

New method for analyzing dark matter direct detection dataJonathan H. Davis,¹ Torsten Enßlin,² and Céline Boehm^{1,3}¹*Institute for Particle Physics Phenomenology, Durham University, Durham, DH1 3LE, United Kingdom*²*Max-Planck-Institut für Astrophysik Karl-Schwarzschild-Str.,**1. Postfach 13 17 85741 Garching, Germany*³*LAPTH, Université de Savoie, CNRS, BP 110, 74941 Annecy-Le-Vieux, France*

(Received 27 August 2013; published 11 February 2014)

The experimental situation of dark matter direct detection has reached an exciting crossroads, with potential hints of a discovery of dark matter (DM) from the CDMS, CoGeNT, CRESST-II and DAMA experiments in tension with null results from xenon-based experiments such as XENON100 and LUX. Given the present controversial experimental status, it is important that the analytical method used to search for DM in direct detection experiments is both robust and flexible enough to deal with data for which the distinction between signal and background points is difficult, and hence where the choice between setting a limit or defining a discovery region is debatable. In this article we propose a novel (Bayesian) analytical method, which can be applied to all direct detection experiments and which extracts the maximum amount of information from the data. We apply our method to the XENON100 experiment data as a worked example, and show that firstly our exclusion limit at 90% confidence is in agreement with their own for the 225 live days data, but is several times stronger for the 100 live days data. Secondly we find that, due to the two points at low values of S_1 and S_2 in the 225 days data set, our analysis points to either weak consistency with low-mass dark matter or the possible presence of an unknown background. Given the null result from LUX, the latter scenario seems the more plausible.

DOI: [10.1103/PhysRevD.89.043505](https://doi.org/10.1103/PhysRevD.89.043505)

PACS numbers: 95.35.+d

I. INTRODUCTION

Despite convincing gravitational evidence for the existence of dark matter (DM) in our Universe (from galactic to cluster scales) its nature remains a mystery. Yet great progress has been made. In particular direct detection experiments have set progressively stronger limits on the properties of dark matter [1,2], gaining several orders of magnitude in less than a decade for masses in the 10 GeV to TeV range.

Several direct detection experiments have reported dark matter-like events in their data (e.g. CoGeNT [3], CRESST-II [4] and DAMA [5]), with the most recent positive result coming from the CDMS-Si experiment [6]. Such hints are in tension with the limits published by the LUX [7] and XENON100 [8] collaborations. However several authors have claimed that the systematic uncertainties inherent in their analysis may provide a way of reducing such tension [9–11]. In addition if one moves beyond the most basic model of DM-quark scattering and considers e.g. inelastic scattering or isospin-violating DM, where the coupling to neutrons and protons is different, then such tension can also be greatly reduced [12–16].

Given the present situation, it is essential to exploit all the information contained in the data. In this article we propose a Bayesian approach, based on the information Hamiltonian, with a view to providing the community with a novel and robust interpretation of these conflicting experimental signals. This is not the first Bayesian analysis

of direct detection data [17]; however our method is distinct in that it extracts the maximum amount of information from the available data, by exploiting the differences between expected signal and background events. For the purpose of illustration, we will make use of data from the XENON100 calibration [8,18]. This is an independent analysis of XENON100 data, and will enable us to check and also confront our new method with the collaboration's approach. This example is also highly relevant for the LUX experiment, which works under a similar principle.

As we will show for the case where there are signallike points¹ in the data our method is particularly powerful, since one can simultaneously set an exclusion limit and define a potential signal region using Bayesian regions of credibility. This is in contrast to current analytical approaches, which usually involve methods designed only to set limits, such as the p_{\max} method [19], or the profile likelihood analysis with the CL_s method [20]. We do not claim that our method is technically superior for all cases; however our approach is particularly transparent and easily generalized to many different data sets.

In Sec. II we first introduce our method and show how to apply it to direct detection experimental data in general; this

¹We define “signallike” data as those consistent with a signal from DM; however we do not wish to make any explicit claim as to their origin, since they may also be consistent with a background interpretation.

includes a discussion of when to set limits or claim discovery. In Sec. III we apply our method to data from the XENON100 experiment [8,18] as a worked example and conclude in Sec. IV.

II. INFORMATION THEORY

Our method is inspired from information theory, in the sense that it employs Bayesian techniques (see [21] for a review) with the aim to fully exploit the different expected distributions of signal and background events. Ultimately this should either enhance the characteristics of a potential signal (and therefore the evidence for a dark matter discovery), or place stringent bounds on dark matter models.

Before proceeding, we would like to clarify the distinction between this approach, and the profile likelihood method used by e.g. the LUX [7] and XENON100 [8] collaborations to set upper limits on the DM-nucleon cross section (and also by CDMS to fit to their data [22]). The major difference is that our approach is Bayesian and the profile likelihood is frequentist, and hence for example both methods have different ways of dealing with nuisance parameters. However, in most cases, with the same likelihood function the Bayesian and profile likelihood results should agree, and each can provide an important cross-check of the other.

In Sec. III we discuss the XENON100 experiment as a worked example, and when referring to the profile likelihood in this context we mean specifically the likelihood used by the XENON100 Collaboration [8,18] to analyze their data. Indeed, as an alternative, the LUX Collaboration [7] also use a profile likelihood method, but not necessarily the same likelihood function as XENON100.² In the absence of any nuisance parameters, a profile likelihood analysis performed with our likelihood function should give similar limits to those derived in this work using a Bayesian approach. Even so, the two approaches are distinct and should be considered complementary to each other.

A. Dividing the data space into a grid

Our general strategy is to treat any 2D data set effectively as an image, which we pixelate and exploit using pattern recognition. Said differently, we map the data contained in a 2D plot onto a 2D data space Ω . A point x in this space is identified by its two coordinates α and β , the coordinates of the initial plot and in fact the discrimination parameters used to identify events (e.g. scintillation intensity, ionisation, phonon signals).³

²We do not have enough information to make a statement about the likelihood function used by the LUX Collaboration.

³We have chosen a two-dimensional data space here; however our method is easily extended to data with only one parameter or several.

The next key step is to then grid the data space by pixelating it into M two-dimensional bins of equal size in α - β given by $\Delta x_j = (\Delta\alpha, \Delta\beta)$ and labeled with the index j . If such 2D bins are chosen to be small enough, the ability of the analysis to discriminate between signal and background will be maximized. Within a pixel j at position $x_j = (\alpha_j, \beta_j)$ in the α - β plane there will be a certain number n_j of *experimental* data points, each of which are identified by their coordinates x_i^{data} (with i running from 0 to N , the total number of data points in the whole space). For the same pixel, the *theoretically expected* number of points is given by $\lambda_j = \lambda(x_j)\Delta x_j$. Hence we can compare n_j to λ_j given fluctuations in the latter, which we assume obey Poisson statistics. The function $\lambda(x_j)$ is the expected distribution of events, which constitutes the theoretical expectation of both the background and possible signal in a pixel x_j .⁴

B. Defining a likelihood and posterior

We can now analyze the data using the method described above. The main issue is to find for which theoretical parameters λ_j is closest to n_j for all pixels j , within Poisson fluctuations. If there is no DM signal in the data, one expects that for the configuration where λ_j is closest to n_j that the former is equal to the theoretically expected number of background events in each pixel.

For this purpose, we will define a Poisson likelihood to describe the theoretical number of background and signal-like events in each pixel j . Here λ_j represents the mean expectation value of the number of points expected in each pixel j . Such a likelihood is given by

$$\mathcal{P}(d|s) = \prod_{j=1}^M \frac{\lambda_j^{n_j} e^{-\lambda_j}}{n_j!}. \quad (1)$$

In this expression, d represents the data and s the signal.

To make the interpretation easier, we decompose $\lambda(x_j)$ into a DM component $\mathcal{F}(x_j)$ composed entirely of nuclear recoils (NR) and a background component $b(x_j)$ [dominantly electronic recoils (ERs), but with a possible NR component], leading to $\lambda(x_j) = \mathcal{F}(x_j) + b(x_j)$. The predominance of the signal $\mathcal{F}(x_j)$ over the background $b(x_j)$ essentially depends on the number of signal events with respect to that from the background, at a given location in the data space x_j . Since both the number of events and the location are important, and since the location depends on the DM mass (i.e. can be computed once for each mass), we have explicitly separated out these two contributions. Our calculations are therefore significantly speeded up by using the decomposition:

$$\lambda(x_j) = f(x_j)r + b(x_j) \quad (2)$$

⁴The experimental data can be thought of as a discrete sample of the theoretical distribution $\lambda(x)$.

where the term $f(x_j)$ represents the signal position (or shape) in the data space and r its magnitude (or intensity). For the standard picture of a nonrelativistic Weakly Interacting Massive Particle (WIMP), the interaction rate depends linearly on cross section σ , and hence $r \propto \sigma$.

The number of events is governed by the interaction cross section σ between the dark matter and the nucleons of the detector. If the shape of the signal matches that of the data points (above background), then an inspection of the number of events should reveal the value of the cross section, and therefore the strength of the DM interactions.

On the other hand, if the shape does not match the data-point distribution, one can set a limit on the DM interaction cross section. In practice the finite experimental sensitivity means we can only exclude values of σ which would lead to too large a signal. Hence it is convenient to start with a value that is already excluded from previous experimental searches, namely $\sigma = \sigma_0$, and decrease it until one reaches the experimental sensitivity. For this reason we will work with the ratio $r \equiv \sigma/\sigma_0$ where $\sigma_0 \gg \sigma$, so that $r \equiv \sigma/\sigma_0$ provides us with a direct measurement of the intensity of the signal. An exclusion limit is then set by determining the smallest $r = r_{\text{limit}}$ value that still leads to too many signal-like events, so that all $r > r_{\text{limit}}$ are excluded, while keeping values of r which the experiment is not sensitive to.

The number of expected signal events in a pixel at x_j is therefore given by $f_j r = f(x_j) r \Delta x_j$.⁵ To proceed, we must now define a prior for r . We have no theoretical prejudice on its value and therefore consider a flat prior i.e. assign to all possible cross section values $r \in [0, 1]$ the same *a priori* probability density function $\mathcal{P}(s(r)) = \text{const}$.⁶

We can now combine the likelihood $\mathcal{P}(d|s)$ and prior $\mathcal{P}(s)$ into the joint data and signal probability $\mathcal{P}(d, s) = \mathcal{P}(d|s)\mathcal{P}(s)$. We will work with the information Hamiltonian,

$$H = -\ln \mathcal{P}(d, s) = \sum_{\text{pixel}j} (\lambda_j - n_j \ln \lambda_j) + \dots, \quad (3)$$

where the \dots indicates signal-independent terms, which do not contribute to the determination of the ratio r . Inserting our decomposition for $\lambda(x_j)$ [cf. Eq. (2)] and rearranging we obtain

$$H = \sum_{\text{pixel}j} \left[f_j r - n_j \ln \left(1 + \frac{f_j r}{b_j} \right) \right] + \dots \quad (4)$$

⁵We will assume here that the overall normalization for the background is known. However in cases where this is not true one can parametrize the unknown normalization with a nuisance parameter and associated prior, and then marginalize over it.

⁶If we had absolutely no prejudice on the prior value of σ , we would have to take $\sigma_0 \rightarrow \infty$. However in practice we can take σ_0 to be very large but finite, such that we are confident that the probability of finding DM with this interaction strength is vanishingly small, given previous experimental knowledge.

The limit can now be taken where $\Delta x_j \rightarrow 0$, so that each pixel can only contain either 1 or 0 data points. Hence in this limit n_j tends to a delta function and the Hamiltonian becomes

$$H = \int_{\Omega} dx \left[f(x)r - \ln \left(1 + \frac{f(x)r}{b(x)} \right) \delta^N(x - x_i^{\text{data}}) \right] + \dots \quad (5)$$

where the δ function picks out the positions of the N data points x_i^{data} . We define $F = \int_{\Omega} dx f(x)$, the total number of reference signal (nuclear recoil from dark matter) events in the data space calculated at σ_0 .

C. Setting limits and signal regions

With this likelihood we are ready to look for a dark matter signal in our data and we now outline this process explicitly (see also [23]).

As with standard χ^2 methods, we seek to minimize the Hamiltonian. There is a positive identification of a DM signal in the experimental data only when the Hamiltonian possesses a minimum. In this case the shape of the signal $f(x)$ matches the distribution of the data points, in some region of data space where $b(x)$ is expected to be small. The strength of the DM-nucleon interaction is given by the intensity of the signal, r_{best} , corresponding to $\partial_r H(d, s_{\text{best}}) = 0$, with s_{best} representing the properties of the signal that fit the data best.

To define the goodness of the fit in the standard approach, one would then consider all r (or equivalently σ) values leading to $\chi^2 = \chi_{\text{best}}^2 + \delta$ where δ is fixed by the confidence level that one wants to have. Here we shall proceed slightly differently (but ultimately this is equivalent): we define the significance of the signal by integrating the posterior distribution

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d, s)}{\mathcal{P}(d)} \stackrel{\text{f.p.}}{=} \mathcal{P}(d|s), \quad (6)$$

over r , retaining in particular r values around r_{best} .

Note that the last equality holds only for flat priors (f.p.'s), and assuming that $\mathcal{P}(d) = \mathcal{P}(s)$. However, in the following we will take out the normalization of $\mathcal{P}(d|s)$ explicitly, such that

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)}{\int dr \mathcal{P}(d|s)}. \quad (7)$$

Hence in our case a discovery will be established at a confidence level X by using the definition,

$$\int_{r_{\text{low}}}^{r_{\text{best}}} dr \mathcal{P}(s(r)|d) = \int_{r_{\text{best}}}^{r_{\text{up}}} dr \mathcal{P}(s(r)|d) = X/2, \quad (8)$$

where the discovery region is bounded from below by r_{low} and from above by r_{up} . Such a region is therefore a

two-sided region of credibility, while an exclusion limit by contrast is said to be one-sided. One could also relate our two-sided Bayesian region of credibility to a frequentist confidence interval with a certain number of “sigmas,” though this is only strictly possible for a Gaussian likelihood and posterior.⁷

However one may find that the Hamiltonian possesses no minimum. In this case there is no value of r for which the data is compatible with the signal distribution, no matter how intense this distribution becomes. One cannot completely rule out dark matter however, since we know that our experiment has finite sensitivity, but we can set a limit, hereafter referred to as r_{limit} , on the DM interactions.

Since the experiment is not sensitive to DM cross section values smaller than $\sigma_{\text{limit}} = r_{\text{limit}} \times \sigma_0$, all r values below r_{limit} are equally good (or equally bad). Hence there is a region of the parameter space corresponding to $r < r_{\text{limit}}$ where the Posterior probability $\mathcal{P}(s|d)$ is practically constant, as the experiment cannot discriminate between these values of the cross section (for a given exposure).

The allowed region below r_{limit} is thus characterized by $\mathcal{P}(s|d) = cst$ while the excluded region above r_{limit} (where one expects too much signal) is identified by a sharp cutoff in the posterior probability. To determine the exclusion limit (i.e. r_{limit}), we thus seek to quantify this cutoff. We have some freedom in choosing its value: it will depend on the confidence with which we set out limit. For example to set an exclusion limit at a confidence of Y (e.g. for 90% confidence we take $Y = 0.9$), we define r_{limit} analogously with our best-fit region, as

$$\int_0^{r_{\text{limit}}} dr \mathcal{P}(s(r)|d) = Y. \quad (9)$$

By integrating the constant region of the posterior probability until the integration reaches the value that we set, we identify r_{limit} and the cutoff.

Note also that for ease of calculation we tend to use the Hamiltonian in the form of

$$H = Fr - \sum_{i=1}^N \ln(1 + w_i r), \quad (10)$$

where i sums over all N data points at positions x_i and w_i are data weights with $w_i = f(x_i)/b(x_i)$. For setting a limit the first term in Eq. (10) Fr is data independent and gives the absolute limit in the case where no signallike events are observed in the data, while the second term accounts for potential signallike events present in the data, and weakens the limit.

⁷In such a case the size of the region between r_{low} and r_{up} could be directly related to the distance from the best-fit point r_{best} in units of the Gaussian variance i.e. a number of sigmas.

The statistical treatment is largely similar for setting limits or claiming discovery, and our method provides a natural transition between the two, though the approach to how one thinks about regions of credibility is different in either case. Indeed both a signal region and an exclusion limit are equally valid regions of credibility, and so one may wish to highlight both if there is a hint of signal present in the data, but one wishes to remain conservative as to its interpretation.

III. WORKED EXAMPLE: XENON100 DATA

The strongest limits (for $m > 10$ GeV) on the spin-independent cross section for dark matter elastic scattering with nuclei have been set by xenon-based experiments i.e. XENON100 and LUX [7]. We focus on the XENON100 experiment [8,18,24] as a worked example, which operates using both liquid and gaseous xenon with a fiducial mass of 34 kg (for the most recent data set [8]). The XENON100 detector identifies events by using two distinct signals [25]: primary (S1) and secondary (S2) scintillation, the former of which is due to scintillation light originating from the liquid part of the detector, while the latter comes from ionized electrons, which drift to the gaseous part of the detector under an electric field. The LUX detector [7] operates on a similar principle, but with a larger fiducial mass. The LUX Collaboration also employs different cuts (e.g. a cut at $S1 = 2$ PE, instead of 3 PE) and potentially a different likelihood function for their own analysis. Otherwise, the following discussion should be interesting for an understanding of the analysis of LUX data, as well as XENON100.

In order to derive limits on the spin-independent cross section as a function of dark matter mass, the XENON100 Collaboration employs a profile likelihood approach [20,26]. Such a method takes advantage of the distinct signatures in S1-S2 of electronic and nuclear recoils by splitting the data space into a number of bands (23 in [20] and 12 in [26]). We can contrast this approach with our method, where the data space is split into a grid of rectangular pixels, which are associated with a point in the data space $x = (\alpha, \beta)$. Hence, we expect our gridded approach to perform better than this method of bands used by the XENON100 Collaboration, since we can exploit the difference between signal and background to the maximum amount, while they are limited by the rather coarse-grained resolution of their bands.⁸ This application should serve as a clear demonstration of the advantages to *any* direct detection experiment of using our method.

⁸The likelihood used by the XENON100 Collaboration is claimed to be able to exploit the spectra of events within each band using a separate term in the likelihood [20,26]. However in practice it is not clear how effective this actually is, and we do not believe it exploits the difference between signal and background as well as our gridded method.

We can identify $S1$ and $S2$ with our discrimination parameters α and β from Sec. II A, though here we choose instead to take $\alpha = S1$, $\beta = \text{Log}(S2/S1)$, to match more closely the method used by the XENON100 Collaboration themselves (and also the LUX Collaboration [7]). We will proceed first to discuss the determination of the signal $f(x)$ and background $b(x)$ distributions for the XENON100 experiment, before applying our method to data, both the more recent 225 live days data (225LD) [8] and the older 100 live days data set (100LD) [18].

A. Signal distribution

1. WIMP recoil spectrum

Potential WIMP events are characterized by their recoil spectra $\frac{dR}{dE}$, parametrized as [27,28],

$$\frac{dR}{dE} = \frac{\sigma(E)}{2m\mu^2} \rho \eta(E), \quad (11)$$

where $\sigma(E)$ is the WIMP-nucleus cross section as a function of energy E , μ is the WIMP-nucleus reduced mass, $\rho = 0.3 \text{ GeV cm}^{-3}$ is the local dark matter density and $\eta(E) = \int_{v_{\min}}^{\infty} d^3v \frac{f(v+u_e)}{v}$ is the WIMP mean velocity. The mean velocity is integrated over the distribution of WIMP velocities in the galaxy $f(v)$ boosted into the reference frame of the Earth by u_e . The lower limit of the integration is $v_{\min}(E)$, which is the minimum WIMP velocity required to induce a recoil of energy E . We assume the standard halo model such that $f(v)$ is given by a Maxwell-Boltzmann distribution cut off at an escape velocity of $v_{\text{esc}} = 544 \text{ kms}^{-1}$.

We assume that WIMPs interact identically with protons and neutrons⁹ giving $\sigma(E) = \sigma(\frac{\mu}{\mu_p})^2 A^2 F^2(E)$, where σ is the zero-momentum WIMP-nucleon cross section, A is the atomic mass of xenon, μ_p is the WIMP-proton reduced mass and $F(E)$ is the Helm nuclear form factor [27]. For 225LD (100LD) we use a value of 224.6 days (100.9 days) for the exposure and 34 kg (48 kg) for the target mass.

2. Calculation of $S1$ and $S2$ for nuclear recoils

At a given nuclear-recoil energy E the expected primary ($S1_{\text{exp}}$) and secondary ($S2_{\text{exp}}$) scintillation signals are obtained from the following formulas [10,26,31,32]:

$$S1_{\text{exp}} = P \left(E \cdot L_{\text{eff}}(E) \cdot L_y \cdot \frac{S_n}{S_e} \right) \quad (12)$$

⁹We take a simple model of DM-nucleon elastic scattering here for convenience; however our method is easily generalized to more complicated models (e.g. [14–16,29,30]) by replacing $\sigma(E)$ and dR/dE .

$$S2_{\text{exp}} = Y \cdot P(E \cdot Q_y(E)), \quad (13)$$

where $P(x)$ represents a Poisson distribution with expectation value x , $L_y = 2.20 \pm 0.09 \text{ PE keV}^{-1}$, $\frac{S_n}{S_e} = 0.95/0.58$, Y is a Gaussian-distributed value with mean 19.5 PE per electron and width $\sigma = 6.7 \text{ PE}/e^-$ [33], $L_{\text{eff}}(E)$ is the relative scintillation efficiency and $Q_y(E)$ is the ionization yield. For Q_y there is a degree of uncertainty on its functional form [33]; we use the model of [32] in this work; however we have obtained similar results with the best-fit curve from [33]. L_{eff} is obtained from a cubic spline fit to data from [34–37].

To obtain the $S1_{\text{obs}}$ and $S2_{\text{obs}}$ signals observed in the detector, we must include the finite detector resolution and the cuts imposed by the XENON100 Collaboration on the data [18,26,33]. Both $S1_{\text{exp}}$ and $S2_{\text{exp}}$ are blurred with a Gaussian of width $0.5\sqrt{n}$ for n photoelectrons (PEs) to take account of the finite photomultiplier (PMT) resolution [20]. The effect of cuts is then implemented using the cut-acceptance curve as a function of $S1$ [26,33] after applying the resolution effect. Additionally an $S2$ threshold cut is applied before Gaussian blurring, cutting away all points with $S1 < 1 \text{ PE}$ [8].

3. Expected dark matter signal in XENON100

The expected signal distribution for a given WIMP mass in the data space $f(x)$ can now be calculated using dR/dE of Sec. III A 1, at a value of the reference cross section $\sigma_0 = 10^{-35} \text{ cm}^2$ (or 10^{-34} cm^2 for $m < 10 \text{ GeV}$). The energy range between 1 keV and 60 keV is separated into bins of size $\Delta E = 0.01 \text{ keV}$. For each binned energy E_{rec} we calculate $S1_{\text{obs}}$ and $S2_{\text{obs}}$ a total of N_{rec} times, where $N_{\text{rec}} = \frac{dR}{dE}(\sigma_0, E_{\text{rec}}) \Delta E$, to obtain the full signal distribution as expected in XENON100. The result is shown for two different masses in Fig. 1. Similar simulations of the signal distribution expected from XENON100 have been performed in [11,33,38]; however our method goes further and directly links these to the analysis through the weight function $w(x) = f(x)/b(x)$, as shown in Fig. 1.

B. Background distribution in XENON100

The expected distribution of electronic-recoil background events $b_{\text{ER}}(x)$ is determined from fits to ^{60}Co calibration data,¹⁰ as is done in [18,20]. Although the electronic recoil events appear mostly Gaussian distributed, the XENON100 Collaboration noticed the presence of an

¹⁰Our determination of $b_{\text{ER}}(x)$ would improve were we to use the ^{232}Th calibration data (especially for the anomalous component), collected by the XENON100 Collaboration for their most recent analysis [8]; however this is not currently publicly available.

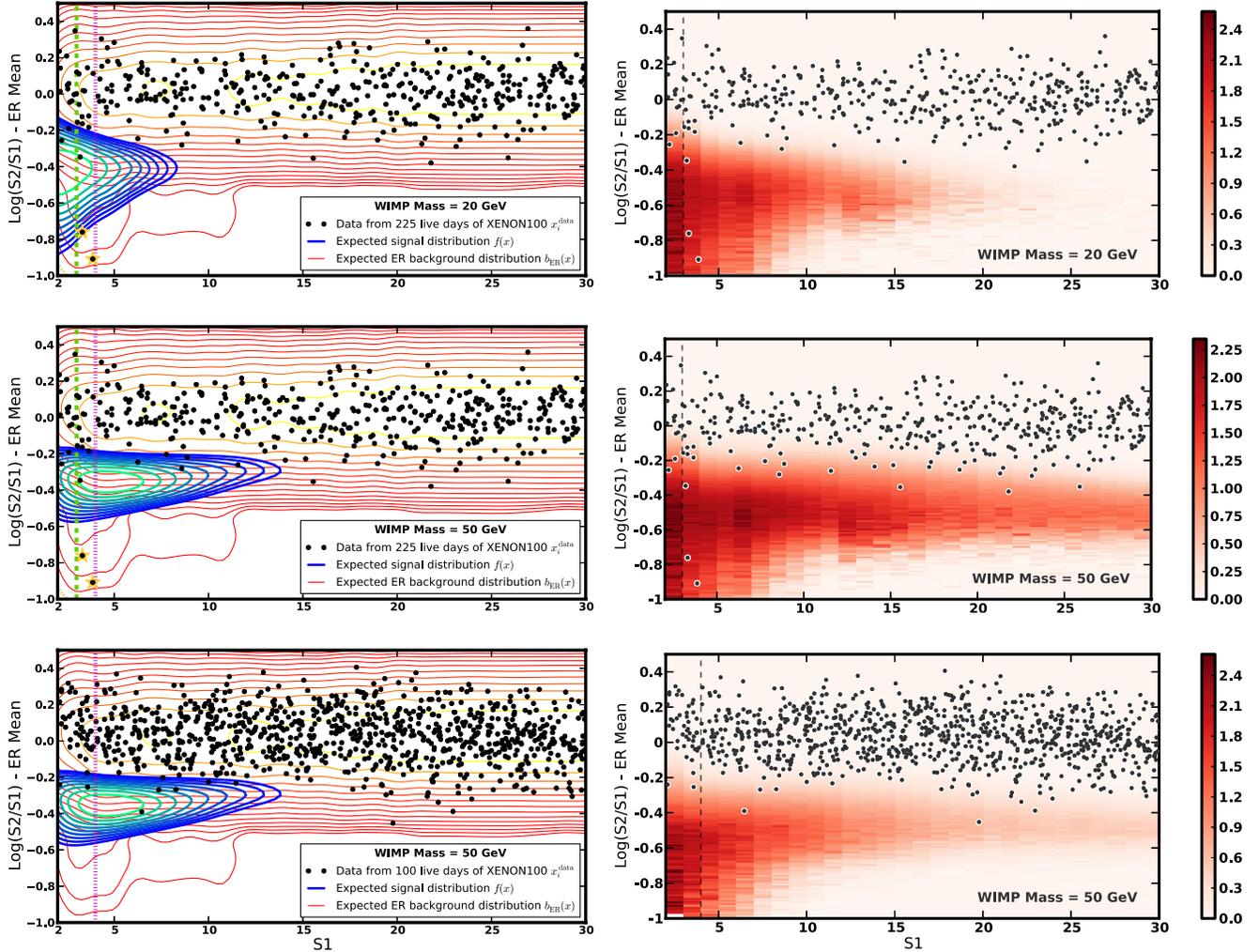


FIG. 1 (color online). The upper four panels show the 225 live days data set [8], while the lower two display the data for 100 live days of the XENON100 experiment [18]. The left panels show the expected signal $f(x)$ and background $b(x)$ distributions used for our analysis. For the signal distribution, each contour is 1.2 times less than the previous, from light to dark blue, while for the background the ratio is 1.5 from orange to red. The data are shown as black circles. Note that only the ER background is shown here for convenience, where one can also see the anomalous background component at low $S1$, as discussed in Sec. III B. For the 225 live days data, the two most signallike points have been highlighted with yellow stars and are referred to as “hint” points in the text. In the right panels we show the function $\text{Ln}(1 + w(x) \cdot s)$, where $s = 10^{-8}$ here and $w(x) = f(x)/b(x)$, the weight distribution of Eq. (10). We bin $w(x)$ in units of $\Delta S1 = 0.5$ and $\Delta \text{Log}(S2/S1) = 0.01$, and interpolate between these bins for the analysis. The y axis is shifted by the mean of the electronic-recoil band, as shown by “ER Mean.” For the 225 live days data we make use of two sets of cuts on the data space: the first is to consider only points between $S1_{\text{up}} = 3$ PE and $S1_{\text{up}} = 30$ PE, while the second moves the lower cut to $S1_{\text{low}} = 4$ PE. The former is referred to as the full data set in the text, while the latter removes the hint points and is referred to as the hint-removed data set. For the 100 live days data $S1_{\text{low}} = 4$ PE and $S1_{\text{up}} = 30$ PE. Additionally the 225 live days data is bounded from below by $S2 = 150$ PE, while for 100 live days this moves to 300 PE.

anomalous (non-Gaussian) background component [18]. This could be due to double-scatter gamma events, where only one of the gammas contributes to the $S2$ signal. Both such components of the ER background are included; indeed the anomalous component can be seen in Fig. 1 predominantly at low $S1$. The distribution is normalized by the total number of expected background events, whose rate takes the constant value of 0.0053 counts per day per kg per keV_{ee} [8,39]. For 100LD the background is

larger due to krypton contamination in the experimental apparatus, taking a value of 0.022 counts per day per kg per keV_{ee} [18].

We also model the nuclear-recoil background due to neutrons $b_{\text{NR}}(x)$. The distribution is calculated as for the signal distribution, but replacing dR/dE with the expected energy spectrum of neutron scatters in the detector [40]. Hence the total background distribution is $b(x) = b_{\text{ER}}(x) + b_{\text{NR}}(x)$.

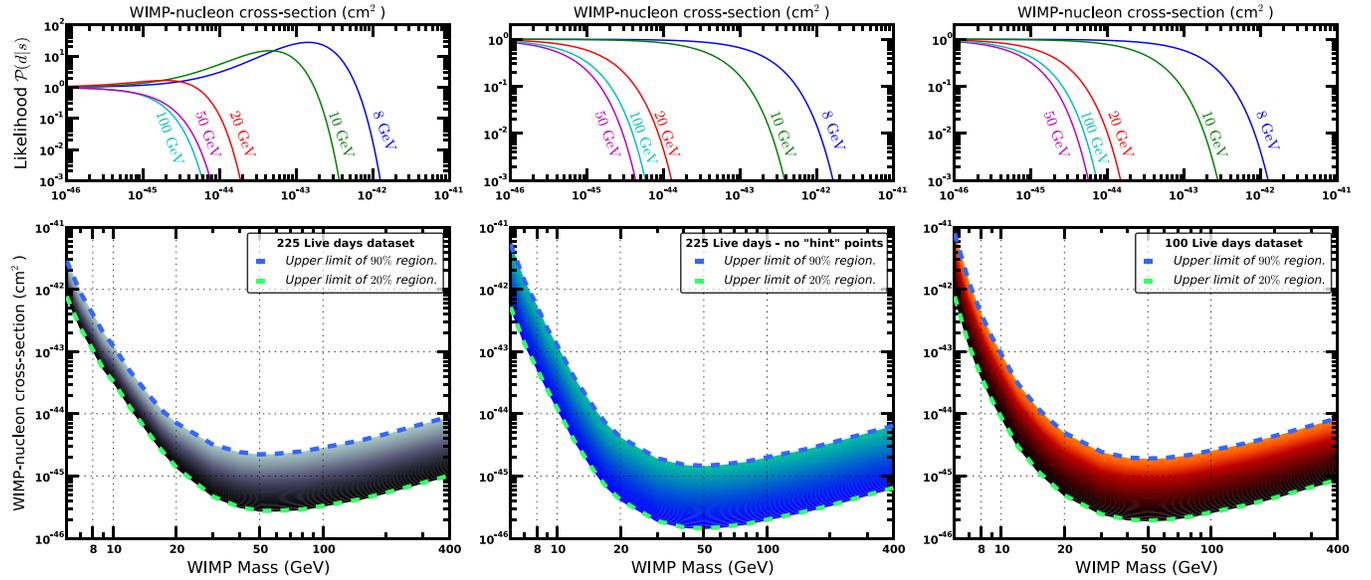


FIG. 2 (color online). Plots showing exclusion limits and regions of credibility, derived from applying our analysis to data from the XENON100 experiment [8]. For the leftmost 225LD analysis, there is a weak preference for low-mass DM, which vanishes under more stringent cuts (central) or with the 100LD data (right). The upper panels show examples of the (un-normalized) likelihood function $\mathcal{P}(d|s)$ for various WIMP masses, while the lower panels show the result of integrating the posterior from $s = 0$ up to some limiting value, in order to define an exclusion limit for a given significance. The region between the two dashed lines shows exclusion curves with significance increasing linearly from darker to lighter shading. One can indeed consider this region as one of 70% significance. For the left panels we have used the full 225LD data set (all points between $S1_{\text{low}} = 3$ PE and $S1_{\text{up}} = 30$ PE), while for the central panels the analysis has been performed with the two most signallike (labeled as “hint”) data points removed by cutting off the data space below $S1_{\text{low}} = 4$ PE. The rightmost panels show results for the 100LD data.

C. Example exclusion limits and posterior scans

1. Signals and limits from XENON100 data

Now that we know how to calculate the expected signal and background distributions $f(x)$ and $b(x)$, we are ready to apply our method to the data from the XENON100 experiment. All relevant ingredients are displayed in Fig. 1; the left panels show the regions where the expected signal and background are expected to be largest, while the right panels show plots of $\text{Ln}(1 + w(x) \cdot s)$ as used directly for our analysis. The discrimination between signal and background is maximized provided the two-dimensional bins for $w(x) = f(x)/b(x)$ are small enough: data points where $w(x)$ is large are more likely to be due to signal than background, while the opposite is true for points located where $w(x)$ is small. This is then fed directly into our analysis; hence Fig. 1 contains all of the main ingredients of our method.

Shown in Fig. 2 are the results of applying our method to the data. In order to understand the effect of data points consistent with a signal interpretation, we have performed the analysis with both the full data set (with a lower cut on $S1$ at $S1_{\text{low}} = 3$ PE), and with a reduced data set, where the two hint data points (i.e. the starred points in Fig. 1) have been removed by cutting away the data space below

$S1_{\text{low}} = 4$ PE.¹¹ The former is displayed in the left panel of Fig. 2, while the results for the reduced data set are shown on the central panel. Results from the 100LD data are shown on the right.

As discussed in Sec. II C we can define regions of credibility (either exclusion limits or potential discovery regions) by integrating under the normalized posterior $\mathcal{P}(s|d)$. Hence in the lower panels of Fig. 2 we show exclusion limits for various levels of confidence, between 20% and 90%, calculated by integrating the posterior from $s = 0$ up to the limiting value of s . One can equivalently consider the parameter space between these limits as a region of 70% credibility. The 90% limit for the full 225LD data set can be compared with the result from [8], while the shaded band represents how the limit changes with different confidence.

The upper panels show the dependence of the likelihood $\mathcal{P}(d|s)$ as a function of σ for various WIMP masses. One can see directly that for the full 225LD data set the likelihood function has a maximum (corresponding to a minimum in the Hamiltonian), indicating a preference for

¹¹We could instead have moved the low- $S2$ cut from 150 to 300 PE, as for the 100LD data set, which would remove one of these points.

the data of a particular value of σ , which is strongest for lighter WIMPs. Indeed this can also be observed in the exclusion curve as we change the significance value: particularly for lighter WIMPs the region of credibility between the 20% and 90% limits is denser as compared to heavier DM. This is due directly to the presence of a maximum in the posterior and likelihood.

This is particularly interesting in the context of the potential hints of light DM in CDMS [6] and CoGeNT [3] (and to some extent DAMA). However the significance of such a hint is weak. Indeed the 50% credible region for an 8 GeV WIMP lies between 6.15×10^{-44} and 2.15×10^{-43} cm², with a best-fit cross section at 1.40×10^{-43} cm². Of course the cross section is still inconsistent with the best-fit region from CDMS [6], unless one changes the systematic parameters to a rather extreme degree [11] or considers less standard interactions [12].

Claims that these points are consistent with a DM signal are likely to be overly optimistic. The significance of the signal is comparable to a 1σ fluctuation,¹² and hence these data points may just be events from the non-Gaussian ER background, which we already model. We can additionally compute the Bayes factor \mathcal{B} [41] for e.g. an 8 GeV WIMP, by calculating the ratio of the joint signal and data probability $\mathcal{P}(d, s)$ integrated over all r , to $\mathcal{P}(d, s(r=0))$ i.e. the no-DM scenario, where $\sigma = 0$. Hence the size of \mathcal{B} should tell us to what degree a positive signal of DM is preferred, relative to the scenario where no signal is present (see [41] for details). We calculate $\mathcal{B} = 3.18$, which is just on the boundary of being a positive result. Hence, again we can conclude there is only a weak hint of signal for a low-mass WIMP. There are also systematic uncertainties from L_{eff} and Q_y , though they are unlikely to result in a significant enhancement of the signal significance.

Indeed, as can be seen from 1 if one attributes these points to a WIMP signal, one must also explain why no data is seen where the signal from DM is expected to be even larger, at lower values of S1 for example. Even so, the presence of consistency with signal, however weak, indicates some sort of new phenomenon may be present: either DM or an unknown (or possibly misunderstood) background. Hence an interpretation of these points in terms of dark matter is possible but premature; however they are instructive as an example of the effect of signallike points on our ability to set limits on light DM.

¹²Since our method is Bayesian, a comparison with frequentist confidence intervals is not directly possible. However if one considers a 1σ confidence interval as (roughly) comparable to a 68% region of credibility, then we actually find the significance to be a bit less than 1σ . Indeed our choice of 50% was motivated by the fact that it is close to the largest two-sided interval we could set around the maximum-likelihood value of the cross section. The sigma level is only approximate though, as our likelihood is non-Gaussian (see Fig. 2).

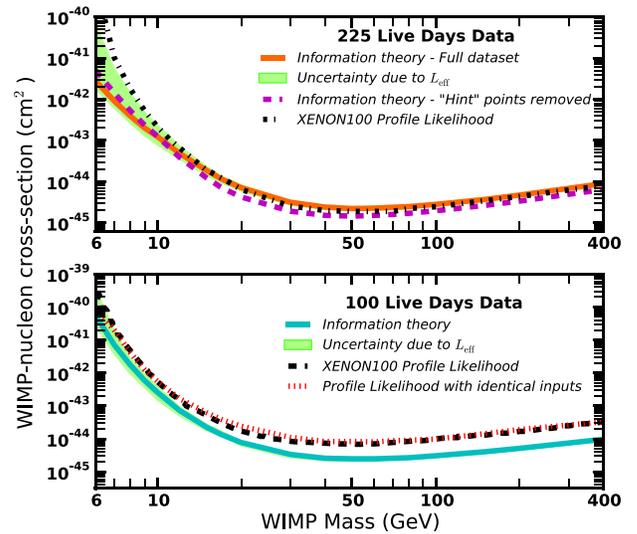


FIG. 3 (color online). A comparison of various limits set with either 225 live days [8] or 100 live days [18] of XENON100 data. Limits from information theory refer to those derived using the method presented in this work. For the 100 live days data we also compare the result of a profile likelihood analysis performed by the XENON100 Collaboration with that from an analysis we have done using the same profile likelihood method, but where the inputs are identical to those for our Bayesian method, such as $f_{\text{NR}}(x)$ and $b(x)$. The limit from our Bayesian information theory method agrees with the XENON100 published limit for 225LD, but is several times stronger for 100LD.

By contrast when the two hint data points are removed from the analysis by the more stringent low-S1 cut (see Fig. 1 for details), there is no maximum in the likelihood and posterior for any WIMP mass, as one would expect since all points are in a region where the weight $w(x) = f(x)/b(x)$ is small. Indeed the density of the posterior is now less for all masses than for the full data set, with the contrast particularly stark for lighter DM. The same is seen for 100LD, for which no hint of signal is present. In addition, the limits without the hint points are stronger since the data are now almost completely consistent with a negative result. If the XENON100 Collaboration were to observe additional signallike points in their data, one would expect the density of the posterior to increase around the best-fit region.

In any case this demonstrates the ability of our method to accurately set limits or define potential discovery regions. All of the relevant information is contained within the posterior $\mathcal{P}(s|d)$, which can be integrated over to define the degree of belief that a given region of parameter space is consistent with the data.

2. Comparison with results from XENON100

Before forming any firm conclusions on the efficacy of our method in searching for dark matter signals in direct detection data, we must compare our results to those

previously found by the XENON100 Collaboration. Shown in Fig. 3 is our 90% confidence limit (identical to the one in Fig. 2), compared with the limit derived by the XENON100 Collaboration with the same 225 live days data set [8], but their own profile likelihood analysis [20]. Uncertainties due to the relative scintillation efficiency L_{eff} are shown as a shaded region around our limit (see e.g. [9,10] for a review).

In addition, in the lower panel of Fig. 3 we also show the results of applying our method to the 100 live days data set, along with the limit from the XENON100 Collaboration using their profile likelihood method, and a limit we have independently derived using the same method, but with identical inputs to our information theory analysis.

The exclusion limit derived with our information Hamiltonian method agrees with that derived by the XENON100 Collaboration for the 225 live days data set for large masses. For lighter WIMPs our limit is stronger, though this is likely due to uncertainty in the low-energy extrapolation of L_{eff} [9]. Indeed the XENON100 Collaboration employs the most conservative approach and cuts L_{eff} to zero below $3 \text{ keV}_{\text{nr}}$, where no data are available [8]. Our limit is derived using a constant extrapolation instead, though the uncertainty band shows the limit under different parametrizations of L_{eff} [9]. Hence one can consider our result as an independent cross-check of the limit published by the XENON100 Collaboration.

There are undoubtedly other small differences between our inputs and those used by the XENON100 Collaboration; however the agreement of both limits indicates that our method does indeed perform correctly when analyzing direct detection data. Note also that for the hint-removed data set, where the low-S1 cut is moved to $S1_{\text{low}} = 4 \text{ PE}$, the limit is stronger for heavy WIMPs due to the removal of the signallike points by the cut. This is not so for lighter WIMPs, since much of the region where one expects to see signal is cut away in addition to the hint points.

We note however that when applying our method to the 100LD data [18] that our information theory limit is stronger than that derived using the profile likelihood analysis, both performed directly by the XENON100 Collaboration and from an independent analysis we have carried out. Since the latter two limits are in agreement, it would be difficult to blame the inputs of the analysis on this discrepancy between the limits; hence it is likely that the coarse-graining¹³ of the profile likelihood analysis has resulted in the derivation of an overconservative limit. To reiterate, we refer specifically to the profile likelihood analysis used by XENON100 here. The issue is not with the frequentist method itself, but rather with the choice of

likelihood function used by the collaboration. Hence, our limit is more accurate because we use a likelihood which exploits the whole data space, and this should also be reflected in a profile likelihood analysis which followed the same principles.

The reason for this discrepancy arising only for the 100LD data set is not entirely clear, though it is likely that the increased background in this data set relative to that from 225 live days [8] (due to the krypton leakage) has effectively fooled the analysis into treating too many points as potential signal, thereby weakening the limit. Hence we believe that this demonstrates the robustness of our method as compared to such a profile likelihood analysis, since it is less susceptible to leakage of background points into the signal region.

IV. CONCLUSION

In this work we have introduced a Bayesian method of analyzing data from dark matter direct detection experiments. Our method takes as input the data set itself and the expected signal and background distributions, defined over the whole data space, which is divided into a grid of two-dimensional pixels. This enables us to take full advantage of the distinct expected distributions signal and background events, and hence to set limits (or discovery regions) without resorting to conservative approximations.

Using data from the XENON100 experiment [8] as a worked example we demonstrated how one would apply our method to direct detection data. This has direct relevance also to the LUX experiment [7], and any future runs of XENON100. We have shown that there is merit in looking beyond the 90% confidence limit, as hints of signal may be affecting the structure of the likelihood and posterior in a nontrivial manner. Indeed an analysis of the XENON100 data from 225 live days indicates a weak preference in the data for a light DM particle. At 50% confidence the best-fit cross section is in between 6.15×10^{-44} and $2.15 \times 10^{-43} \text{ cm}^2$ for an 8 GeV WIMP; the error bars being relatively large, it is very premature to argue that this is evidence for dark matter. Similar regions can be obtained for any dark matter particle with a mass below $\sim 20 \text{ GeV}$, with a possible evidence for a dark matter signal in the data vanishing for masses above about 20 GeV. If indeed these points are due to a detection of dark matter, more data from the XENON100 experiment should increase the confidence level and shrink the error bars on the cross section. Alternatively, these events may be found to be due to an additional background process or the anomalous component of the ER background, in which case the signal significance would vanish with more data. Considering the recent null result from the LUX experiment [7], the latter would seem to be a more plausible explanation.

¹³Specifically we refer to the splitting of the data space into a finite number of bands for the profile likelihood method used by the XENON100 Collaboration, which necessarily limits the amount of information extracted from the data, as opposed to our method where the data space is pixelated (see Fig. 1).

We also demonstrated that our new method can produce a complementary analysis to the one currently used by the XENON100 Collaboration, where the data are placed into bands. Indeed our limit and theirs agree for the most recent 225 live days data set [8]; however ours is several times stronger for the data from 100 live days [18]. The reason for this disagreement for the older data set is not clear. However it is possible that since the background was higher due to krypton contamination, there was a greater proportion of background events leaking into the region where signal was expected (i.e. the more signallike bands of the analysis used by the XENON100 Collaboration), which may have fooled their analysis into setting too weak a limit. Additionally our method could be even more robust, especially if one exploits the full detector volume (with f and b now depending on physical positions in the detector).

Our analysis can be seen as an independent analysis of the XENON100 data, and more importantly could be employed by any present or forthcoming experimental collaboration for such a purpose. In particular, our method can be easily applied to the LUX experiment [7], since it

operates on a similar principle to XENON100. In this case one should hope to find agreement with our Bayesian results and the frequentist method used by the LUX Collaboration, which should provide an important cross-check of the LUX results. Future experiments such as XENON1T [42], LZ [43] and SuperCDMS [44] could also benefit from a Bayesian cross-check.

The use of our formalism should be very convenient to set limits and potential regions of discovery simultaneously, allowing scenarios where the presence of a signal is ambiguous to be studied without bias. Additionally, our method can be used to go beyond the conservative approach, and to set the strongest limit possible by exploiting the different distributions of signal and background events. With a consistent analytical method used by all dark matter direct detection experiments, the current constraints on the WIMP cross section should be both stronger and clearer.

ACKNOWLEDGMENTS

J. H. D. and C. B. are supported by the STFC.

-
- [1] Z. Ahmed *et al.* (CDMS Collaboration, EDELWEISS Collaboration), *Phys. Rev. D* **84**, 011102 (2011).
 - [2] D. Y. Akimov *et al.*, *Phys. Lett. B* **709**, 14 (2012).
 - [3] C. Aalseth *et al.*, *Phys. Rev. Lett.* **107**, 141301 (2011).
 - [4] G. Angloher *et al.*, *Eur. Phys. J. C* **72**, 1971 (2012).
 - [5] R. Bernabei *et al.* (DAMA Collaboration, LIBRA Collaboration), *Eur. Phys. J. C* **67**, 39 (2010).
 - [6] R. Agnese *et al.* (CDMS Collaboration), *Phys. Rev. Lett.* **111**, 251301 (2013).
 - [7] D. Akerib *et al.* (LUX Collaboration), [arXiv:1310.8214](https://arxiv.org/abs/1310.8214) [*Phys. Rev. Lett.* (to be published)].
 - [8] E. Aprile *et al.* (XENON100 Collaboration), *Phys. Rev. Lett.* **109**, 181301 (2012).
 - [9] J. H. Davis, C. Boehm, N. Oppermann, T. Ensslin, and T. Lacroix, *Phys. Rev. D* **86**, 015027 (2012).
 - [10] C. Savage, G. Gelmini, P. Gondolo, and K. Freese, *Phys. Rev. D* **83**, 055002 (2011).
 - [11] D. Hooper, *J. Cosmol. Astropart. Phys.* **09** (2013) 035.
 - [12] M. T. Frandsen, F. Kahlhoefer, C. McCabe, S. Sarkar, and K. Schmidt-Hoberg, *J. Cosmol. Astropart. Phys.* **07** (2013) 023.
 - [13] M. T. Frandsen, F. Kahlhoefer, S. Sarkar, and K. Schmidt-Hoberg, *J. High Energy Phys.* **09** (2011) 128.
 - [14] T. Schwetz and J. Zupan, *J. Cosmol. Astropart. Phys.* **08** (2011) 008.
 - [15] S. Chang, J. Liu, A. Pierce, N. Weiner, and I. Yavin, *J. Cosmol. Astropart. Phys.* **08** (2010) 018.
 - [16] D. Hooper and C. Kelso, *Phys. Rev. D* **84**, 083001 (2011).
 - [17] C. Arina, J. Hamann, and Y. Y. Wong, *J. Cosmol. Astropart. Phys.* **09** (2011) 022.
 - [18] E. Aprile *et al.* (XENON100 Collaboration), *Phys. Rev. Lett.* **107**, 131302 (2011).
 - [19] S. Yellin, *Phys. Rev. D* **66**, 032005 (2002).
 - [20] E. Aprile *et al.* (XENON100 Collaboration), *Phys. Rev. D* **84**, 052003 (2011).
 - [21] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, England, 2003).
 - [22] J. Billard, [arXiv:1312.7737](https://arxiv.org/abs/1312.7737).
 - [23] R. D. Cousins, *Am. J. Phys.* **63**, 398 (1995).
 - [24] E. Aprile *et al.* (XENON100 Collaboration), *Phys. Rev. Lett.* **105**, 131302 (2010).
 - [25] E. Aprile *et al.*, *Astropart. Phys.* **35**, 573 (2012).
 - [26] E. Aprile *et al.*, [arXiv:1207.3458](https://arxiv.org/abs/1207.3458).
 - [27] J. Lewin and P. Smith, *Astropart. Phys.* **6**, 87 (1996).
 - [28] D. G. Cerdeno and A. M. Green, [arXiv:1002.1912](https://arxiv.org/abs/1002.1912).
 - [29] E. Del Nobile, C. Kouvaris, P. Panci, F. Sannino, and J. Virkajarvi, *J. Cosmol. Astropart. Phys.* **08** (2012) 010.
 - [30] R. Foot, *Phys. Lett. B* **728**, 45 (2014).
 - [31] P. Sorensen, *J. Cosmol. Astropart. Phys.* **09** (2010) 033.
 - [32] F. Bezrukov, F. Kahlhoefer, and M. Lindner, *Astropart. Phys.* **35**, 119 (2011).
 - [33] E. Aprile *et al.* (XENON100 Collaboration), *Phys. Rev. D* **88**, 012006 (2013).
 - [34] V. Chepel, V. Solovov, F. Neves, A. Pereira, P. J. Mendes, C. P. Silva, A. Lindote, J. Pinto da Cunha, M. I. Lopes, and S. Kossionides, *Astropart. Phys.* **26**, 58 (2006).

- [35] A. Manzur, A. Curioni, L. Kastens, D. N. McKinsey, K. Ni, and T. Wongjirad, *Phys. Rev. C* **81**, 025808 (2010).
- [36] G. Plante, E. Aprile, R. Budnik, B. Choi, K.-L. Giboni, L. W. Goetzke, R. F. Lang, K. E. Lim, and A. J. Melgarejo Fernandez, *Phys. Rev. C* **84**, 045805 (2011).
- [37] E. Aprile, L. Baudis, B. Choi, K. L. Giboni, K. Lim, A. Manalaysay, M. E. Monzani, G. Plante, R. Santorelli, and M. Yamashita, *Phys. Rev. C* **79**, 045807 (2009).
- [38] P. Sorensen, *Phys. Rev. D* **86**, 101301 (2012).
- [39] E. Aprile *et al.*, *Phys. Rev. D* **83**, 082001 (2011).
- [40] E. Aprile *et al.* (XENON100 Collaboration), *J. Phys. G* **40**, 115201 (2013).
- [41] R. E. Kass and A. E. Raftery, *J. Am. Stat. Assoc.* **90**, 773 (1995).
- [42] E. Aprile *et al.* (XENON1T Collaboration), [arXiv:1206.6288](https://arxiv.org/abs/1206.6288).
- [43] D. Malling *et al.*, [arXiv:1110.0103](https://arxiv.org/abs/1110.0103).
- [44] R. Agnese *et al.* (SuperCDMS Collaboration), [arXiv:1309.3259](https://arxiv.org/abs/1309.3259) [*Phys. Rev. Lett.* (to be published)].