FERMILAB-PUB-23-612-T (accepted) DOI:10.1088/1475-7516/2024/03/047

FERMILAB-PUB-23-612-T

Prepared for submission to JCAP

Extracting Halo Independent Information from Dark Matter Electron Scattering Data

Elias Bernreuther,[a] **Patrick J. Fox,**[a] **Benjamin Lillard,**[b,c] **Anna-Maria Taki,**[c] **and Tien-Tien Yu**[b,c]

[a]Particle Theory Department, Theory Division, Fermilab, Batavia, IL 60510, USA
[b]Department of Physics, University of Oregon, Eugene, OR 97403, USA
[c]Institute for Fundamental Science, University of Oregon, Eugene, OR 97403, USA

E-mail: ebernreu@fnal.gov, pjfox@fnal.gov, blillard@uoregon.edu, annamariataki@gmail.com, tientien@uoregon.edu

Abstract. Direct detection experiments and the interpretation of their results are sensitive to the velocity structure of the dark matter in our galactic halo. In this work, we extend the formalism that deals with such astrophysics-driven uncertainties, originally introduced in the context of dark-matter-nuclear scattering, to include dark-matter-electron scattering interactions. Using mock data, we demonstrate that the ability to determine the correct dark matter mass and velocity distribution is depleted for recoil spectra which only populate a few low-lying bins, such as models involving a light mediator. We also demonstrate how this formalism allows one to test the compatibility of existing experimental data sets (*e.g.* SENSEI and EDELWEISS), as well as make predictions for possible future experiments (*e.g.* GaAs-based detectors).

arXiv:2311.04957v2 [hep-ph] 5 Apr 2024

Contents

1 Introduction **1**

2 Formalism **2**
- 2.1 Dark-Matter-Electron Scattering Rate 3
- 2.2 Halo-Independent Method 4
- 2.3 Statistics 5
- 2.4 Linear Algebra and Flat Directions 6

3 Application to simulated rates **8**
- 3.1 Best-fit dark matter particle and astrophysical properties 8
- 3.2 Limits on the dark matter mass 11

4 Application to data **14**

5 Discussion and Conclusions **18**

A Halo Models **21**

B The Pseudoinverse **22**

1 Introduction

Despite considerable progress on several fronts, and a concerted effort experimentally and theoretically, the properties of particle dark matter (DM) are still unknown. One of the key tools in the hunt for DM is direct detection experiments. These experiments search for the signs of DM in our galactic halo scattering off Standard Model (SM) particles in the (typically underground) lab. Motivated by the WIMP paradigm, these experiments originally focused on DM scattering off nuclei and were best suited for DM masses above ~ 1 GeV. More recently there has been a surge of interest and progress in DM-electron scattering experiments, to search for sub-GeV dark matter [1].

The expected scattering rate at direct detection experiments depends upon details of the target material, the nature of the DM-SM couplings, and the local DM distribution. The first of these, while challenging to compute, can be determined from experiment and theory. The second is presently unknown but at the low energies involved in the scattering there are only a finite number of possibilities to explore. The last unknown consists of the DM's distribution in both position and velocity space. Typical distributions can be determined from numerical simulations [2–5] and in our own galaxy the average DM density at the Sun's galactic radius is constrained by observations of stellar kinematics [6]. However, neither approach has the resolution to determine the DM distribution in our local neighborhood. When interpreting results from direct detection experiments, it is typically assumed that the speed of the DM impinging on the detector follows a Maxwell-Boltzmann distribution.

For DM scattering off nuclei, techniques have been developed to allow interpretation of direct detection results without needing to make this assumption [7–14] and thus circumventing the large uncertainty on the DM distribution. This is done by noticing that for

nuclear scattering the rate factorizes into a product of a target-dependent term and a target-independent, but halo-dependent, term. This then allows the extraction of (integrals of) the halo velocity distribution from experimental data and the direct comparison of two direct detection rates measured, or bounded, in different materials.

As with nuclear recoil experiments, astrophysical uncertainties have a significant effect on the interpretation of DM-electron scattering data [15, 16]. For DM-electron scattering the target electron's momentum is unknown and must be integrated over. For a given observed recoil energy the range of possible electron momenta depends upon the incoming DM velocity. This means that the scattering rate no longer cleanly factorizes. Instead, the rate involves a convolution of the electron wavefunction with the halo velocity distribution. However, it is still possible to extend halo-independent techniques to the case of electron scattering [17, 18].

Heuristically, the dependence of the scattering rate for nuclear recoils on astrophysics, *i.e.* the DM velocity distribution in the galactic frame f_χ, is

$$\frac{dR_n}{dE} \propto \int_{v_{\rm min}^{\rm nuc}(E)}^{v_{\rm max}} d^3v \frac{f_\chi(\vec{v}+\vec{v}_E)}{v} \ , \tag{1.1}$$

with $v_{\rm max}$ the maximal DM speed in the Earth's frame and we have suppressed the terms independent of astrophysics. There is a one-to-one correspondence between the nuclear recoil energy E_R and the minimum necessary incoming DM speed $v_{\rm min}^{\rm nuc} = \sqrt{\frac{m_N E_R}{2\mu_{\chi N}^2}}$, with m_N denoting the mass of the nucleus and $\mu_{\chi N}$ the reduced mass of the DM-nucleus system. Thus, the results of different experiments can be compared by analyzing their results in $v_{\rm min}$-space.

This situation is to be contrasted to electron recoils where the velocity distribution enters in a convolution,

$$\frac{dR_e}{dE} \propto \int dq \int_{v_{\rm min}(E,q)}^{v_{\rm max}} d^3v \frac{f_\chi(\vec{v}+\vec{v}_E)}{v} \ , \tag{1.2}$$

with $v_{\rm min}$ given below, in (2.4).

In section 2 we describe the formalism and our statistical approach to finding the best-fit DM mass and velocity distributions. In this work, we focus on semiconductor targets, but our approach can be readily applied to other detector materials. In section 3 we apply the method to mock data generated for several different DM models. Our ability to uncover the DM parameters used to generate the mock data improves with the number of populated electron-hole bins. In section 4 we use our approach to compare the consistency of results from SENSEI and EDELWEISS, concluding that their excesses cannot be simultaneously explained as coming entirely from dark matter events. In addition, we predict the rate in a future GaAs detector under the assumption that the SENSEI observations are due to dark matter. We conclude in section 5.

2 Formalism

In this section we present the method for calculating the scattering rate of DM off of electrons for an arbitrary velocity distribution. A general velocity distribution can be thought of as a sum of DM streams. Following the approach previously applied to nuclear recoils [7–14], we show how an observed signal can be used to extract the parameters associated with these streams. We discuss the statistical techniques we use to do this, along with their associated uncertainties.

2.1 Dark-Matter-Electron Scattering Rate

To begin, we review the calculations behind the DM-electron scattering rate. The general form for the differential scattering rate is given by [19]

$$\frac{dR}{d\ln E_e} = \frac{\rho_\chi}{m_\chi}\frac{\overline{\sigma}_e}{8\mu_{\chi e}^2}\int dq\, q\, |F_{\rm DM}(q)|^2 |f_{\rm res}(E_e,q)|^2 \eta\left(v_{\rm min}\right) \tag{2.1}$$

where E_e and q are the final electron's energy and the momentum transfer, respectively. m_χ is the DM mass, while $\rho_\chi = 0.4$ GeV/cm^3 is the local DM density. $\mu_{\chi e}$ is the reduced mass of the DM-e system. $\overline{\sigma}_e$ denotes the DM-free electron scattering cross section at fixed momentum transfer $q_0 = \alpha m_e$, where m_e is the electron mass. $F_{\rm DM}(q) = (\alpha m_e/q)^n$ encodes the momentum dependence of the interaction, where $n = 0\,(2)$ corresponds to a heavy (light) mediator. We assume that the interaction cross section does not directly depend on the DM velocity. $f_{\rm res}(E_e, q)$ is the material-dependent, dimensionless response function for an electron excitation with momentum q and energy E_e. For a crystal target, we have,

$$|f_{\rm res}(E_e,q)|^2 \equiv \frac{8\alpha m_e^2 E_e}{q^3} \times |f_{\rm crystal}(E_e,q)|^2\,, \tag{2.2}$$

where $\alpha \simeq 1/137$ is the fine-structure constant and $f_{\rm crystal}(E_e, q)$ is the dimensionless crystal form factor[1] as defined in [19]. Lastly, the DM velocity distribution in the galactic frame, $f_\chi(\vec{v})$, enters through $\eta(v_{\rm min})$,

$$\eta(v_{\rm min}) = \int_{v_{\rm min}} d^3v \frac{f_\chi(\vec{v}+\vec{v}_E)}{v}\,, \tag{2.3}$$

where $\vec{v}_E$ is the Earth's velocity, also in the galactic frame, and we have ignored the small amount of annual time dependence induced by the Earth's orbit around the Sun. For DM-electron scattering the minimum speed for scattering is

$$v_{\rm min}(q) = \frac{q}{2m_\chi} + \frac{E_e}{q}\,. \tag{2.4}$$

Thus, in contrast to the DM-nuclear case where $v_{\rm min}^{\rm nucl} = \sqrt{\frac{m_N E_R}{2\mu_{\chi N}^2}}$, there is not a clear mapping between recoil energy and $v_{\rm min}$, as (2.4) also depends on q. Instead, a fixed E_e leads to a range of $v_{\rm min}$ as a function of q. Alternatively, for a fixed v and E_e, q must lie in the range $q_- \le q \le q_+$ with

$$q_\pm(v) = m_\chi v \pm \sqrt{m_\chi^2 v^2 - 2E_e m_\chi}\,. \tag{2.5}$$

Furthermore, at fixed recoil energy there is a smallest possible speed allowed for scattering, and an associated momentum transfer to the electron,

$$v_*^2 = \frac{2E_e}{m_\chi} \quad \text{and} \quad q_* = m_\chi v_*\,. \tag{2.6}$$

An example of the region in (q, v) space that must be integrated over, at fixed recoil energy, is shown in figure 1. Note that the response function $f_{\rm res}$ is a relatively peaked

[1] For an atomic target, $|f_{\rm res}(E_e,q)|^2 \equiv |f_{nl}^{\rm ion}(E_e,q)|^2$, where (n,l) are the quantum numbers for the initial-state bound electron [20, 21].

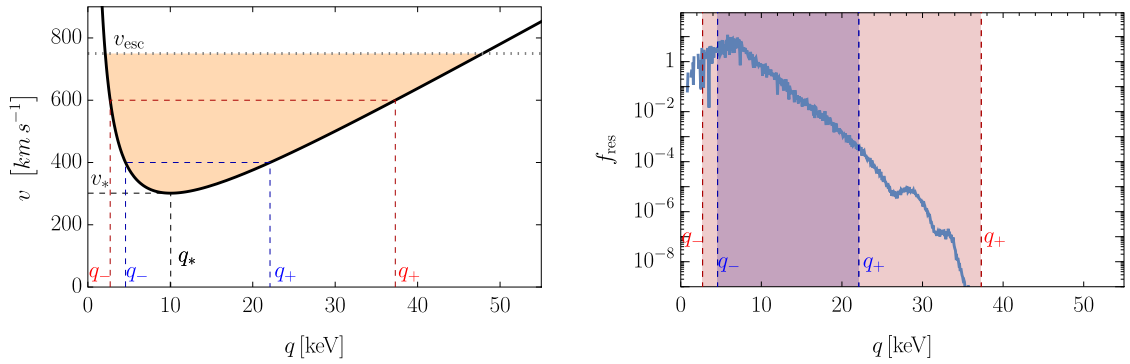

Figure 1: **Left:** the region in $q - v$ space that must be integrated over (shaded) for a 10 MeV DM candidate scattering off germanium, depositing 5.05 eV energy. The black line is $v_{\rm min}$ as a function of momentum exchange q. At fixed q the integral over v is from $v_{\rm min}$ to $v_{\rm esc}$, denoted by the dotted line. At fixed v the integral is over $q_- \leq q \leq q_+$, denoted by the dashed red/blue lines. The smallest SM speed probed and its corresponding momentum exchange are marked with v_*, q_*, respectively. **Right:** The dimensionless response function $f_{\rm res}$ and the ranges of q probed for the two choices of v in the left panel.

function. In the limit that $f_{\rm res}(q) \to \delta(q - q_0)$ the relationship between $v_{\rm min}$ and the recoil energy becomes one-to-one. In this case the behavior is much like for nuclear recoils where experiments that probe the same range of $v_{\rm min}$-space provide cross-checks of one another. The ability to directly compare the results of different experiments depends upon the details of the broader (physical) response functions. Furthermore, the relationship between recoil energy and $v_{\rm min}$ (2.4) becomes independent of DM mass when this mass is large and thus it is not possible to separate between different mass hypotheses in this regime.

Typically, one imposes that the velocity distribution of the DM halo, $f_\chi(v)$, follows a standard Maxwell-Boltzmann form, which depends on the circular velocity v_0 and galactic escape velocity $v_{\rm esc}$. However, as demonstrated in [15], the DM-electron scattering rate and, accordingly, the cross-section constraints and projections are highly sensitive to the choice of underlying halo model and halo parameters. Therefore, we propose a method in which one can interpret the results of DM-electron scattering experiments in a halo-independent manner.

2.2 Halo-Independent Method

We begin by parameterizing the rate as follows,

$$\frac{dR}{d\ln E_e} \equiv \frac{1}{8\mu_{\chi e}^2} \int dq\, q\, |F_{\rm DM}(q)|^2 |f_{\rm res}(E_e, q)|^2 \tilde{\eta}\,(v_{\rm min}) \tag{2.7}$$

where

$$\tilde{\eta}(v_{\rm min}) = \frac{\rho_\chi \overline{\sigma}_e}{m_\chi} \eta(v_{\rm min})\,. \tag{2.8}$$

Since $\eta(v_{\rm min})$ is given as the integral (2.3) over the velocity distribution, $\eta(v_{\rm min})$ and, therefore, $\tilde{\eta}(v_{\rm min})$ are monotonically-decreasing functions of $v_{\rm min}$. We adopt a conservative ansatz by approximating $\tilde{\eta}(v_{\rm min})$ as a series of step functions [11–13]. The intervals are defined by

dividing up $v_{\rm min}$-space into N_s intervals of the form $[v_i, v_{i+1}]$, with $v_i \leq v_{i+1}$. This corresponds to the scenario in which the velocity distribution of the DM halo is given by a series of streams with speeds $\{v_i\}$ and thus

$$\tilde{\eta}(v_{\rm min}) = \sum_i \tilde{\eta}_i \, \Theta(v_i - v_{\rm min}) \,. \tag{2.9}$$

Note that, by construction, $\tilde{\eta}$ is monotonically decreasing with $v_{\rm min}$. As $v_{\rm min}$ decreases, the step across speed v_i is of height $\tilde{\eta}_i$. We discuss the technical details of our numerical procedure below, see section 2.3.

For several sub-GeV DM experiments, the observed event rates are given as a function of the number of detected electrons, n_e. We use a simple model to determine the correspondence between the electron energy and n_e, given by

$$n_e = 1 + \lfloor (E_e - E_{\rm gap})/\varepsilon \rfloor \,, \tag{2.10}$$

where $\lfloor x \rfloor$ rounds x down to the nearest integer, $E_{\rm gap}$ is the bandgap energy of the material, and ε is the mean energy per electron-hole pair. For silicon, $\{E_{\rm gap}, \varepsilon\} = \{1.2, 3.8\}$ eV, while for germanium $\{E_{\rm gap}, \varepsilon\} = \{0.67, 3.0\}$ eV. The observed rate for n_e detected electrons is given by integrating the differential rate (2.7) over the energies associated with that number of electrons,

$$R(n_e) = \frac{1}{8\mu_{\chi e}^2} \int_{E_{\rm min}}^{E_{\rm max}} \frac{dE_e}{E_e} \int_{q_-}^{q_+} dq \, q \, |F_{\rm DM}(q)|^2 |f_{\rm res}(E_e, q)|^2 \sum_i \tilde{\eta}_i \Theta(v_i - v_{\rm min}) \,, \tag{2.11}$$

where $E_{\rm min} = E_{\rm gap} + (n_e - 1)\varepsilon$ and $E_{\rm max} = E_{\rm gap} + n_e \varepsilon$. The limits of q are determined by kinematics, and are given by $q_\pm$ in (2.5).

2.3 Statistics

In order to determine the constants $\tilde{\eta}_i$, we introduce a likelihood function which compares the observed number of events to the number predicted for a given choice of $\tilde{\eta}_i$. For our purposes, we will use a binned likelihood function for an experiment with N bins,

$$-2 \log \mathcal{L} = 2 \sum_{i=1}^{N} \left[R_i + B_i - N_i + N_i \log \frac{N_i}{R_i + B_i} \right] \,, \tag{2.12}$$

where R_i, B_i, N_i are the expected signal, expected background, and the observed number of events for the i-th bin of the experiment, respectively. The expected signal rate R_i in each bin is determined using (2.11), and any known sources of background can be accounted for with B_i. For the case where $N_i = 0$, we set $\left[N_i \log \frac{N_i}{R_i + B_i} \right] = 0$. For more than one experiment we label each likelihood with an index α, and the results of all experiments are combined by multiplying (adding) their (log-) likelihoods,

$$\mathcal{L} = \prod_\alpha \mathcal{L}^{(\alpha)} \,. \tag{2.13}$$

Given this likelihood, we can determine, for a fixed m_χ, the step heights $\tilde{\eta}_i$ which minimize $-2 \log \mathcal{L}$ (*i.e.* maximize the likelihood). From this, we can define the profile likelihood function,

$$\hat{\mathcal{L}}(m_\chi) = \max_{\tilde{\eta}} \mathcal{L}(m_\chi, \tilde{\eta}) \,. \tag{2.14}$$

The linearity of the rate (2.11) with respect to all $\tilde{\eta}_i$ ensures that any minimum of the likelihood in $\tilde{\eta}$ space is a global minimum. To solve this minimization problem in practice, we discretize the function $\tilde{\eta}(v_{\rm min})$ in 100 velocity steps of equal size $\Delta v_{\rm min}$ between $v_{\rm min} = 0$ km/s and $v_{\rm min}$=1000 km/s. In the galactic frame the maximal speed of dark matter bound in our halo is expected to be $\mathcal{O}(500 - 700)$km/s. In the Earth's frame the maximal speed is about 200 km/s higher. We consider 1000 km/s a conservative cut off, but it is simple to extend the range to higher speeds. Concretely, we take the step-function ansatz (2.9) and write the step heights as

$$\tilde{\eta}_i \equiv \exp(\tilde{g}_i) \times 1\,\text{year}^{-1}\,, \tag{2.15}$$

with real parameters $\tilde{g}_i$, where $i = 1 \ldots 100$. This parameterization ensures that the step heights are non-negative, as dictated by the monotonicity of $\tilde{\eta}(v_{\rm min})$.

To find the values $\tilde{g}_i$ that minimize the log-likelihood (2.12), we use the ADAM optimizer [22] implemented in TensorFlow 2.10.0 [23] with its default settings, a learning rate of 0.001, and the log-likelihood as the loss function. After removing an overall normalization of 10^{-20} year^{-1}, we initialize the $\tilde{g}_i$ to a random value between 0 and 1. We then perform 10^5 minimization steps, which was sufficient for the optimizer to converge in all cases studied in this work.

While we always use the log-likelihood as a loss function for our optimizer, we will also use the χ^2 as a test statistic, which is given by

$$\chi^2 = \sum_{i=1}^{N} \frac{(N_i - R_i - B_i)^2}{R_i + B_i}\,, \tag{2.16}$$

where all quantities are defined like in (2.12). Compared with the likelihood, the χ^2 has the advantage that it provides an absolute measure of the compatibility of the data with a given model hypothesis.

2.4 Linear Algebra and Flat Directions

The rate (2.11) depends linearly on $\tilde{\eta}_i$. So for fixed dark matter mass the expected rate for a given velocity distribution, corresponding to $\tilde{\eta}$, is given by

$$R_i(\tilde{\eta}) = \sum_j \mathcal{A}_{ij}\tilde{\eta}_j, \tag{2.17}$$

with

$$\mathcal{A}_{ij}(m_\chi) = \frac{1}{8\mu_{\chi e}^2} \int_{E_{\rm min}^{(i)}}^{E_{\rm max}^{(i)}} \frac{dE_e}{E_e} \int_{q_-}^{q_+} dq\, q\, |F_{\rm DM}(q)|^2 |f_{\rm res}(E_e, q)|^2 \Theta(v_j - v_{\rm min})\,. \tag{2.18}$$

The index i refers to the different observables, and $\mathcal{A}(m_\chi)$ encodes how much the jth velocity bin contributes to the ith observable, assuming some fixed value for m_χ. If $\mathcal{A}$ is an invertible square matrix, as would be the case if the number of observed energy bins was equal to the number of steps in the velocity distribution, then a best-fit solution for $\tilde{\eta}_j$ given some observed rates R_i could be found simply by inverting $\mathcal{A}$: $\tilde{\eta}_j = (\mathcal{A}^{-1})_{ji} R_i$.

When the number, n, of continuous parameters $\tilde{\eta}_i$ is larger than the number of observables, m, the matrix $\mathcal{A}$ is not invertible in the usual fashion. Instead one may introduce the

concept of a pseudoinverse[2] $\mathcal{A}^+$. The pseudoinverse can be found from the singular value decomposition (SVD) of $\mathcal{A}$. The SVD identifies orthonormal matrices U and V such that

$$\mathcal{A} = U\mathcal{D}V^T \ . \tag{2.19}$$

If $\mathcal{A}$ is of size $(m \times n)$ then $U(V)$ are square matrices of size $m(n)$, respectively, and $\mathcal{D} = D \oplus \mathbb{0}$ is of size $(m \times n)$, with D a diagonal $(n \times n)$ matrix and $\mathbb{0}$ a zero matrix. Defining $\mathcal{D}^+ = D^{-1} \oplus \mathbb{0}$ the pseudoinverse of $\mathcal{A}$ is defined as

$$\mathcal{A}^+ = V\mathcal{D}^+U^T \ . \tag{2.20}$$

Using the pseudoinverse a solution can be found:

$$\tilde{\eta}_* = \mathcal{A}^+ R \ . \tag{2.21}$$

It is shown, in appendix B, that this solution minimizes $|R - \mathcal{A}\tilde{\eta}|^2$.[3] For the case of interest here, with more parameters than constraints, this minimum is at zero. Thus, $\tilde{\eta}$ is the best-fit solution. However, there is no guarantee that the solution is physical *i.e.* that the resulting $\tilde{\eta}$ is monotonic. Furthermore, there are flat directions in the parameter space: perturbations to the vector $\tilde{\eta}$ that leave every observable R_i unchanged. These flat directions are spanned by the vectors that lie in the kernel of $\mathcal{A}$. Thus, there is a family of solutions related to $\tilde{\eta}_*$ by

$$\tilde{\eta}_* \to \tilde{\eta}_* + \left(\mathbb{1} - \mathcal{A}^+\mathcal{A}\right)\vec{\alpha} \ , \tag{2.22}$$

with $\vec{\alpha}$ an arbitrary vector that is projected down to $\ker(\mathcal{A})$. By moving around in these flat directions an equally good, but physical, solution for $\tilde{\eta}$ may be found.

Bands of 1σ or 2σ can be generated by a related two-step procedure. First, starting from a best-fit point $\tilde{\eta}_*$, the vector $\tilde{\eta}$ is varied in all of the directions perpendicular to $\ker(\mathcal{A})$, until the χ^2 of the fit reaches a 1σ (or 2σ) threshold. These results for $\tilde{\eta}$ generate a partial contour, surrounding the best-fit point. Next, every point on this contour is extended along the flat directions, using the same method as (2.22). This complete 1σ or 2σ contour encloses a volume, with $\tilde{\eta}_*$ and all of the points generated by (2.22) contained in the center.

With a large number of flat directions, determining in which direction to move to go from the unphysical solution $\tilde{\eta}_*$ to a physical one can be numerically challenging. However, once a physically allowed solution has been determined, for instance using the techniques outlined in the previous subsection, the pseudoinverse and its associated flat directions can be employed to determine the boundaries of the physically allowed region which yields the same rate as the original solution. In high-dimensional applications it may be more tractable to explore a subspace of the flat directions. In section 3 we employ a coarse-graining method, where k neighboring bins in the distribution are varied in parallel, $\delta\eta_i = \delta\eta_{i+1} = \ldots = \delta\eta_{i+k}$. This makes it easier to explore a larger range of the physically-allowed region, or at least those perturbations to $\tilde{\eta}_*$ that do not involve significant changes at the smallest scales. Through this technique, we can then define a Region of Equivalent Statistical Test (REST) that encapsulates the degeneracy in the underlying velocity distribution by containing all velocity distributions that yield the same rate.

[2] For a brief introduction on the properties of the pseudoinverse, see appendix B.

[3] By rescaling R and $\tilde{\eta}$ the quantity being minimized can be made into χ^2. In the limit where the number of events is large – where the Poisson distribution approaches a Gaussian – this quantity approaches $-2\log\mathcal{L}$.

3 Application to simulated rates

As an application of the formalism described in the previous sections, we simulate mock data under various DM model assumptions and apply the formalism to this data set. This provides an opportunity to measure the utility of this approach, under controlled conditions. We calculate mock data for both silicon and germanium targets. To compute the rate (2.1) we use crystal form factors (2.2) calculated by `QEdark` [19], but the procedure can be applied using rates from other codes, such as `DarkELF` [24], `EXCEED-DM` [25], and `QCDark` [26]. We simulated two representative halo models: the Standard Halo Model (SHM) and a stream of DM at a fixed velocity.

We take the SHM to have the form [27]

$$f_\chi(\vec{v}_\chi) = \frac{1}{K_{SHM}} e^{-v_\chi^2/v_0^2}\, \Theta(v_{\rm esc} - v_\chi)\,, \tag{3.1}$$

where $\vec{v}_\chi$ is the DM velocity in the galactic frame. We take the velocity dispersion to be $v_0 = 220$ km/s and $v_{\rm esc} = 544$ km/s for the escape velocity. The distribution is normalized such that $\int d^3v_\chi f_\chi(v_\chi) = 1$. The analytic expression for the normalization factor K_{SHM} is derived in appendix A. To determine the rate in a direct detection experiment we must boost to the Earth's frame. To do so, we use the average of the Earth's speed relative to the DM halo, $v_E = 232$ km/s, and ignore its time dependence.

In the galactic frame a stream with zero velocity dispersion is a delta function in velocity space, $f_\chi(\vec{v}_\chi) = \delta^{(3)}(\vec{v}_\chi - \vec{v}_{\rm str}^{\rm gal})$. Here, we model the DM velocity distribution as a Gaussian with small dispersion, σ. Furthermore, we define the stream's mean velocity in the Earth's frame $\vec{v}_{\rm str} = \vec{v}_{\rm str}^{\rm gal} - \vec{v}_E$. Thus, in the Earth's frame the velocity distribution for the stream is

$$f(\vec{v}) = \frac{1}{\sqrt{8\pi^3\sigma^3}} e^{-\frac{(\vec{v}-\vec{v}_{\rm str})^2}{2\sigma^2}}\,, \tag{3.2}$$

where we take the speed of the stream to be $v_{\rm str} = 300$ km/s and the dispersion $\sigma = 20$ km/s. We are interested in the limit where the stream's dispersion is small and its speed is far from the escape speed, thus we ignore the cutoff at the galactic escape speed. See appendix A for more discussion about the function η for the cases of a SHM and a stream.

The DM can couple to electrons either through a contact interaction or through the exchange of a light mediator. For these two cases we parameterize the DM form factor [28] as $F_{\rm DM} = 1$ and $F_{\rm DM} = (\alpha m_e/q)^2$, respectively.

Using these velocity distributions, the DM form factors, and the crystal form factors (2.2) calculated by `QEdark`, we simulate four different DM models, whose parameter choices are listed in table 1. In this analysis of mock data, for simplicity, we do not include backgrounds and hence always set $B_i = 0$.

3.1 Best-fit dark matter particle and astrophysical properties

Armed with the mock data, we can now employ our formalism to find the best-fit DM velocity distribution for the two elements, silicon and germanium, independently as well as the joint fit of the two rates. In figures 2 and 3, we show the value of χ^2 profiled over the DM velocity distribution as a function of the hypothesized DM mass for the mock data from models A–D. In each fit of figure 2, the DM form factor $F_{\rm DM}$ is set to its true value for the respective model, while in figure 3, the DM form factor is set to the wrong value.

model name	m_χ [MeV]	$F_{\rm DM}$	halo model
A	50	1	SHM
B	10	1	SHM
C	50	$(\alpha m_e/q)^2$	SHM
D	50	1	stream

Table 1: The DM mass, DM form factor, and DM halo model assumed for each of the generated mock data sets.

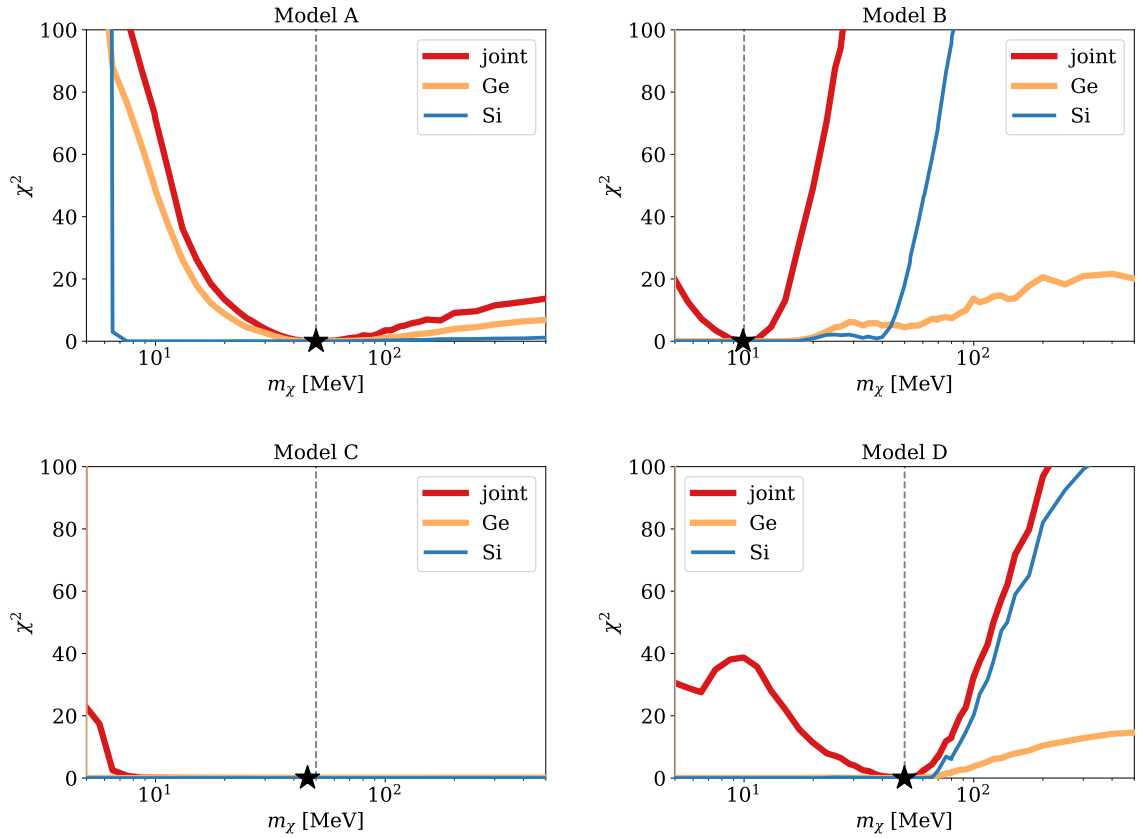

Figure 2: Best-fit χ^2 as a function of m_χ from a fit (profiled over the DM velocity distribution) of simulated Si and Ge rates at 1 kg·yr exposure for different DM and halo models. Shown are separate fits of the Si rate (blue) and the Ge rate (orange), and a joint fit of both (red). Each fit is carried out with the correct $F_{\rm DM}$ for the respective DM model. The vertical dashed line shows the true DM mass used in creating the mock data, while $\star$ denotes the DM mass that gives the best joint fit to the two mock data sets. For a version of these plots with a logarithmic y-axis, see the dashed curves in Figure 3.

Using the method described in section 2, we carry out separate fits of simulated silicon and germanium rates as well as a joint fit of the events from both elements. For the case of the joint fit with the correct $F_{\rm DM}$, we see that the minimum of the χ^2 statistic, *i.e.* the best fit, lies exactly at the true DM mass for models A, B and D, and nearly at the correct

mass for model C. In each case, the χ^2 value at the minimum is negligible. That is, for the correct mass we can fit the simulated data perfectly, as expected. At incorrect masses, on the other hand, a perfect joint fit of both silicon and germanium data cannot be achieved with any physical velocity distribution for models A, B and D. This demonstrates that our formalism is successful at identifying the correct DM mass for models with a heavy mediator, *i.e.* $F_{\rm DM} = 1$.

Only for model C, which has a light mediator, can we find some velocity distribution that fits the data perfectly for any DM mass $m_\chi \gtrsim 10$ MeV. Here, the light mediator, corresponding to $F_{\rm DM} \sim 1/q^2$, biases the electron recoil spectrum towards small energies largely irrespective of m_χ, erasing information about the DM mass from the resulting event rates. The value of χ^2 only increases at $m_\chi \lesssim 10$ MeV, where fitting the event rates of model C would require velocities above 1000 km/s, which we impose as a hard cutoff in our fits.

For a heavy DM mediator, our method correctly infers the underlying DM mass via a joint fit of silicon and germanium data. However, when fitting silicon and germanium rates individually, we find in almost every case (the Ge fit of model A being the only exception) a broad range of masses for which some velocity distribution yields negligible χ^2. This demonstrates that it is essential to combine event rates from multiple electron-scattering experiments with different materials to infer DM properties in a halo-independent manner.

Results for incorrect choices of the form factor (*i.e.* $F_{\rm DM} \sim 1/q^2$ for models A, B, D, and $F_{\rm DM} = 1$ for model C) are displayed in figure 3. As in the previous figure, we again show χ^2 for the best-fit velocity distribution for each DM mass. For comparison, we also show the χ^2 curves with the correct form factor from figure 2 as dashed lines.

When fitting silicon and germanium data simultaneously with the incorrect form factor choices, we observe best-fit values of χ^2 between 10^3 and 10^4 over the entire mass range for models A, B and D. Only for the light-mediator model C there is a narrow range of masses below approximately 10 MeV where we obtain a good fit despite the wrong DM form factor. This indicates that a light DM particle with a heavy mediator can mimic the event rates of a heavier DM particle with a light mediator. This is not unexpected, since both scenarios lead to a concentration of events at low energies.

The results for models A, B, and D, on the other hand, show that the combined rates in silicon and germanium from a heavy-mediator model cannot be explained by a light-mediator model with any DM mass or velocity distribution. This implies that our halo-independent analysis method can correctly identify that the coupling is through a heavy mediator even without any input about the DM mass or velocity.

Having studied the ability of our formalism to identify the particle properties of dark matter, we next turn to the DM velocity distribution. The best-fit results for $\tilde{\eta}$ for the correct masses of models A–D are shown in figure 4 alongside the $\tilde{\eta}$ of the true velocity distributions for each model. These are the Standard Halo Model defined in (3.1) for models A–C and a stream localized at $v_{\rm str} = 300$ km/s for model D, see (3.2). Around each best-fit velocity distribution we also show the corresponding REST, which accounts for the flat directions in velocity space, as discussed in section 2.4.

For all four models considered here we find that the best-fit velocity distribution follows the true velocity distribution nearly perfectly. Note that the part of the velocity spectrum below the threshold $v_*(E_{\rm gap})$ (as defined in (2.6)) does not affect the observed event numbers and therefore remains at its arbitrary initial value throughout the minimization process. Hence the best-fit $\tilde{\eta}$ curves reach a plateau at the lowest velocities shown in figure 4. Accounting for this, the $\tilde{\eta}$ of the true velocity distribution above threshold is contained inside

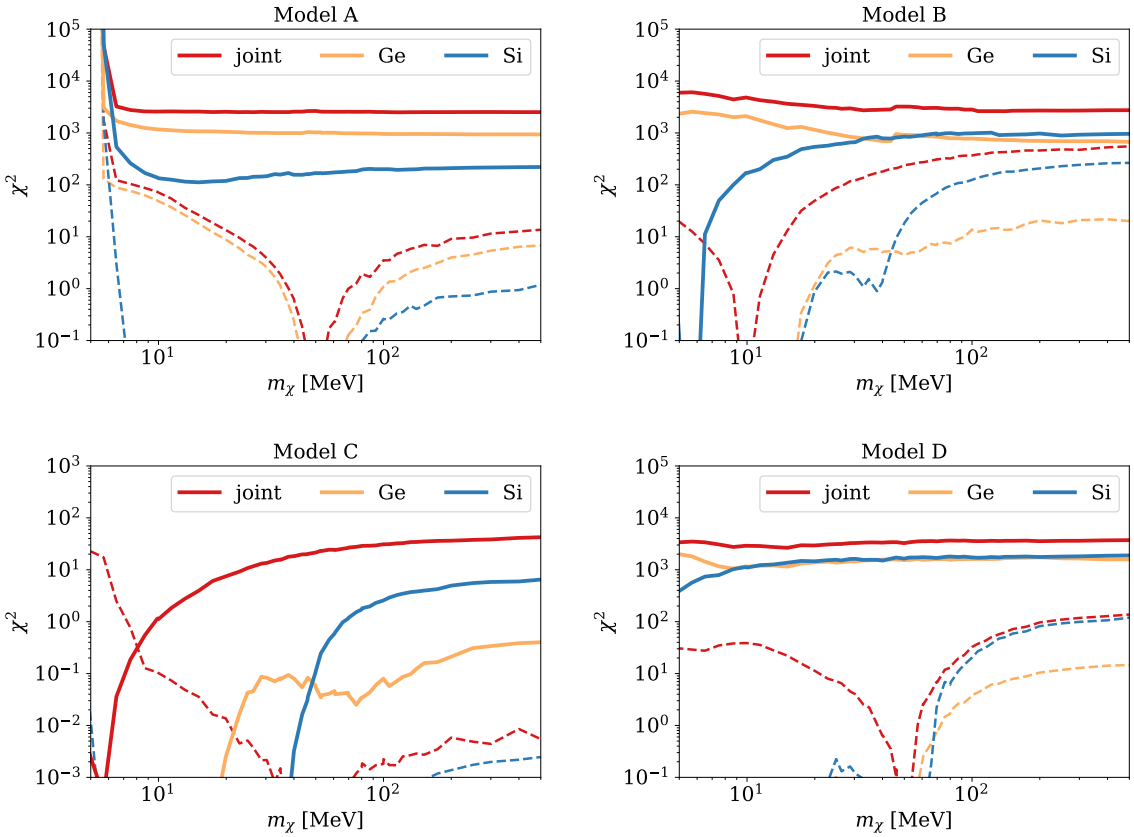

Figure 3: Same as figure 2, but with the incorrect DM form factors, i.e. $F_{\rm DM} \sim 1/q^2$ for models A, B and D, and $F_{\rm DM} = 1$ for model C. The results from figure 2, which use the correct form factors, are shown with dashed lines (note the change to a logarithmic scale).

the REST for all models.

3.2 Limits on the dark matter mass

So far we have discussed the best-fit χ^2 only in terms of its qualitative behavior as a function of DM mass. In this section, we proceed to use it quantitatively as a test statistic in order to set limits on the DM mass based on our simulated data. Figure 5 shows the χ^2 curves of the Asimov[4] datasets for models A–D (dashed red lines) together with the χ^2 value that corresponds to a 90% CL exclusion for each dataset (dotted black line). The number of degrees of freedom of the χ^2 distribution underlying each of these thresholds is equal to the sum of the number of silicon bins and the number of germanium bins taken into account in the analysis. Here, we only use bins with an expected rate ≥ 5 events to ensure that our test statistic is χ^2-distributed to good approximation. The resulting numbers of degrees of freedom and the corresponding 90% CL χ^2 thresholds for each model are summarized in table 2.

[4]In the Asimov dataset [29] the number of observed events is exactly the number of expected events, i.e. the Poisson distribution in the limit of infinite exposure.

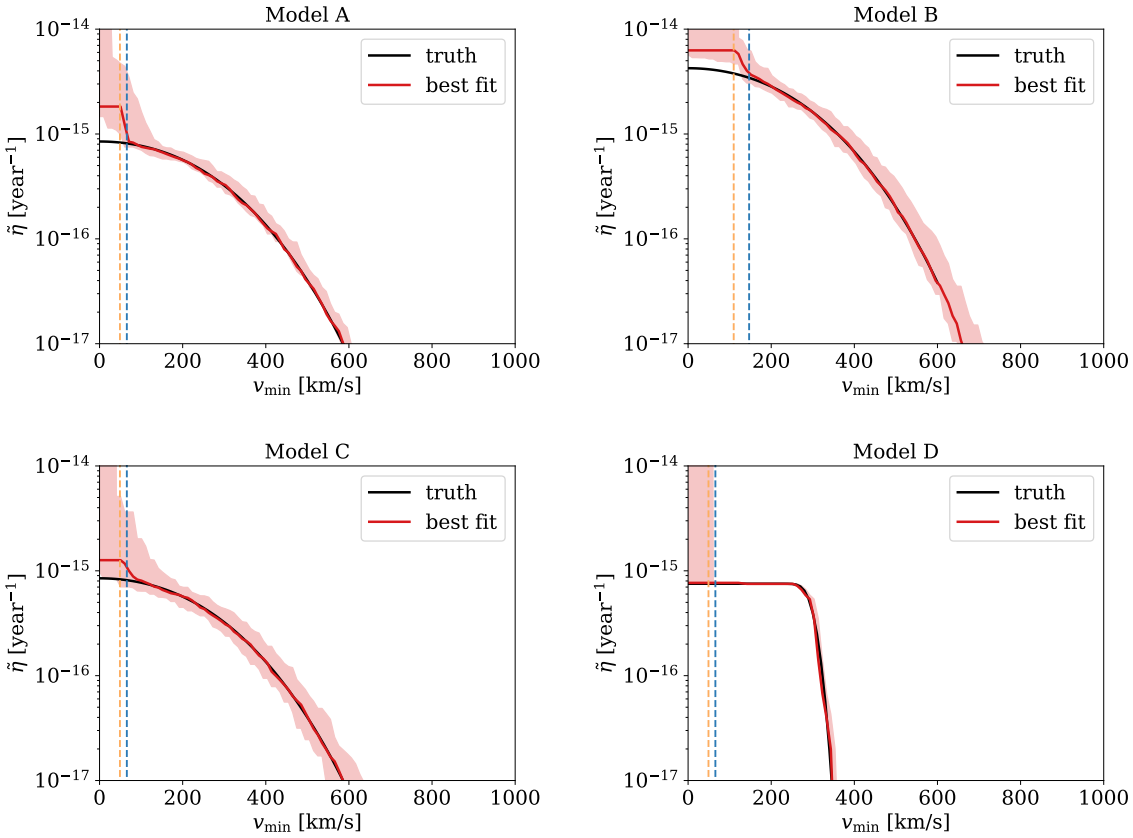

Figure 4: Best-fit velocity distributions, shown in the form of $\tilde{\eta}(v_{\min})$, for the correct DM masses and DM form factors for the models from figure 2 inferred from a joint fit of simulated Si and Ge rates. The threshold velocities (as given in (2.6)) for Si and Ge are shown as blue and orange dashed lines, respectively. The bands are defined by the REST procedure described in section 2.4.

model	d.o.f. (Si+Ge)	$\chi^2_{90\%}$	m_χ [MeV]	m_χ lower limit [MeV]	m_χ upper limit [MeV]
A	9+10	27.2	50	15	–
B	6+7	19.8	10	5	16
C	5+5	16.0	50	6	–
D	5+5	16.0	50	17	83

Table 2: Number of degrees of freedom relevant for setting a limit on m_χ, associated 90% CL χ^2 threshold, and upper and lower bounds from the Asimov dataset at an exposure of 1 kg-yr for each model.

Using the best-fit χ^2 curve as a function of m_χ, we can translate the χ^2 thresholds into 90% CL exclusion limits on the DM mass. For the Asimov datasets of models A–D, this yields the exclusion shaded in dark gray in figure 5. The lower and upper bound of the allowed mass range for each model are given in the last column of table 2.

In the Asimov dataset the number of observed events is exactly equal to the expected

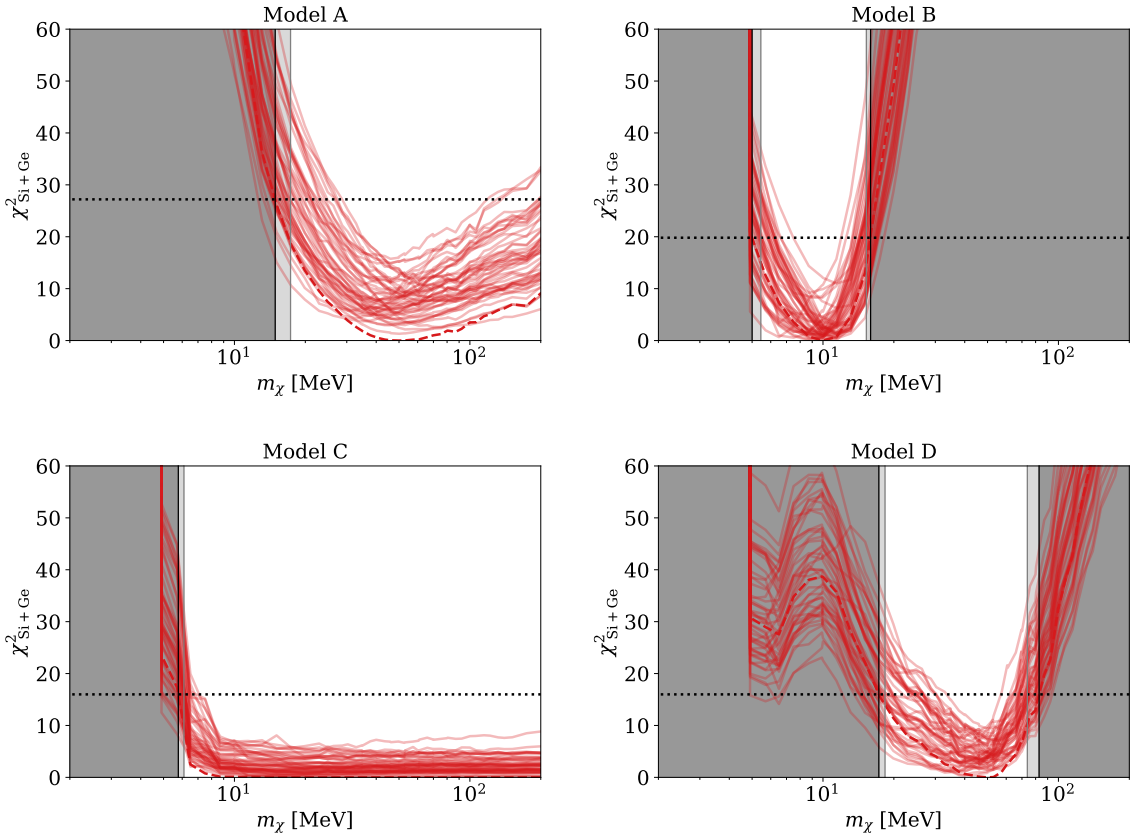

Figure 5: Joint Ge and Si χ^2 for the best-fit velocity distribution as a function of the DM mass for the Asimov dataset (dashed red line) and 50 Poisson draws (solid red lines) at an exposure of 1 kg-yr. The dashed black line indicates the χ^2 that corresponds to a 90% CL exclusion limit for the relevant number of degrees of freedom (see table 2). The dark gray shaded region indicates the associated excluded mass range based on the Asimov dataset. The light gray shaded region is the median excluded mass range derived from the Poisson draws.

rate. To check how the mass limits obtained from this idealized construction compare to a more realistic scenario with Poisson fluctuations in the data, we generate 50 Poisson draws from the expected rates for each model. To this end, we draw event numbers for each model and experimental bin from the Poisson distribution with mean equal to the expected rate in the respective bin. We then derive mass limits based on each associated χ^2 curve, which are all shown in figure 5. The median excluded mass range, corresponding to the median of the lower limits of the 50 draws and the median of their upper limits, is shaded in gray. For all of our models we find that the median exclusion from the Poisson draws is almost identical to the exclusion constructed from the Asimov dataset. This confirms that the latter is robust under statistical fluctuations and hence provides a good approximation to a realistic projected limit.

For models B and D we obtain an upper as well as a lower bound at 90% CL from the Asimov dataset for the exposure of 1 kg-yr used in figure 5. For model A an approximately

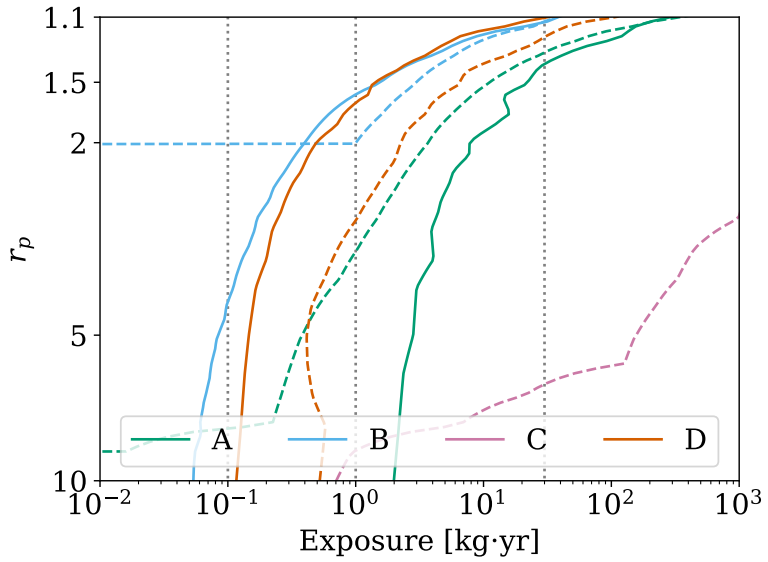

Figure 6: Relative precision r_p for the upper bound (solid lines) and lower bound (dashed lines) on the mass at 90% CL as a function of exposure for each model. The relative precision is defined as the ratio of the limit to the true mass (the true mass to the limit) for the upper (lower) bound (see main text for details). Future exposure targets of 0.1 kg-yr, 1 kg-yr and 30 kg-yr are indicated by vertical dotted lines.

twice as large exposure is necessary to find an upper bound. This can be read off figure 6, which for each model shows the minimum exposure that yields a certain relative precision for the upper and lower limits on the DM mass. For the upper and lower bound we define the relative precision as $r_p \equiv m_\chi^{\max}/m_\chi$ and $r_p \equiv m_\chi/m_\chi^{\min}$, respectively, where $m_\chi^{\max}$ denotes the upper limit on the DM mass, $m_\chi^{\min}$ the lower limit and m_χ the true value. Larger exposures yield smaller values of r_p, i.e. more precise limits.

For the light-mediator model C no upper bound can be obtained for any exposure since arbitrarily large DM masses still yield a perfect fit to the Asimov dataset for some velocity distribution (see figure 5). For the other models there exists a minimum exposure above which an upper limit at a given confidence level can be found. In the case of a 90% CL bound, this minimum exposure lies below the value of 1 kg-yr chosen for figure 5 for models B and D, and above it for model A.

4 Application to data

The mock-data based examples in the previous section provide a closure test which demonstrates that our procedure can reliably find the true dark matter model. We now apply our procedure to actual experimental data. There are multiple experiments which have searched for DM-electron recoils: DAMIC [30], DarkSide-50 [31], EDELWEISS [32], PandaX [33], SENSEI [34], SuperCDMS [35], XENON10 [21], and XENON1T [36, 37]. So far, none has found definitive evidence for DM-electron recoils. Rather than carry out a comprehensive comparison across all experiments we illustrate our technique by focusing on two

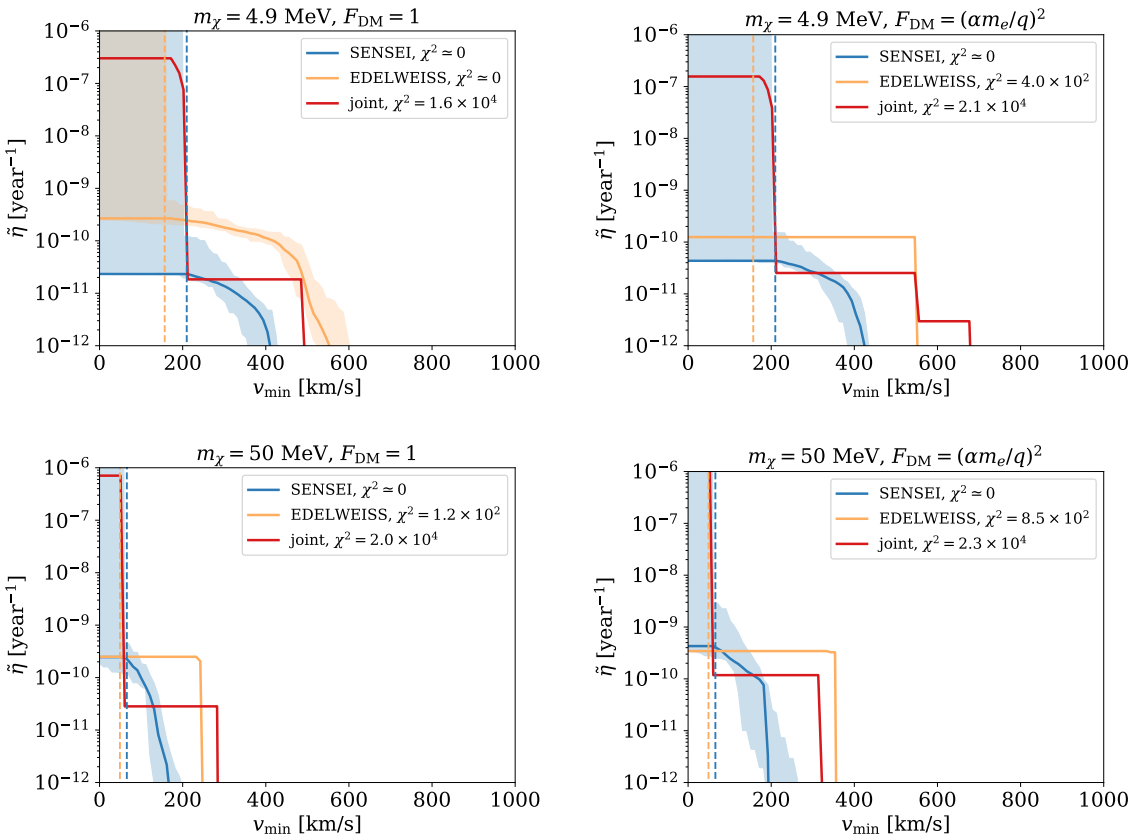

Figure 7: Best-fit velocity distributions inferred from fits of data from SENSEI, EDELWEISS, and both experiments simultaneously, shown for $m_\chi = 4.9$ MeV (top) $m_\chi = 50.0$ MeV (bottom) and for a heavy (left) or light (right) mediator. The bands denote the REST regions, as defined in section 2.4. We show these only for best fits with negligible χ^2.

recent semiconductor-based searches, one using silicon (SENSEI) and one using germanium (EDELWEISS) as the target. However, our procedure can be easily expanded to include more searches.

SENSEI uses silicon Skipper CCDs to search for low-energy electron recoils [38]. Silicon has a band gap of 1.2 eV and requires 3.8 eV for the creation of an additional electron-hole pair. In 2020, the collaboration took varying amounts of data up to the 4-electron bin [34]. After adjusting for exposure and misclassification of $1e^-$ events the observed number of events in the $(1e^-, 2e^-, 3e^-, 4e^-)$ bins was $(1311.7, 5, 0, 0)$. After subtracting spurious-charge background events from the first bin, this translates[5] to a rate of $(549, 2.39, 0, 0)/(\text{g} \cdot \text{day})$. Note that the properties of the materials used in experiments, *e.g.* the band gap and the crystal form factor, are subject to systematic uncertainties. For simplicity, we do not vary these quantities within their uncertainties in the present demonstration of the halo-independent

[5]To be conservative, we subtract the 2σ lower bound on the spurious-charge background, 554 events. Note that to convert the number of observed events to a rate the exposure, which varies across electron bins, must be taken into account.

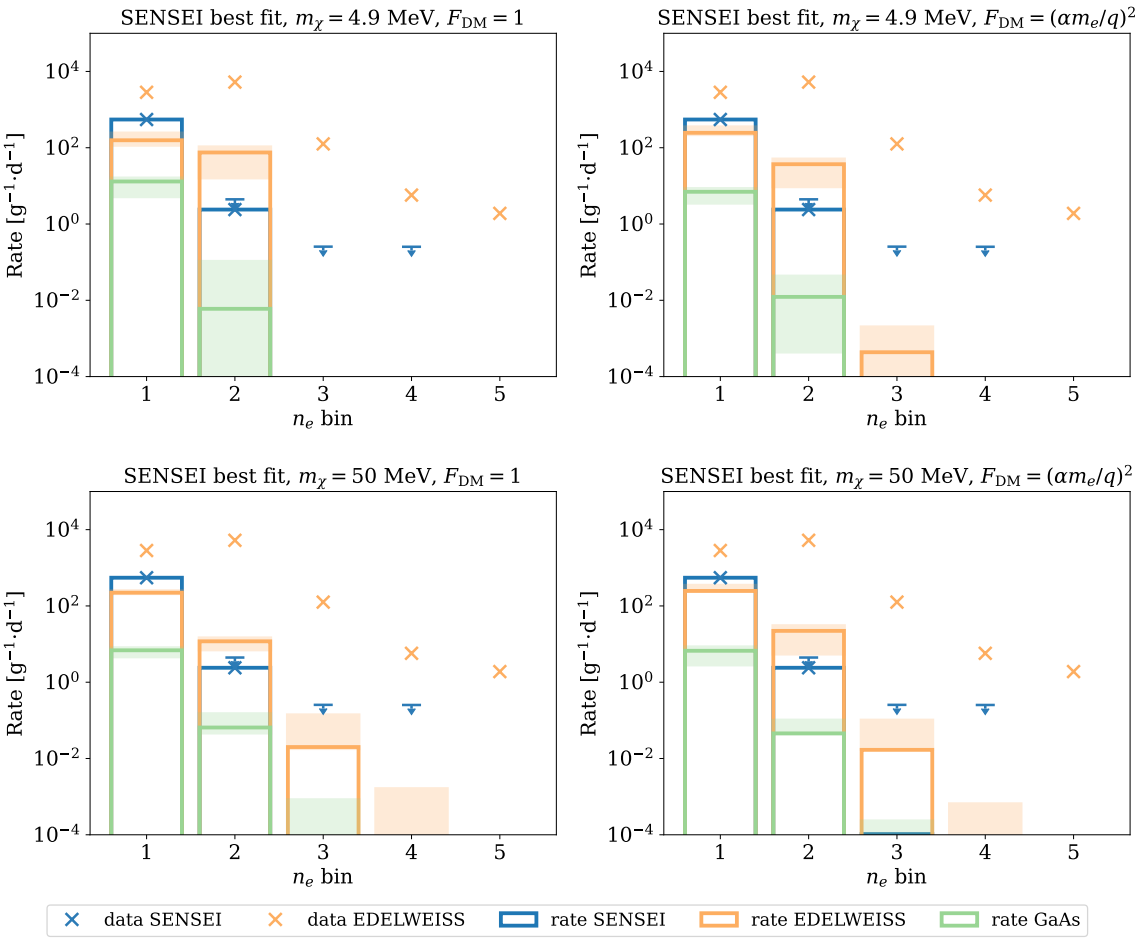

Figure 8: Predictions of the signal rates in EDELWEISS and a future GaAs detector based on the best-fit velocity distribution from SENSEI data, shown for $m_\chi = 4.9$ MeV (top) and $m_\chi = 50$ MeV (bottom) and for a heavy (left) or light (right) mediator. For the second, third and fourth SENSEI bin the 90% CL upper limit on the rate is indicated with downward arrows. The shaded areas indicate the span of the predictions for EDELWEISS and GaAs from the REST band of the SENSEI fit.

method, but a combined scan over material form factors and velocity distributions could be conducted efficiently using the vector space integration method of [39]

EDELWEISS uses germanium detectors utilizing Neganov-Trofimov-Luke amplification to go beyond electron-hole pair resolution [40, 41]. Germanium has a somewhat smaller band gap and excitation energy than silicon, of 0.67 eV and 3 eV, respectively. In 2020, the EDELWEISS collaboration performed a search for DM-electron scattering using 33.4 g of germanium [32] and 58 hours of data collection. In this search, the DM signal region extended to 30 eV_{ee}, but here we only consider the data from the first 5 electron bins. Adjusting for the efficiency of signals to pass selection cuts and integrating the observed energy spectrum across each electron bin, we find a rate for electron recoils of $(1e^-, 2e^-, 3e^-, 4e^-, 5e^-) = (2840, 5260, 126, 5.7, 1.9)$ events/(g · day). Since there is presently not a model for possible

backgrounds, we do not carry out any background subtraction.

There are hypothesized sources of background for these experiments, see *e.g.* [42, 43] for discussions of potential sources of single-electron events in Skipper-CCDs. Nevertheless, one can ask what DM velocity distribution is required if these rates are pure signal and whether they can be compatible with one another.

For the SENSEI dataset it is possible to find a physically viable velocity distribution that perfectly fits the observed data for any DM mass; for both a contact interaction $F_{\rm DM} = 1$ and a long-range interaction $F_{\rm DM} \sim q^{-2}$. EDELWEISS has a considerably larger event rate and has events out to higher energies than SENSEI. Thus, their best-fit velocity distribution extends to higher speeds, and has a larger normalization. Moreover, only a heavy mediator ($F_{\rm DM} = 1$) and DM masses $m_\chi \lesssim 20$ MeV result in a good fit to the EDELWEISS data. For other DM hypotheses, a portion of the observed events has to originate from background, independent of the DM velocity distribution. Finally, when taken together, both experiments cannot be simultaneously fit for any choice of mass or coupling structure.

We illustrate these findings in figure 7, where we show the best-fit velocity curves for SENSEI and EDELWEISS, as well as the joint fit for a selection of DM parameter choices. A DM mass of 4.9 MeV, shown in the top panels, provided the best joint fit, while a heavier mass of 50 MeV is shown for comparison in the lower panels. We see that even for the best-fit point of $m_\chi = 4.9$ MeV, with the REST bands, the velocity distributions are incompatible, as demonstrated by the large χ^2 value of $\chi^2 = 1.6\ (2.1) \times 10^4$ for the heavy (light) mediator, respectively. Although we did not calculate beyond REST, in principle one can go beyond REST by following the procedure discussed in section 2.4. However, given the large size of the χ^2 value in this example, going beyond REST does not provide any additional information.

Many of the best-fit velocity distributions in figure 7 are very different from the Standard Halo Model. In fact, only the $m_\chi = 4.9\,\text{MeV}$, $F_{\rm DM} = 1$ fit to the EDELWEISS data resembles the SHM (see models A–C in figure 4 for a comparison). With $m_\chi = 50\,\text{MeV}$, for example, the best fit to the SENSEI data is a co-rotating halo or disc with unexpectedly low dispersion. The power of the halo-independent approach is its ability to check the consistency of possible signals from multiple experiments, without imposing any prior for the distribution other than it be physical. Thus, if the halo-independent approach finds two datasets to be inconsistent there is no physically allowed distribution that explains them both.

When comparing two experiments, there is another question one can ask; given a best-fit velocity profile for one set of experimental data, what should we expect in a different experiment? Specifically, we use the smaller velocity distribution required by SENSEI to predict a rate at EDELWEISS, under the (unlikely) assumption that SENSEI has no background events in their data. The expected rates at EDELWEISS for a DM particle of 4.9 MeV and 50 MeV are shown in figure 8, for two choices of the DM form-factor. The intervals between the lowest and the highest rate predicted by velocity distributions within the REST band of the fit to SENSEI data are shown as shaded areas in each bin. The large shortfall between the predicted EDELWEISS rate and the observed rate points to a substantial background.

As a complementary exercise, we also show the predicted rate in a third, hypothetical experiment using GaAs. We see that the predicted rate for GaAs is lower than that of SENSEI. Thus, if this hypothetical GaAs experiment observes a higher rate than this prediction, then the experiment has a non-trivial background. On the other hand, if the GaAs experiment observes a lower rate, then this implies that the observed events in the SENSEI data are also contaminated by some background.

5 Discussion and Conclusions

The astrophysical DM-velocity halo distribution is a key ingredient to predicting the rate at dark matter direct detection experiments. Equivalently, converting direct detection experimental data into insights about the particle physics properties of dark matter, such as its mass and couplings, depends upon assumptions about the astrophysics. However, the halo model is subject to large uncertainties. These uncertainties propagate to the predictions for DM direct detection rates, which motivates a formalism in which one can present the results of DM direct detection experiments in a halo-independent manner.

In this work, we presented a procedure which extends the halo-independent formalism, previously applied to DM-nuclear recoils, to DM-electron scattering. We use this approach to find the best-fit halo velocity distribution for a given set of data and DM parameters. Profiling over these velocity distributions it is possible to determine the DM parameters that yield the best fit to the data, as well as exclude masses which do not fit the data for any physically-allowed velocity distribution. Our procedure includes the Region of Equivalent Statistical Test (REST), which encapsulates the velocity models which report the same best-fit χ^2 values and thus cannot be distinguished on the basis of the data (see section 2.4).

We first performed a closure test in which we created mock data and then demonstrated that our procedure can accurately determine the true DM parameters and halo velocity distribution used to simulate the data. Secondly, we applied our procedure to two current experiments, EDELWEISS and SENSEI, to determine if there exists a viable DM model and halo velocity distribution which can simultaneously explain all the observed events at both experiments.

From the closure test, we observed that the procedure works well for DM models with a heavy mediator, but is not as constraining for a light mediator. Rates from DM models with a light mediator are enhanced at low-momentum transfer, and therefore low energies. As a consequence, they will only populate the lowest few energy bins. In contrast, a heavy mediator will populate more energy bins, which provides more information for the analysis. Generally, the efficacy of our halo-independent approach increases with the number of non-zero data bins. For similar reasons our procedure provides tighter constraints as the dark matter mass increases.

The result of our halo-independent analysis of the SENSEI and EDELWEISS data indicates that the data cannot be explained by DM-electron scattering for *any* DM velocity distribution if both experiments are free of background. Furthermore, if the lower rate observed at SENSEI is due entirely to dark matter, EDELWEISS has a substantial background. Finally, we predict that a GaAs experiment searching for DM-electron scattering [44, 45] should see a DM rate between ~ 3 events/(g-day) and ~ 20 events/(g-day), if SENSEI's rate is entirely due to dark matter.

Although we only analyzed data from semiconductor experiments, the technique outlined above is applicable to noble gas detectors, once the analogous ion form factors are calculated. The rates in this work were generated using the `QEDark` code. However, our procedure can be readily adapted to other codes, which can also compute $f_{\rm res}$ (2.2). The remainder of the procedure is unaffected by the choice of DM-electron scattering code. The procedure we outline parameterizes the velocity distribution as a piecewise-constant monotonic function with $\mathcal{O}(100)$ coefficients that must be determined. Despite the large number of parameters, this minimization procedure is fast and takes only $\mathcal{O}(10$ min) per model point.

The halo-independent DM-electron scattering formalism presented here can also be

extended to other types of DM models, such as inelastic DM, or to anisotropic detector materials. The formalism for inelastic DM-nuclear scattering was studied in [46, 47], where the authors extended the elastic nuclear recoil halo-independent formalism to account for the fact that the nuclear recoil energy does not uniquely map to $v_{\rm min}$, but instead has 2 solutions. For anisotropic detector materials, the wavelet-harmonic method of [48] provides a suitable basis for the 3d velocity distributions and detector form factors, analogous to the linear basis of velocity functions in (2.9).

If a robust DM signal is observed, the proposed formalism should be used to extract rigorous conclusions about the properties of DM. Additionally, it will be the only way to measure the local DM velocity distribution. The insights gained through this formalism could then inform the next experimental steps, including which other materials should be used to build follow-up detectors.

Acknowledgments

We thank Felix Kahlhoefer and Volodymyr Takhistov for helpful discussions. Fermilab is administered by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The work of B.L. was supported in part by the U.S. Department of Energy under grant number DE-SC0011640. T.-T.Y. is supported in part by NSF CAREER grant PHY-1944826. T.-T.Y. also thanks the Center for Cosmology and Particle Physics at NYU for hospitality and support, where a portion of this work was completed.

References

[1] R. Essig et al., *Snowmass2021 Cosmic Frontier: The landscape of low-threshold dark matter direct detection in the next decade*, in *Snowmass 2021*, 3, 2022, 2203.08297.

[2] J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter et al., *Clumps and streams in the local dark matter distribution*, *Nature* **454** (2008) 735.

[3] A. A. Klypin, S. Trujillo-Gomez and J. Primack, *Halos and galaxies in the standard cosmological model: results from the bolshoi simulation*, *The Astrophysical Journal* **740** (2011) 102.

[4] J. Guedes, S. Callegari, P. Madau and L. Mayer, *Forming realistic late-type spirals in a λcdm universe: the eris simulation*, *The Astrophysical Journal* **742** (2011) 76.

[5] P. F. Hopkins, A. Wetzel, D. Kereš, C.-A. Faucher-Giguère, E. Quataert, M. Boylan-Kolchin et al., *Fire-2 simulations: physics versus numerics in galaxy formation*, *Monthly Notices of the Royal Astronomical Society* **480** (2018) 800.

[6] J. Bovy and S. Tremaine, *On the local dark matter density*, *Astrophys. J.* **756** (2012) 89 [1205.4033].

[7] P. J. Fox, G. D. Kribs and T. M. P. Tait, *Interpreting Dark Matter Direct Detection Independently of the Local Velocity and Density Distribution*, *Phys. Rev. D* **83** (2011) 034007 [1011.1910].

[8] P. J. Fox, J. Liu and N. Weiner, *Integrating Out Astrophysical Uncertainties*, *Phys. Rev. D* **83** (2011) 103514 [1011.1915].

[9] P. Gondolo and G. B. Gelmini, *Halo independent comparison of direct dark matter detection data*, *JCAP* **12** (2012) 015 [1202.6359].

[10] E. Del Nobile, G. B. Gelmini, P. Gondolo and J.-H. Huh, *Halo-independent analysis of direct detection data for light WIMPs*, *JCAP* **10** (2013) 026 [1304.6183].

[11] B. Feldstein and F. Kahlhoefer, *A new halo-independent approach to dark matter direct detection analysis*, *JCAP* **08** (2014) 065 [1403.4606].

[12] P. J. Fox, Y. Kahn and M. McCullough, *Taking Halo-Independent Dark Matter Methods Out of the Bin*, *JCAP* **10** (2014) 076 [1403.6830].

[13] B. Feldstein and F. Kahlhoefer, *Quantifying (dis)agreement between direct detection experiments in a halo-independent way*, *JCAP* **12** (2014) 052 [1409.5446].

[14] A. J. Anderson, P. J. Fox, Y. Kahn and M. McCullough, *Halo-Independent Direct Detection Analyses Without Mass Assumptions*, *JCAP* **10** (2015) 012 [1504.03333].

[15] A. Radick, A.-M. Taki and T.-T. Yu, *Dependence of Dark Matter - Electron Scattering on the Galactic Dark Matter Velocity Distribution*, *JCAP* **02** (2021) 004 [2011.02493].

[16] T. N. Maity, T. S. Ray and S. Sarkar, *Halo uncertainties in electron recoil events at direct detection experiments*, *Eur. Phys. J. C* **81** (2021) 1005 [2011.12896].

[17] M. Chen, G. B. Gelmini and V. Takhistov, *Halo-independent dark matter electron scattering analysis with in-medium effects*, *Phys. Lett. B* **841** (2023) 137922 [2209.10902].

[18] M. Chen, G. B. Gelmini and V. Takhistov, *Halo-independent analysis of direct dark matter detection through electron scattering*, *JCAP* **12** (2021) 048 [2105.08101].

[19] R. Essig, M. Fernandez-Serra, J. Mardon, A. Soto, T. Volansky and T.-T. Yu, *Direct Detection of sub-GeV Dark Matter with Semiconductor Targets*, *JHEP* **05** (2016) 046 [1509.01598].

[20] 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 [1206.2644].

[21] R. Essig, T. Volansky and T.-T. Yu, *New Constraints and Prospects for sub-GeV Dark Matter Scattering off Electrons in Xenon*, *Phys. Rev. D* **96** (2017) 043017 [1703.00910].

[22] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *arXiv preprint arXiv:1412.6980* (2014) .

[23] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.

[24] S. Knapen, J. Kozaczuk and T. Lin, *python package for dark matter scattering in dielectric targets*, *Phys. Rev. D* **105** (2022) 015014 [2104.12786].

[25] T. Trickle, *Extended calculation of electronic excitations for direct detection of dark matter*, *Phys. Rev. D* **107** (2023) 035035 [2210.14917].

[26] C. E. Dreyer, R. Essig, M. Fernandez-Serra, A. Singal and C. Zhen, *Fully ab-initio all-electron calculation of dark matter–electron scattering in crystals with evaluation of systematic uncertainties*, 2306.14944.

[27] D. Baxter et al., *Recommended conventions for reporting results from direct dark matter searches*, *Eur. Phys. J. C* **81** (2021) 907 [2105.00599].

[28] R. Essig, J. Mardon and T. Volansky, *Direct Detection of Sub-GeV Dark Matter*, *Phys. Rev. D* **85** (2012) 076007 [1108.5383].

[29] 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 [1007.1727].

[30] DAMIC-M collaboration, *First Constraints from DAMIC-M on Sub-GeV Dark-Matter Particles Interacting with Electrons*, *Phys. Rev. Lett.* **130** (2023) 171003 [2302.02372].

[31] DarkSide collaboration, *Search for Dark Matter Particle Interactions with Electron Final States with DarkSide-50*, *Phys. Rev. Lett.* **130** (2023) 101002 [2207.11968].

[32] EDELWEISS collaboration, *First germanium-based constraints on sub-MeV Dark Matter with the EDELWEISS experiment*, *Phys. Rev. Lett.* **125** (2020) 141301 [2003.01046].

[33] PandaX-II collaboration, *Search for Light Dark Matter-Electron Scatterings in the PandaX-II Experiment*, *Phys. Rev. Lett.* **126** (2021) 211803 [2101.07479].

[34] SENSEI collaboration, *SENSEI: Direct-Detection Results on sub-GeV Dark Matter from a New Skipper-CCD*, *Phys. Rev. Lett.* **125** (2020) 171802 [2004.11378].

[35] SuperCDMS collaboration, *First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector*, *Phys. Rev. Lett.* **121** (2018) 051301 [1804.10697].

[36] XENON collaboration, *Light Dark Matter Search with Ionization Signals in XENON1T*, *Phys. Rev. Lett.* **123** (2019) 251801 [1907.11485].

[37] XENON collaboration, *Emission of single and few electrons in XENON1T and limits on light dark matter*, *Phys. Rev. D* **106** (2022) 022001 [2112.12116].

[38] SENSEI collaboration, *Single-electron and single-photon sensitivity with a silicon Skipper CCD*, *Phys. Rev. Lett.* **119** (2017) 131802 [1706.00028].

[39] B. Lillard, *Vector Space Integration for Dark Matter Scattering*, 2310.01480.

[40] P. N. Luke, *Voltage-assisted calorimetric ionization detector*, *Journal of Applied Physics* **64** (1988) 6858 [https://pubs.aip.org/aip/jap/article-pdf/64/12/6858/8002883/6858_1_online.pdf].

[41] B. S. Neganov and V. N. Trofimov, *Colorimetric method measuring ionizing radiation*, *Otkryt. Izobret.* **146** (1985) 215.

[42] P. Du, D. Egana-Ugrinovic, R. Essig and M. Sholapurkar, *Sources of Low-Energy Events in Low-Threshold Dark-Matter and Neutrino Detectors*, *Phys. Rev. X* **12** (2022) 011009 [2011.13939].

[43] SENSEI collaboration, *SENSEI: Characterization of Single-Electron Events Using a Skipper Charge-Coupled Device*, *Phys. Rev. Applied* **17** (2022) 014022 [2106.08347].

[44] S. Derenzo, R. Essig, A. Massari, A. Soto and T.-T. Yu, *Direct Detection of sub-GeV Dark Matter with Scintillating Targets*, *Phys. Rev. D* **96** (2017) 016026 [1607.01009].

[45] B. Penning, "The TESSERACT Project." Contribution to Dark Matter 2023.

[46] N. Bozorgnia, J. Herrero-Garcia, T. Schwetz and J. Zupan, *Halo-independent methods for inelastic dark matter scattering*, *JCAP* **07** (2013) 049 [1305.3575].

[47] S. Scopel and K. Yoon, *A systematic halo-independent analysis of direct detection data within the framework of Inelastic Dark Matter*, *JCAP* **08** (2014) 060 [1405.0364].

[48] B. Lillard, *Wavelet-Harmonic Integration Methods*, 2310.01483.

[49] R. Penrose, *A generalized inverse for matrices*, *Mathematical Proceedings of the Cambridge Philosophical Society* **51** (1955) 406–413.

A Halo Models

Here, we present the analytic expressions for our two representative halo models, used in generating our mock data.

Standard Halo Model

In the galactic frame the Standard Halo Model is a Maxwell-Boltzmann distribution with a cutoff

$$f_\chi(\vec{v}_\chi) = \frac{1}{K_{SHM}} e^{-v_\chi^2/v_0^2}\,\Theta(v_{\rm esc} - v_\chi). \tag{A.1}$$

The normalization factor K_{SHM} is determined by requiring $\int d^3v f_\chi(\vec{v}) = 1$, giving

$$K_{SHM} = v_0^3\pi\left[\sqrt{\pi}{\rm Erf}\left(\frac{v_{\rm esc}}{v_0}\right) - 2\frac{v_{\rm esc}}{v_0}e^{-\left(\frac{v_{\rm esc}}{v_0}\right)^2}\right]. \tag{A.2}$$

We define the function $\eta(v_{\rm min})$, as in (2.3),

$$\eta(v_{min}) = \int d^3v\,\frac{e^{-(\vec{v}+\vec{v}_E)^2/v_0^2}}{v\,K_{SHM}}\,\Theta(v - v_{\rm min})\,\Theta(v_{\rm esc} - |\vec{v}+\vec{v}_E|)\,. \tag{A.3}$$

The result of the integral takes a different form depending on whether $v_{\rm min}$ is bigger or smaller than $v_{\rm esc} - v_E$. For $v_{\rm min} < v_{\rm esc} - v_E$,

$$\begin{aligned}\eta(v_{\rm min}) = \frac{v_0^2\pi}{2v_E K_{SHM}}\Big[&-4e^{-v_{\rm esc}^2/v_0^2}v_E \\ &+\sqrt{\pi}v_0\left(\mathrm{erf}\left(\frac{v_{\rm min}+v_E}{v_0}\right) - \mathrm{erf}\left(\frac{v_{\rm min}-v_E}{v_0}\right)\right)\Big]\,,\end{aligned} \tag{A.4}$$

while for $v_{\rm esc} - v_E < v_{\rm min} < v_{\rm esc} + v_E$

$$\begin{aligned}\eta(v_{\rm min}) = \frac{v_0^2\pi}{2v_E K_{SHM}}\Big[&-2e^{-v_{\rm esc}^2/v_0^2}(v_{\rm esc} - v_{\rm min} + v_E) \\ &+\sqrt{\pi}v_0\left(\mathrm{erf}\left(\frac{v_{\rm esc}}{v_0}\right) - \mathrm{erf}\left(\frac{v_{\rm min}-v_E}{v_0}\right)\right)\Big]\,.\end{aligned} \tag{A.5}$$

Stream

We approximate the stream's velocity distribution as a narrow Gaussian, which in the Earth's frame takes the form

$$f(\vec{v}) = \frac{1}{\sqrt{8\pi^3\sigma^3}}e^{-\frac{(\vec{v}-\vec{v}_{\rm str})^2}{2\sigma^2}}\,. \tag{A.6}$$

We are focused on the case where the stream's velocity is well below the galactic escape speed, $v_{\rm esc} - |\vec{v}_{\rm str} - \vec{v}_E| \gg \sigma$, and it is a reasonable approximation to integrate over all speeds in the Earth's frame, *i.e.* we do not include a Heaviside cutoff. As before, the velocity distribution is normalized such that $\int d^3v f(v) = 1$. The function η can be found by using (A.5) and making the substitutions $v_E \to v_{\rm str}$, $v_0 \to \sqrt{2}\sigma$ and taking the limit $v_{\rm esc} \to \infty$ resulting in

$$\eta(v_{\rm min}) = \frac{1}{2v_{\rm str}}\left(\mathrm{erf}\left(\frac{v_{\rm min}+v_{\rm str}}{\sqrt{2}\sigma}\right) - \mathrm{erf}\left(\frac{v_{\rm min}-v_{\rm str}}{\sqrt{2}\sigma}\right)\right)\,. \tag{A.7}$$

B The Pseudoinverse

Consider an $(m \times n)$ matrix A with rank k whose, singular value decomposition is

$$A = UDV^T\,, \tag{B.1}$$

with U an orthogonal $(m \times m)$ matrix, V an orthogonal $(n \times n)$ matrix and D a diagonal $(m \times n)$ matrix, of the form

$$D = \begin{pmatrix} D_k & 0 \\ 0 & 0 \end{pmatrix} \ . \tag{B.2}$$

The rank k square matrix D_k is diagonal, with $(D_k)_{ii} \neq 0$. The singular values of A are the diagonal entries of D, while D_k includes only the nonzero singular values of A. By definition, $\det D_k \neq 0$. The pseudoinverse of D is defined as

$$D^+ = \begin{pmatrix} D_k^{-1} & 0 \\ 0 & 0 \end{pmatrix} \ . \tag{B.3}$$

The pseudoinverse of A is defined in terms of D^+,

$$A^+ = V D^+ U^T \ . \tag{B.4}$$

This pseudoinverse is the unique matrix [49] with the properties

$$AA^+A = A \ , \qquad A^+AA^+ = A^+ \ , \tag{B.5}$$

$$(AA^+)^T = AA^+ \ , \qquad (A^+A)^T = A^+A \ . \tag{B.6}$$

Furthermore, AA^+ acts as a projection operator, so that any vector can be projected into orthogonal image spaces using AA^+ and $\mathbb{1} - AA^+$, for example $(\mathbb{1} - AA^+)x \in \ker(A)$.

Finally, the least squares solution to the equation $Ax = b$ is given by $x_0 = A^+b$. This can be seen by using the properties of the pseudoinverse to write

$$Ax - b = AA^+ \left[A \left(x - A^+ b \right) \right] - \left(\mathbb{1} - AA^+ \right) b \ , \tag{B.7}$$

where we have used the projection operators to split the expression into pieces sitting in the two orthogonal spaces. Thus, the norm of $Ax - b$ is

$$|Ax - b|^2 = |A \left(x - A^+ b \right)|^2 + |Ax_0 - b|^2 \geq |Ax_0 - b|^2 \ . \tag{B.8}$$

This shows that the pseudoinverse provides us with the least squares solution, and x_0 is the solution of minimum norm. There may be other solutions, with larger norm, related to x_0 by an arbitrary vector in the kernel of A:

$$x \to x_0 + \left(\mathbb{1} - AA^+ \right) y \ . \tag{B.9}$$

For the case of an underdetermined system $(n > m)$ an exact solution $Ax_0 = b$ can always be found, provided the SVD of A has at least m nonzero singular values. For an overdetermined system $(n < m)$ there is no guarantee of an exact solution but x_0 is the least squares solution.