Lo ringrazio per tutti gli utili consigli, dal punto professionale e umano, e per avermi aiutato a capire cosa fare da grande. Auguro tutto il bene possibile a lui e alla sua famiglia. Io e Migliorati non vediamo l'ora della briscolata tutti insieme.

Infine, vorrei spendere un paio di parole per ringraziare tutte le persone con cui ho avuto il piacere di collaborare professionalmente e che sono sempre state gentili con me: ringrazio Valerio Ippolito e il prof. Sandro De Cecco per l'aiuto che mi hanno dato nell'integrarmi nella collaborazione DarkSide e nella comprensione dell'esperimento e della precedente attività della collaborazione; ringrazio Davide Pinci e la prof. Elisabetta Baracchini per avermi accolto calorosamente in Cygno, e per avermi permesso di partecipare alle attività di R&D in LIME; infine ringrazio il collega Giorgio Dho, insieme al quale ho avuto modo di lavorare allo studio preliminare sulla sensitività del futuro esperimento e con il quale condivido la passione per i manga e, nello specifico, per One Piece.

Un saluto speciale va a tutti i miei amici e colleghi, con i quali ho trascorso storiche serate che non dimenticherò mai e che mi hanno permesso di non impazzire del tutto in questi tre anni. In particolare, ringrazio tutti i colleghi della stanza

117/118 all'edificio Fermi, per la loro gentilezza e disponibilità a risolvere i problemi (e ce ne sono stati parecchi in questi anni, soprattutto durante il lockdown): una menzione speciale va a Gabriele Barca, ispirato Dungeon Master e incredibile compagno di bevute. Ringrazio i miei amici e colleghi storici Andrea Sabatucci, detto Er Pertica, Lorenzo Roberti e Alessandro Trinca con i quali ho passato gran parte degli anni universitari, innumerevoli birre, e a cui auguro una brillante carriera. Senza di loro l'università non sarebbe stata il luogo spensierato, stimolante e familiare che ho vissuto. Ringrazio i miei amici Franco Risi, Francesco Ritarossi, il mitico Tomm (con *almeno* due "m") Renzi e Marco Terziarol del server Discord Quarantena Delocalizzata: è grazie a loro se non ho perso la testa durante il lockdown, e se ho imparato che il miglior posto dove atterrare dopo essersi paracadutati da un aereo in volo è l'ospedale. Infine ringrazio il mio migliore amico Simone "Sghì" Battaglia, per essermi sempre stato vicino fin dal primo anno di liceo, soprattutto nei momenti peggiori, e per tutte le idee pazze che abbiamo realizzato insieme. Non vedo l'ora di organizzare con lui la prossima abbuffata.

Infine, vorrei ringraziare la mia famiglia per tutto quello che hanno fatto e che fanno ogni giorno per me. Se oggi sono quello che sono lo devo ai miei genitori, ai loro insegnamenti, al loro impegno, al loro supporto e al loro amore. Sono fiero di mia sorella Francesca, e sono fortunato a condividere la vita con una persona come lei, che condivide molte delle mie passioni e che è la mia migliore amica. Infine, anche se è l'ultimo arrivato, vorrei rivolgere un pensiero affettuoso al mio cane Whisky, per l'enorme affetto che dimostra verso di me e per il sorriso che riesce sempre a strapparmi, anche nei momenti più difficili. Sono orgoglioso di quelli che siamo, mai disuniti nei momenti peggiori e sempre insieme quando si tratta di affrontare una difficoltà. Vi amo.

Bibliography

[1] **DarkSide** Collaboration, P. Agnes et al., *Low-Mass Dark Matter Search with the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018), no. 8 081307, [`arXiv:1802.06994`].

[2] G. Grilli di Cortona, A. Messina, and S. Piacentini, *Migdal effect and photon Bremsstrahlung: improving the sensitivity to light dark matter of liquid argon experiments*, *JHEP* **11** (2020) 034, [`arXiv:2006.02453`].

[3] M. Ibe, W. Nakano, Y. Shoji, and K. Suzuki, *Migdal Effect in Dark Matter Direct Detection Experiments*, *JHEP* **03** (2018) 194, [`arXiv:1707.07258`].

[4] A. Messina, M. Nardecchia, and S. Piacentini, *Annual modulations from secular variations: not relaxing DAMA?*, *JCAP* **04** (2020) 037, [`arXiv:2003.03340`].

[5] **Planck** Collaboration, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [`arXiv:1807.06209`].

[6] V. C. Rubin and W. K. Ford, Jr., *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, *Astrophys. J.* **159** (1970) 379–403.

[7] K. C. Freeman, *On the disks of spiral and SO Galaxies*, *Astrophys. J.* **160** (1970) 811.

[8] K. G. Begeman, A. H. Broeils, and R. H. Sanders, *Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics*, *Mon. Not. Roy. Astron. Soc.* **249** (1991) 523.

[9] W. J. G. de Blok, F. Walter, E. Brinks, C. Trachternach, S.-H. Oh, and R. C. Kennicutt, Jr., *High-Resolution Rotation Curves and Galaxy Mass Models from THINGS*, *Astron. J.* **136** (2008) 2648–2719, [`arXiv:0810.2100`].

[10] J. Einasto, *On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters*, *Trudy Astrofizicheskogo Instituta Alma-Ata* **5** (1965) 87–100.

[11] J. F. Navarro, C. S. Frenk, and S. D. M. White, *A Universal density profile from hierarchical clustering*, *Astrophys. J.* **490** (1997) 493–508, [`astro-ph/9611107`].

[12] J. P. Ostriker and P. J. E. Peebles, *A Numerical Study of the Stability of Flattened Galaxies: or, can Cold Galaxies Survive?*, *Astrophys. J.* **186** (Dec., 1973) 467–480.

[13] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, *Helv. Phys. Acta* **6** (1933) 110–127.

[14] S. Smith, *The Mass of the Virgo Cluster*, *Astrophys. J.* **83** (1936) 23–30.

[15] B. Ryden, *Introduction to cosmology.* Cambridge University Press, 1970.

[16] J. A. Tyson, G. P. Kochanski, and I. P. Dell'Antonio, *Detailed mass map of CL0024+1654 from strong lensing*, *Astrophys. J. Lett.* **498** (1998) L107, [`astro-ph/9801193`].

[17] T. Treu, *Strong Lensing by Galaxies*, *Ann. Rev. Astron. Astrophys.* **48** (2010) 87–125, [`arXiv:1003.5567`].

[18] M. Kilbinger, *Cosmology with cosmic shear observations: a review*, *Rept. Prog. Phys.* **78** (2015) 086901, [`arXiv:1411.0115`].

[19] B. Paczynski, *Gravitational microlensing by the galactic halo*, *Astrophys. J.* **304** (1986) 1–5.

[20] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, *A direct empirical proof of the existence of dark matter*, *Astrophys. J. Lett.* **648** (2006) L109–L113, [`astro-ph/0608407`].

[21] "X-ray: NASA/CXC/CfA/M.Markevitch et al.; Optical: NASA/STScI; Magellan/U.Arizona/D.Clowe et al.; Lensing Map: NASA/STScI; ESO WFI; Magellan/U.Arizona/D.Clowe et al.."

[22] E. W. Kolb and M. S. Turner, *The Early Universe*, vol. 69. CRC Press, 1990.

[23] L. Perivolaropoulos and F. Skara, *Challenges for ΛCDM: An update*, 5, 2021.

[24] M. S. Turner and J. A. Tyson, *Cosmology at the millennium*, *Rev. Mod. Phys.* **71** (1999) S145–S164, [`astro-ph/9901113`].

[25] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *Astrophys. J.* **707** (2009) 916–920, [`arXiv:0911.1955`].

[26] **DES** Collaboration, T. M. C. Abbott et al., *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing*, 2022.

[27] **Particle Data Group** Collaboration, P. A. Zyla et al., *Review of Particle Physics*, *PTEP* **2020** (2020), no. 8 083C01.

[28] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, *Ann. Rev. Astron. Astrophys.* **48** (2010) 495–545, [`arXiv:1003.0904`].

[29] R. J. Scherrer and M. S. Turner, *On the Relic, Cosmic Abundance of Stable Weakly Interacting Massive Particles*, *Phys. Rev. D* **33** (1986) 1585. [Erratum: Phys.Rev.D 34, 3263 (1986)].

[30] R. Bernabei et al., *Dark matter particles in the Galactic halo: Results and implications from DAMA/NaI*, *Int. J. Mod. Phys. D* **13** (2004) 2127–2160, [`astro-ph/0501412`].

[31] R. Bernabei et al., *Final model independent result of DAMA/LIBRA-phase1*, *Eur. Phys. J. C* **73** (2013) 2648, [`arXiv:1308.5109`].

[32] R. Bernabei et al., *First model independent results from DAMA/LIBRA-phase2*, *Nucl. Phys. Atom. Energy* **19** (2018), no. 4 307–325, [`arXiv:1805.10486`].

[33] J. L. Feng and J. Kumar, *The WIMPless Miracle: Dark-Matter Particles without Weak-Scale Masses or Weak Interactions*, *Phys. Rev. Lett.* **101** (2008) 231301, [`arXiv:0803.4196`].

[34] T. Lin, H.-B. Yu, and K. M. Zurek, *On Symmetric and Asymmetric Light Dark Matter*, *Phys. Rev. D* **85** (2012) 063503, [`arXiv:1111.0293`].

[35] S. Knapen, T. Lin, and K. M. Zurek, *Light Dark Matter: Models and Constraints*, *Phys. Rev. D* **96** (2017), no. 11 115021, [`arXiv:1709.07882`].

[36] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[37] Y. Aoki et al., *FLAG Review 2021*, `arXiv:2111.09849`.

[38] J. F. Gunion, P. C. McNamee, and M. D. Scadron, *Chiral Symmetry Breaking and the Quark Model: Unification of Baryon and Meson Constraints*, *Nucl. Phys. B* **123** (1977) 445–492.

[39] R. D. Peccei, *The Strong CP problem and axions*, *Lect. Notes Phys.* **741** (2008) 3–17, [`hep-ph/0607268`].

[40] G. G. Raffelt, *Axions in astrophysics and cosmology*, in *30th Rencontres de Moriond: Euroconferences: Dark Matter in Cosmology, Clocks and Tests of Fundamental Laws*, pp. 159–168, 1995. `hep-ph/9502358`.

[41] **XENON** Collaboration, E. Aprile et al., *Excess electronic recoil events in XENON1T*, *Phys. Rev. D* **102** (2020), no. 7 072004, [`arXiv:2006.09721`].

[42] S. Roy Choudhury and S. Hannestad, *Updated results on neutrino mass and mass hierarchy from cosmology with Planck 2018 likelihoods*, *JCAP* **07** (2020) 037, [`arXiv:1907.12598`].

[43] S. Dodelson and L. M. Widrow, *Sterile-neutrinos as dark matter*, *Phys. Rev. Lett.* **72** (1994) 17–20, [`hep-ph/9303287`].

[44] R. E. Shrock, *Electromagnetic Properties and Decays of Dirac and Majorana Neutrinos in a General Class of Gauge Theories*, Nucl. Phys. B **206** (1982) 359–379.

[45] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, and S. W. Randall, *Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters*, Astrophys. J. **789** (2014) 13, [arXiv:1402.2301].

[46] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*. Princeton University Press, 2008.

[47] J. Bovy, "Dynamics and Astrophysics of Galaxies." https://galaxiesbook.org/.

[48] P. F. de Salas, K. Malhan, K. Freese, K. Hattori, and M. Valluri, *On the estimation of the Local Dark Matter Density using the rotation curve of the Milky Way*, JCAP **10** (2019) 037, [arXiv:1906.06133].

[49] P. F. de Salas and A. Widmark, *Dark matter local density determination: recent observations and future prospects*, 2021.

[50] M. J. Reid et al., *Trigonometric Parallaxes of Massive Star Forming Regions: VI. Galactic Structure, Fundamental Parameters and Non-Circular Motions*, Astrophys. J. **700** (2009) 137–148, [arXiv:0902.3913].

[51] A.-C. Eilers, D. W. Hogg, H.-W. Rix, and M. K. Ness, *The Circular Velocity Curve of the Milky Way from 5 to 25 kpc*, Astrophys. J. **871** (Jan, 2019) 120, [arXiv:1810.09466].

[52] T. Piffl et al., *The RAVE survey: the Galactic escape speed and the mass of the Milky Way*, Astron. Astrophys. **562** (2014) A91, [arXiv:1309.4293].

[53] G. Monari, B. Famaey, I. Carrillo, T. Piffl, M. Steinmetz, R. F. G. Wyse, F. Anders, C. Chiappini, and K. Janßen, *The escape speed curve of the Galaxy obtained from Gaia DR2 implies a heavy Milky Way*, Astron. Astrophys. **616** (2018) L9, [arXiv:1807.04565].

[54] E. Del Nobile, *Appendiciario – A hands-on manual on the theory of direct Dark Matter detection*, 4, 2021.

[55] J. D. Lewin and P. F. Smith, *Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil*, Astropart. Phys. **6** (1996) 87–112.

[56] K. Freese, J. A. Frieman, and A. Gould, *Signal Modulation in Cold Dark Matter Detection*, Phys. Rev. D **37** (1988) 3388–3405.

[57] D. G. Cerdeno and A. M. Green, *Direct detection of WIMPs*, arXiv:1002.1912.

[58] D. N. Spergel, *The Motion of the Earth and the Detection of Wimps*, *Phys. Rev. D* **37** (1988) 1353.

[59] G. Heusser, *Cosmic ray interaction study with low-level Ge-spectrometry*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **369** (1996), no. 2 539–543.

[60] **DarkSide** Collaboration, P. Agnes et al., *Results From the First Use of Low Radioactivity Argon in a Dark Matter Search*, *Phys. Rev. D* **93** (2016), no. 8 081101, [`arXiv:1510.00702`]. [Addendum: Phys.Rev.D 95, 069901 (2017)].

[61] **XENON** Collaboration, E. Aprile et al., *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, *Phys. Rev. Lett.* **121** (2018), no. 11 111302, [`arXiv:1805.12562`].

[62] **LUX** Collaboration, D. Akerib et al., *Results of a Search for Sub-GeV Dark Matter Using 2013 LUX Data*, *Phys. Rev. Lett.* **122** (2019), no. 13 131301, [`arXiv:1811.11241`].

[63] **DEAP** Collaboration, R. Ajaj et al., *Search for dark matter with a 231-day exposure of liquid argon using DEAP-3600 at SNOLAB*, *Phys. Rev. D* **100** (2019), no. 2 022004, [`arXiv:1902.04048`].

[64] **CRESST** Collaboration, A. Abdelhameed et al., *First results from the CRESST-III low-mass dark matter program*, *Phys. Rev. D* **100** (2019), no. 10 102002, [`arXiv:1904.00498`].

[65] J. Amare et al., *Annual Modulation Results from Three Years Exposure of ANAIS-112*, *Phys. Rev. D* **103** (2021), no. 10 102005, [`arXiv:2103.01175`].

[66] **EDELWEISS** Collaboration, E. Armengaud et al., *Searching for low-mass dark matter particles with a massive Ge bolometer operated above-ground*, *Phys. Rev. D* **99** (2019), no. 8 082003, [`arXiv:1901.03588`].

[67] E. Daw et al., *The DRIFT Directional Dark Matter Experiments*, *EAS Publ. Ser.* **53** (2012) 11–18, [`arXiv:1110.0222`].

[68] **CUORE** Collaboration, A. Giachero et al., *New results from the CUORE experiment*, *PoS* **ICHEP2020** (2021) 133, [`arXiv:2011.09295`].

[69] E. Behnke et al., *Final Results of the PICASSO Dark Matter Search Experiment*, *Astropart. Phys.* **90** (2017) 85–92, [`arXiv:1611.01499`].

[70] **COUPP** Collaboration, E. Behnke et al., *Dark matter detection with bubble chambers*, *Nucl. Phys. B Proc. Suppl.* **221** (2011) 379.

[71] J. Billard et al., *Direct Detection of Dark Matter – APPEC Committee Report*, `arXiv:2104.07634`.

[72] G. D. Starkman, A. Gould, R. Esmailzadeh, and S. Dimopoulos, *Opening the Window on Strongly Interacting Dark Matter*, *Phys. Rev. D* **41** (1990) 3594.

[73] J. Collar and F. Avignone, *Diurnal modulation effects in cold dark matter experiments*, *Phys. Lett. B* **275** (1992) 181–185.

[74] J. Collar and I. Avignone, F.T., *The Effect of elastic scattering in the Earth on cold dark matter experiments*, *Phys. Rev. D* **47** (1993) 5238–5246.

[75] C. Kouvaris and I. M. Shoemaker, *Daily modulation as a smoking gun of dark matter with significant stopping rate*, *Phys. Rev. D* **90** (2014) 095011, [arXiv:1405.1729].

[76] B. J. Kavanagh, R. Catena, and C. Kouvaris, *Signatures of Earth-scattering in the direct detection of Dark Matter*, *JCAP* **01** (2017) 012, [arXiv:1611.05453].

[77] C. Kouvaris and J. Pradler, *Probing sub-GeV Dark Matter with conventional detectors*, *Phys. Rev. Lett.* **118** (2017), no. 3 031803, [arXiv:1607.01789].

[78] **CYGNO** Collaboration, E. Baracchini et al., *A 1 m^3 Gas Time Projection Chamber with Optical Readout for Directional Dark Matter Searches: the CYGNO Experiment*, *2019 IEEE Nuclear Science Symposium (NSS) and Medical Imaging Conference (MIC)* (2019) [arXiv:2001.02453].

[79] A. Migdal, *Ionization of atoms accompanying α- and β-decay*, *J. Phys. USSR* **4** (1941) 449.

[80] R. Landau and E. M. Lifshits, *Quantum mechanics: Non-Relativistic Theory Vol. 3.* Butterworth-Heinemann, Oxford, 1991.

[81] T. Ruijgrok, B. Nijboer, and M. Hoare, *Recoil-induced excitation of atoms by neutron scattering*, *Physica A: Statistical Mechanics and its Applications* **120** (1983), no. 3 537 – 544.

[82] L. Vegh, *Multiple ionisation effects due to recoil in atomic collisions*, *Journal of Physics B: Atomic and Molecular Physics* **16** (nov, 1983) 4175–4189.

[83] G. Baur, F. Rosel, and D. Trautmann, *Ionisation induced by neutrons*, *Journal of Physics B: Atomic and Molecular Physics* **16** (jul, 1983) L419–L423.

[84] P. Sharma, *Role of nuclear charge change and nuclear recoil on shaking processes and their possible implication on physical processes*, *Nucl. Phys. A* **968** (2017) 326–341.

[85] M. J. Dolan, F. Kahlhoefer, and C. McCabe, *Directly detecting sub-GeV dark matter with electrons from nuclear scattering*, *Phys. Rev. Lett.* **121** (2018), no. 10 101801, [arXiv:1711.09906].

[86] N. F. Bell, J. B. Dent, J. L. Newstead, S. Sabharwal, and T. J. Weiler, *Migdal effect and photon bremsstrahlung in effective field theories of dark matter direct detection and coherent elastic neutrino-nucleus scattering*, *Phys. Rev. D* **101** (2020), no. 1 015012, [arXiv:1905.00046].

[87] **CDEX** Collaboration, Z. Liu et al., *Constraints on Spin-Independent Nucleus Scattering with sub-GeV Weakly Interacting Massive Particle Dark Matter from the CDEX-1B Experiment at the China Jinping Underground Laboratory*, *Phys. Rev. Lett.* **123** (2019), no. 16 161301, [`arXiv:1905.00354`].

[88] **EDELWEISS** Collaboration, E. Armengaud et al., *Search for sub-GeV Dark Matter via Migdal effect with an EDELWEISS germanium detector with NbSi TES sensors*, `arXiv:2203.03993`.

[89] D. S. Akerib et al., *Enhancing the sensitivity of the LUX-ZEPLIN (LZ) dark matter experiment to low energy signals*, `arXiv:2101.08753`.

[90] **XENON** Collaboration, E. Aprile et al., *Search for Light Dark Matter Interactions Enhanced by the Migdal Effect or Bremsstrahlung in XENON1T*, *Phys. Rev. Lett.* **123** (2019), no. 24 241803, [`arXiv:1907.12771`].

[91] **SENSEI** Collaboration, L. Barak et al., *SENSEI: Direct-Detection Results on sub-GeV Dark Matter from a New Skipper-CCD*, *Phys. Rev. Lett.* **125** (2020), no. 17 171802, [`arXiv:2004.11378`].

[92] **SuperCDMS** Collaboration, M. Al-Bakry et al., *A Search for Low-mass Dark Matter via Bremsstrahlung Radiation and the Migdal Effect in SuperCDMS*, `arXiv:2203.02594`.

[93] C. P. Liu, C.-P. Wu, H.-C. Chi, and J.-W. Chen, *Model-independent determination of the Migdal effect via photoabsorption*, *Phys. Rev. D* **102** (2020), no. 12 121303, [`arXiv:2007.10965`].

[94] B. H. Bransden and C. J. Joachain, *Physics of Atoms and Molecules; 2nd ed.* Prentice-Hall, Harlow, 2003.

[95] M. F. Gu, *The flexible atomic code*, *Canadian Journal of Physics* **86** (2008), no. 5 675–689, [`https://doi.org/10.1139/p07-197`].

[96] S. Kubota, A. Nakamoto, T. Takahashi, S. Konno, T. Hamada, M. Miyajima, A. Hitachi, E. Shibamura, and T. Doke, *Evidence of the existence of exciton states in liquid argon and exciton-enhanced ionization from xenon doping*, *Phys. Rev. B* **13** (1976), no. 4 1649.

[97] D. Tucker-Smith and N. Weiner, *Inelastic dark matter*, *Phys. Rev. D* **64** (2001) 043502, [`hep-ph/0101138`].

[98] R. Trotta, R. Ruiz de Austri, and L. Roszkowski, *Prospects for direct dark matter detection in the Constrained MSSM*, *New Astron. Rev.* **51** (2007) 316–320, [`astro-ph/0609126`].

[99] L. Roszkowski, R. Ruiz de Austri, J. Silk, and R. Trotta, *On prospects for dark matter indirect detection in the Constrained MSSM*, *Phys. Lett. B* **671** (2009) 10–14, [`arXiv:0707.0622`].

[100] C. Strege, R. Trotta, G. Bertone, A. H. Peter, and P. Scott, *Fundamental statistical limitations of future dark matter direct detection experiments*, *Phys. Rev. D* **86** (2012) 023507, [arXiv:1201.3631].

[101] C. Arina, *Bayesian analysis of multiple direct detection experiments*, *Phys. Dark Univ.* **5-6** (2014) 1–17, [arXiv:1310.5718].

[102] **GAMBIT Dark Matter Workgroup** Collaboration, T. Bringmann et al., *DarkBit: A GAMBIT module for computing dark matter observables and likelihoods*, *Eur. Phys. J. C* **77** (2017), no. 12 831, [arXiv:1705.07920].

[103] S. Liem, G. Bertone, F. Calore, R. Ruiz de Austri, T. M. Tait, R. Trotta, and C. Weniger, *Effective field theory of dark matter: a global analysis*, *JHEP* **09** (2016) 077, [arXiv:1603.05994].

[104] A. Krishak, A. Dantuluri, and S. Desai, *Robust model comparison tests of DAMA/LIBRA annual modulation*, *JCAP* **02** (2020) 007, [arXiv:1906.05726].

[105] A. Krishak and S. Desai, *An independent assessment of significance of annual modulation in COSINE-100 data*, *Open J. Astrophys.* **2** (2019), no. 1 [arXiv:1907.07199].

[106] **DarkSide** Collaboration, P. Agnes et al., *Constraints on Sub-GeV Dark-Matter–Electron Scattering from the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018), no. 11 111303, [arXiv:1802.06998].

[107] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari, *Complementarity of dark matter detectors in light of the neutrino background*, *Phys. Rev. D* **90** (2014), no. 8 083510, [arXiv:1408.3581].

[108] P. Astone and G. D'Agostini, *Inferring the intensity of Poisson processes at the limit of the detector sensitivity (with a case study on gravitational wave burst search)*, hep-ex/9909047.

[109] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [arXiv:1007.1727]. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[110] S. Algeri, J. Aalbers, K. Dundas Morå, and J. Conrad, *Searching for new phenomena with profile likelihood ratio tests*, *Nature Rev. Phys.* **2** (2020), no. 5 245–252.

[111] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[112] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[113] M. Plummer, "Jags: A program for analysis of bayesian graphical models using gibbs sampling." http://mcmc-jags.sourceforge.net/, 2003.

[114] R Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013.

[115] M. Plummer, "R package version 4-10." https://cran.r-project.org/web/packages/rjags/index.html.

[116] F. Beaujean, A. Caldwell, D. Greenwald, D. Kollar, K. Kröninger, and O. Schulz, "Bayesian Analysis Toolkit: BAT." https://github.com/bat/bat.

[117] H. Akaike, *A new look at the statistical model identification*, *IEEE Transactions on Automatic Control* **19** (1974), no. 6 716–723.

[118] G. Schwarz, *Estimating the Dimension of a Model*, *The Annals of Statistics* **6** (1978), no. 2 461 – 464.

[119] A. R. Liddle, *Information criteria for astrophysical model selection*, *Mon. Not. Roy. Astron. Soc.* **377** (2007) L74–L78, [astro-ph/0701113].

[120] H. Jeffreys, *The Theory of Probability.* Oxford Classic Texts in the Physical Sciences. Oxford University Press, 1939.

[121] **DAMA** Collaboration, R. Bernabei et al., *Search for WIMP annual modulation signature: Results from DAMA / NaI-3 and DAMA / NaI-4 and the global combined analysis*, *Phys. Lett. B* **480** (2000) 23–31.

[122] R. Bernabei et al., *Dark matter search*, *Riv. Nuovo Cim.* **26N1** (2003) 1–73, [astro-ph/0307403].

[123] R. Bernabei et al., *Dark matter particles in the Galactic halo: Results and implications from DAMA/NaI*, *Int. J. Mod. Phys. D* **13** (2004) 2127–2160, [astro-ph/0501412].

[124] **DAMA** Collaboration, R. Bernabei et al., *First results from DAMA/LIBRA and the combined results with DAMA/NaI*, *Eur. Phys. J. C* **56** (2008) 333–355, [arXiv:0804.2741].

[125] R. Bernabei et al., *The DAMA project: Achievements, implications and perspectives*, *Prog. Part. Nucl. Phys.* **114** (2020) 103810.

[126] D. Buttazzo, P. Panci, N. Rossi, and A. Strumia, *Annual modulations from secular variations: relaxing DAMA?*, *JHEP* **04** (2020) 137, [arXiv:2002.00459].

[127] S. Piacentini, "`bayes_analysis`." Code available at https://github.com/piacent/bayes_analysis.

[128] Q. e. a. Gronau, "`bridgesampling`: Bridge Sampling for Marginal Likelihoods and Bayes Factors." https://github.com/quentingronau/bridgesampling.

[129] **DarkSide** Collaboration, P. Agnes et al., *The veto system of the DarkSide-50 experiment*, *JINST* **11** (2016), no. 03 P03016, [`arXiv:1512.07896`].

[130] **DarkSide** Collaboration, P. Agnes et al., *First Results from the DarkSide-50 Dark Matter Experiment at Laboratori Nazionali del Gran Sasso*, *Phys. Lett. B* **743** (2015) 456–466, [`arXiv:1410.0653`].

[131] J. Lindhard, *On the properties of a gas of charged particles*, *Kgl. Danske Videnskab, Selskab. Mat. Fys. Medd.* **28** (1954), no. 8.

[132] J. Lindhard, V. Nielsen, M. Scharff, and P. Thomsen, *Integral equations governing radiation effects (Notes on atomic collisions, iii)*, *Kgl. Danske Videnskab, Selskab. Mat. Fys. Medd.* **33** (Jan, 1963).

[133] E. Aprile, A. E. Bolotnikov, A. L. Bolozdynya, and T. Doke, *Noble Gas Detectors*. Wiley, 2008.

[134] J. Thomas and D. A. Imel, *Recombination of electron-ion pairs in liquid argon and liquid xenon*, *Phys. Rev. A* **36** (1987) 614–616.

[135] G. Jaffe, *The columnar theory of ionization*, *Ann. Phys.* **42** (1913) 303.

[136] J. Birks, *Theory and Practice of Scintillation Counting*. Pergamon Press, New York, 1964.

[137] **ICARUS** Collaboration, S. Amoruso et al., *Study of electron recombination in liquid argon with the ICARUS TPC*, *Nucl. Instrum. Meth. A* **523** (2004) 275–286.

[138] F. Bezrukov, F. Kahlhoefer, M. Lindner, F. Kahlhoefer, and M. Lindner, *Interplay between scintillation and ionization in liquid xenon Dark Matter searches*, *Astropart. Phys.* **35** (2011) 119–127, [`arXiv:1011.3990`].

[139] **DarkSide** Collaboration, P. Agnes et al., *DarkSide-50 532-day Dark Matter Search with Low-Radioactivity Argon*, *Phys. Rev. D* **98** (2018), no. 10 102006, [`arXiv:1802.07198`].

[140] C. Salas Redondo, P. Kleine, K. Roszeitis, T. Achenbach, M. Kroll, M. Thomschke, and S. Reineke, *Interplay of Fluorescence and Phosphorescence in Organic Biluminescent Emitters*, *The Journal of Physical Chemistry C* **121** (Jun, 2017) 14946–14953.

[141] **DarkSide** Collaboration, P. Agnes et al., *Simulation of argon response and light detection in the DarkSide-50 dual phase TPC*, *JINST* **12** (2017), no. 10 P10015, [`arXiv:1707.05630`].

[142] **SCENE** Collaboration, T. Alexander et al., *Observation of the Dependence on Drift Field of Scintillation from Nuclear Recoils in Liquid Argon*, *Phys. Rev. D* **88** (2013), no. 9 092006, [`arXiv:1306.5675`].

[143] **SCENE** Collaboration, H. Cao et al., *Measurement of Scintillation and Ionization Yield and Scintillation Pulse Shape from Nuclear Recoils in Liquid Argon*, Phys. Rev. D **91** (2015) 092007, [`arXiv:1406.4825`].

[144] **DarkSide** Collaboration, P. Agnes et al., *Calibration of the liquid argon ionization response to low energy electronic and nuclear recoils with DarkSide-50*, Phys. Rev. D **104** (2021), no. 8 082005, [`arXiv:2107.08087`].

[145] X. Mougeot, *BetaShape: A new code for improved analytical calculations of beta spectra*, EPJ Web Conf. **146** (2017) 12015.

[146] D. Cullen, *Program RELAX: A code designed to calculate atomic relaxation spectra of x-rays and electrons*, Tech. Rep. UCRL-ID.110438 (03, 1992).

[147] D. Cullen, *A Survey of atomic Binding Energies for Use in EPICS2017*, Tech. Rep. IAEA-NDS-0224 (2017).

[148] U. Fano, *Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions*, Phys. Rev. **72** (Jul, 1947) 26–29.

[149] A. G. Santos-Ocampo and D. C. Conway, $\frac{L}{K}$*-Capture Ratio and* $\frac{E_L}{E_K}$ *for* Ar^{37}, Phys. Rev. **120** (Dec, 1960) 2196–2200.

[150] T. Doke, A. Hitachi, S. Kubota, A. Nakamoto, and T. Takahashi, *Estimation of Fano factors in liquid argon, krypton, xenon and xenon-doped liquid argon*, Nucl. Instrum. Meth. **134** (1976) 353–357.

[151] T. Doke, H. J. Crawford, A. Hitachi, J. Kikuchi, P. J. Lindstrom, K. Masuda, E. Shibamura, and T. Takahashi, *Let Dependence of Scintillation Yields in Liquid Argon*, Nucl. Instrum. Meth. A **269** (1988) 291–296.

[152] P. Agnes et al., *Measurement of the liquid argon energy response to nuclear and electronic recoils*, Phys. Rev. D **97** (2018), no. 11 112005, [`arXiv:1801.06653`].

[153] J. P. Biersack and J. F. Ziegler, *The Stopping and Range of Ions in Solids*, in *Ion Implantation Techniques* (H. Ryssel and H. Glawischnig, eds.), (Berlin, Heidelberg), pp. 122–156, Springer Berlin Heidelberg, 1982.

[154] T. H. Joshi et al., *First measurement of the ionization yield of nuclear recoils in liquid argon*, Phys. Rev. Lett. **112** (2014) 171303, [`arXiv:1402.2037`].

[155] G. Moliere, *Theorie der Streuung schneller geladener Teilchen I. Einzelstreuung am abgeschirmten Coulomb-Feld*, Z. Naturforsch. A **2** (1947) 133.

[156] W. Lenz, *Über die Anwendbarkeit der statistischen Methode auf Ionengitter*, Zeitschrift für Physik **77** (1932) 713.

[157] H. Jensen, *Die Ladungsverteilung in Ionen und die Gitterkonstante des Rubidiumbromids nach der statistischen Methode*, Zeitschrift für Physik **77** (1932) 722.

[158] L. C. Angrave, N. E. Booth, R. J. Gaitskell, G. L. Salmon, and M. R. Harston, *Measurement of the Atomic Exchange Effect in Nuclear β Decay*, *Phys. Rev. Lett.* **80** (1998) 1610–1613.

[159] NVIDIA, P. Vingelmann, and F. H. Fitzek, *CUDA, release: 10.2.89*, 2020.

[160] G. Grilli di Cortona, "`DDrates`." Code available at https://github.com/ggrillidc/DDrates.

[161] **UTfit** Collaboration, A. J. Bevan et al., *The UTfit collaboration average of D meson mixing data: Winter 2014*, *JHEP* **03** (2014) 123, [arXiv:1402.1664].

[162] J. de Blas, M. Ciuchini, E. Franco, A. Goncalves, S. Mishima, M. Pierini, L. Reina, and L. Silvestrini, *Global analysis of electroweak data in the Standard Model*, arXiv:2112.07274.

[163] T. J. Loredo, *Accounting for source uncertainties in analyses of astronomical survey data*, *AIP Conf. Proc.* **735** (2004), no. 1 195–206, [astro-ph/0409387].

[164] **XENON** Collaboration, E. Aprile et al., *Light Dark Matter Search with Ionization Signals in XENON1T*, *Phys. Rev. Lett.* **123** (2019), no. 25 251801, [arXiv:1907.11485].

[165] J. Ziegler, "SRIM software." http://www.srim.org/.

[166] M. Foxe, C. Hagmann, I. Jovanovic, A. Bernstein, T. H. Joshi, K. Kazkaz, V. Mozin, S. V. Pereverzev, S. Sangiorgio, and P. Sorensen, *Modeling ionization and recombination from low energy nuclear recoils in liquid argon*, *Astropart. Phys.* **69** (2015) 24–29.

[167] T. Bringmann and M. Pospelov, *Novel direct detection constraints on light dark matter*, *Phys. Rev. Lett.* **122** (2019), no. 17 171801, [arXiv:1810.10543].

[168] Y. Ema, F. Sala, and R. Sato, *Light Dark Matter at Neutrino Experiments*, *Phys. Rev. Lett.* **122** (2019), no. 18 181802, [arXiv:1811.00520].

[169] J.-W. Wang, A. Granelli, and P. Ullio, *Direct Detection Constraints on Blazar-Boosted Dark Matter*, arXiv:2111.13644.

[170] **PandaX-II** Collaboration, X. Cui et al., *A Search for the Cosmic Ray Boosted Sub-GeV Dark Matter at the PandaX-II Experiment*, arXiv:2112.08957.

[171] **CDEX** Collaboration, R. Xu et al., *Constraints on sub-GeV Dark Matter Boosted by Cosmic Rays from CDEX-10 Experiment at the China Jinping Underground Laboratory*, arXiv:2201.01704.

[172] G. Grilli di Cortona, A. Messina, and S. Piacentini, *Search for solar axions with the DarkSide-50 experiment*, *DarkSide DocDB* (9, 2020) 4049.

[173] F. Mayet et al., *A review of the discovery reach of directional Dark Matter detection*, *Phys. Rept.* **627** (2016) 1–49, [arXiv:1602.03781].

[174] K. M. Zurek, *Asymmetric Dark Matter: Theories, Signatures, and Constraints*, *Phys. Rept.* **537** (2014) 91–121, [arXiv:1308.0338].

[175] K. Petraki and R. R. Volkas, *Review of asymmetric dark matter*, *Int. J. Mod. Phys. A* **28** (2013) 1330028, [arXiv:1305.4939].

[176] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, *Model for Thermal Relic Dark Matter of Strongly Interacting Massive Particles*, *Phys. Rev. Lett.* **115** (2015), no. 2 021301, [arXiv:1411.3727].

[177] R. Bernabei et al., *On electromagnetic contributions in WIMP quests*, *Int. J. Mod. Phys. A* **22** (2007) 3155–3168, [arXiv:0706.1421].

[178] G. Cavoto, F. Luchetta, and A. Polosa, *Sub-GeV Dark Matter Detection with Electron Recoils in Carbon Nanotubes*, *Phys. Lett. B* **776** (2018) 338–344, [arXiv:1706.02487].

[179] R. Essig, J. Mardon, and T. Volansky, *Direct Detection of Sub-GeV Dark Matter*, *Phys. Rev. D* **85** (2012) 076007, [arXiv:1108.5383].

[180] R. Essig, A. Manalaysay, J. Mardon, P. Sorensen, and T. Volansky, *First Direct Detection Limits on sub-GeV Dark Matter from XENON10*, *Phys. Rev. Lett.* **109** (2012) 021301, [arXiv:1206.2644].

[181] J. Seguinot, T. Ypsilantis, and A. Zichichi, *A High rate solar neutrino detector with energy determination*, *Conf. Proc. C* **920310** (1992) 289–313.

[182] C. Arpesella, C. Broggini, and C. Cattadori, *A possible gas for solar neutrino spectroscopy*, *Astropart. Phys.* **4** (1996) 333–341.

[183] E. Baracchini, W. Derocco, and G. Dho, *Discovering supernova-produced dark matter with directional detectors*, *Phys. Rev. D* **102** (2020), no. 7 075036, [arXiv:2009.08836].

[184] S. E. Vahsen et al., *CYGNUS: Feasibility of a nuclear recoil observatory with directional sensitivity to dark matter and neutrinos*, arXiv:2008.12587.

[185] J. N. Marx and D. R. Nygren, *The Time Projection Chamber*, *Phys. Today* **31N10** (1978) 46–53.

[186] D. R. Nygren, *The Time Projection Chamber: A New 4 pi Detector for Charged Particles*, *eConf* **C740805** (1974) 58.

[187] W. B. Atwood et al., *Performance of the ALEPH time projection chamber*, *Nucl. Instrum. Meth. A* **306** (1991) 446–458.

[188] J. Alme et al., *The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events*, *Nucl. Instrum. Meth. A* **622** (2010) 316–367, [arXiv:1001.1950].

[189] **ALICE** Collaboration, R. Münzer, *Upgrade of the ALICE Time Projection Chamber*, *Nucl. Instrum. Meth. A* **958** (2020) 162058.

[190] **DUNE** Collaboration, E. Chardonnet, *The DUNE dual-phase liquid argon TPC*, *JINST* **15** (2020), no. 05 C05064.

[191] M. M. F. R. Fraga, F. A. F. Fraga, S. T. G. Fetal, L. M. S. Margato, R. Ferreira-Marques, and A. J. P. L. Policarpo, *The GEM scintillation in He CF_4, Ar CF_4, Ar TEA and Xe TEA mixtures*, *Nucl. Instrum. Meth.* **A504** (2003) 88–92.

[192] L. M. S. Margato, A. Morozov, M. M. F. R. Fraga, L. Pereira, and F. A. F. Fraga, *Effective decay time of CF_4 secondary scintillation*, *JINST* **8** (2013) P07008.

[193] E. Baracchini et al., *First evidence of luminescence in a He/CF_4 gas mixture induced by non-ionizing electrons*, *JINST* **15** (2020), no. 08 P08018, [`arXiv:2004.10493`].

[194] C. M. B. Monteiro, L. M. P. Fernandes, J. F. C. A. Veloso, C. A. B. Oliveira, and J. M. F. dos Santos, *Secondary scintillation yield from GEM and THGEM gaseous electron multipliers for direct dark matter search*, *Phys. Lett. B* **714** (2012) 18–23.

[195] F. Sauli, *GEM: A new concept for electron amplification in gas detectors*, *Nucl. Instrum. Meth. A* **386** (1997) 531–534.

[196] L. M. S. Margato, F. A. F. Fraga, S. T. G. Fetal, M. M. F. R. Fraga, E. F. S. Balau, A. Blanco, R. Ferreira-Marques, and A. J. P. L. Policarpo, *Performance of an optical readout GEM-based TPC*, *Nucl. Instrum. Meth.* **A535** (2004) 231–235.

[197] M. Marafini, V. Patera, D. Pinci, A. Sarti, A. Sciubba, and E. Spiriti, *High granularity tracker based on a Triple-GEM optically read by a CMOS-based camera*, *JINST* **10** (2015), no. 12 P12010, [`arXiv:1508.07143`].

[198] N. S. Phan, E. R. Lee, and D. Loomba, *Imaging ^{55}Fe Electron Tracks in a GEM-based TPC Using a CCD Readout*, *arXiv 1703.09883* (2017) [`arXiv:1703.09883`].

[199] M. Marafini, V. Patera, D. Pinci, A. Sarti, A. Sciubba, and E. Spiriti, *ORANGE: A high sensitivity particle tracker based on optically read out GEM*, *Nucl. Instrum. Meth. A* **845** (2017) 285–288.

[200] V. C. Antochi, E. Baracchini, G. Cavoto, E. D. Marco, M. Marafini, G. Mazzitelli, D. Pinci, F. Renga, S. Tomassini, and C. Voena, *Combined readout of a triple-GEM detector*, *JINST* **13** (2018), no. 05 P05001, [`arXiv:1803.06860`].

[201] M. Marafini, V. Patera, D. Pinci, A. Sarti, A. Sciubba, and N. M. Torchia, *Study of the Performance of an Optically Readout Triple-GEM*, *IEEE Transactions on Nuclear Science* **65** (Jan., 2018) 604–608.

[202] D. Pinci, E. Baracchini, G. Cavoto, E. Di Marco, M. Marafini, G. Mazzitelli, F. Renga, S. Tomassini, and C. Voena, *High resolution TPC based on optically readout GEM*, *Nucl. Instrum. Meth. A* **936** (2019) 453–455.

[203] V. C. Antochi et al., *Performance of an optically read out time projection chamber with ultra-relativistic electrons*, *Nucl. Instrum. Meth. A* **999** (2021) 165209, [`arXiv:2005.12272`].

[204] E. Baracchini et al., *Stability and detection performance of a GEM-based Optical Readout TPC with He/CF$_4$ gas mixtures*, *JINST* **15** (2020), no. 10 P10001, [`arXiv:2007.00608`].

[205] R. Campagnola, *Study and optimization of the light-yield of a triple-GEM detector*. PhD thesis, Sapienza University of Rome, 2018.

[206] R. Veenhof, *Garfield, a drift chamber simulation program*, *Conf. Proc. C* **9306149** (1993) 66–71.

[207] R. Veenhof, *GARFIELD, recent developments*, *Nucl. Instrum. Meth. A* **419** (1998) 726–730.

[208] **CYGNO** Collaboration, Amaro, F. D. and others, *The CYGNO Experiment*, *Instruments* **6** (2022), no. 1.

[209] **GEANT4** Collaboration, S. Agostinelli et al., *GEANT4–a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[210] B. Buonomo, C. Di Giulio, L. G. Foggetta, and P. Valente, *A Hardware and Software Overview on the New BTF Transverse Profile Monitor*, in *5th International Beam Instrumentation Conference*, p. WEPG73, 2017.

[211] P. Valente, B. Buonomo, C. Di Giulio, and L. G. Foggetta, *Frascati Beam-Test Facility (BTF) High Resolution Beam Spot Diagnostics*, in *5th International Beam Instrumentation Conference*, p. MOPG65, 2017.

[212] W. Blum, L. Rolandi, and W. Riegler, *Particle detection with drift chambers*. Particle Acceleration and Detection, ISBN = 9783540766834. Springer, 2008.

[213] W. S. Anderson, J. C. Armitage, E. Dunn, J. G. Heinrich, C. Lu, K. T. McDonald, J. Weckel, and Y. Zhu, *Electron attachment, effective ionization coefficient, and electron drift velocity for CF-4 gas mixtures*, *Nucl. Instrum. Meth. A* **323** (1992) 273–279.

[214] E. Baracchini et al., *A density-based clustering algorithm for the CYGNO data analysis*, *JINST* **15** (2020), no. 12 T12003, [`arXiv:2007.01763`].

[215] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, *A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pp. 226–231, AAAI Press, 1996.

[216] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic Active Contours*, *International Journal of Computer Vision* **22** (1997) 61–79.

[217] P. Marquez-Neila and L. Baumela and L. Alvarez, *A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36** (2014), no. 1 2–17.

[218] E. Baracchini et al., *Identification of low energy nuclear recoils in a gas time projection chamber with optical readout*, *Measur. Sci. Tech.* **32** (2021), no. 2 025902, [arXiv:2007.12508].

[219] I. Abritta Costa et al., *Performance of Prototype of Optically Readout TPC with a ^{55}Fe source*, *J. Phys. Conf. Ser.* **1498** (2020) 012017.

[220] P. Gondolo, *Recoil momentum spectrum in directional dark matter detectors*, *Phys. Rev. D* **66** (2002) 103513, [hep-ph/0209110].

[221] D. Baxter et al., *Recommended conventions for reporting results from direct dark matter searches*, *Eur. Phys. J. C* **81** (2021), no. 10 907, [arXiv:2105.00599].

[222] K. Nakamura et al., *Low pressure gas study for a direction-sensitive dark matter search experiment with MPGD*, *JINST* **7** (2012) C02023.

[223] **CRESST** Collaboration, M. Mancuso et al., *Searches for Light Dark Matter with the CRESST-III Experiment*, *J. Low Temp. Phys.* **199** (2020), no. 1-2 547–555.

[224] **SuperCDMS** Collaboration, R. Agnese et al., *Low-mass dark matter search with CDMSlite*, *Phys. Rev. D* **97** (2018), no. 2 022002, [arXiv:1707.01632].

[225] **SuperCDMS** Collaboration, R. Agnese et al., *First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector*, *Phys. Rev. Lett.* **121** (2018), no. 5 051301, [arXiv:1804.10697]. [Erratum: Phys.Rev.Lett. 122, 069901 (2019)].

[226] M. Willers et al., *Direct dark matter search with the CRESST-III experiment - status and perspectives*, *J. Phys. Conf. Ser.* **888** (2017), no. 1 012209.

[227] C. Savage, G. Gelmini, P. Gondolo, and K. Freese, *Compatibility of DAMA/LIBRA dark matter detection with other searches*, *JCAP* **04** (2009) 010, [arXiv:0808.3607].

[228] C. Bœhm, D. G. Cerdeño, P. A. N. Machado, A. Olivares-Del Campo, E. Perdomo, and E. Reid, *How high is the neutrino floor?*, *JCAP* **01** (2019) 043, [arXiv:1809.06385].

[229] **PICO** Collaboration, C. Amole et al., *Dark Matter Search Results from the Complete Exposure of the PICO-60 C_3F_8 Bubble Chamber*, *Phys. Rev. D* **100** (2019), no. 2 022001, [arXiv:1902.04031].

[230] **DRIFT** Collaboration, J. B. R. Battat et al., *Low Threshold Results and Limits from the DRIFT Directional Dark Matter Detector*, *Astropart. Phys.* **91** (2017) 65–74, [`arXiv:1701.00171`].

[231] R. Yakabe et al., *First limits from a 3D-vector directional dark matter search with the NEWAGE-0.3b' detector*, *PTEP* **2020** (2020), no. 11 113F01, [`arXiv:2005.05157`].

[232] **DarkSide** Collaboration, Agnes, P. and others, *DarkSide-20k Technical Design Report*, 2022.

[233] **DUNE** Collaboration, B. Abi et al., *The DUNE Far Detector Interim Design Report Volume 1: Physics, Technology and Strategies*, `arXiv:1807.10334`.

[234] **LUX-ZEPLIN** Collaboration, D. S. Akerib et al., *Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment*, *Phys. Rev. D* **101** (2020), no. 5 052002, [`arXiv:1802.06039`].

[235] **XENON** Collaboration, E. Aprile et al., *Projected WIMP sensitivity of the XENONnT dark matter experiment*, *JCAP* **11** (2020) 031, [`arXiv:2007.08796`].

[236] J. Billard, L. Strigari, and E. Figueroa-Feliciano, *Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments*, *Phys. Rev. D* **89** (2014), no. 2 023524, [`arXiv:1307.5458`].

[237] E. Bagnaschi et al., *Likelihood Analysis of the pMSSM11 in Light of LHC 13-TeV Data*, *Eur. Phys. J. C* **78** (2018), no. 3 256, [`arXiv:1710.11091`].

[238] M. collaboration, "Migdal In Galactic Dark mAtter expLoration: MIGDAL." `https://migdal.pp.rl.ac.uk`.

[239] K. D. Nakamura, K. Miuchi, S. Kazama, Y. Shoji, M. Ibe, and W. Nakano, *Detection capability of the Migdal effect for argon and xenon nuclei with position-sensitive gaseous detectors*, *PTEP* **2021** (2021), no. 1 013C01, [`arXiv:2009.05939`].

[240] A. Pietropaolo et al., *The Frascati Neutron Generator: A multipurpose facility for physics and engineering*, *J. Phys.: Conf. Ser.* **1021** (2018) 012004.

[241] **n_TOF** Collaboration, N. Colonna et al., *The fission experimental programme at the CERN n_TOF facility: status and perspectives*, *Eur. Phys. J. A* **56** (2020), no. 2 48.