Physics of the Dark Universe 37 (2022) 101118

Contents lists available at ScienceDirect

Physics of the Dark Universe

journal homepage: www.elsevier.com/locate/dark

A machine learning algorithm for direct detection of axion-like particle domain walls

Dongok Kim [a,b], Derek F. Jackson Kimball [c], Hector Masia-Roig [d,e], Joseph A. Smiga [d,e], Arne Wickenbrock [e,d], Dmitry Budker [e,d,f], Younggeun Kim [a,b], Yun Chang Shin [b,*], Yannis K. Semertzidis [b,a]

[a] Department of Physics, Korea Advanced Institute of Science and Technology, 34141, Republic of Korea
[b] Center for Axion and Precision Physics Research, Institute for Basic Science, 34051, Republic of Korea
[c] Department of Physics, California State University – East Bay, Hayward, CA 94542-3084, USA
[d] Johannes Gutenberg-Universität Mainz, 55128 Mainz, Germany
[e] Helmholtz Institut Mainz, Johannes Gutenberg-Universität, 55099 Mainz, Germany
[f] Department of Physics, University of California, Berkeley, CA 94720-7300, USA

ARTICLE INFO

Article history:
Received 7 October 2021
Received in revised form 12 September 2022
Accepted 19 September 2022

Keywords:
Dark matter
Axion
Machine learning
Optical magnetometer
Localized dark matter

ABSTRACT

The Global Network of Optical Magnetometers for Exotic physics searches (GNOME) conducts an experimental search for certain forms of dark matter based on their spatiotemporal signatures imprinted on a global array of synchronized atomic magnetometers. The experiment described here looks for a gradient coupling of axion-like particles (ALPs) with proton spins as a signature of locally dense dark matter objects such as domain walls. In this work, stochastic optimization with machine learning is proposed for use in a search for ALP domain walls based on GNOME data. The validity and reliability of this method were verified using binary classification. The projected sensitivity of this new analysis method for ALP domain-wall crossing events is presented.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Even though there is a considerable amount of evidence for the existence of dark matter, the nature of dark matter is not fully understood [1]. Up to now, there have been a number of hypotheses proposed to explain the existence of dark matter [2–10]. One of the most well-motivated dark matter candidates is an axion, which was originally proposed to solve the strong CP problem in quantum chromodynamics (QCD) [11,12]. These QCD axions are weakly-coupled light pseudo-scalar particles generated from the spontaneously broken Peccei–Quinn $U(1)_{PQ}$ symmetry [13–18]. This concept can be generalized to a class of light pseudo scalar particles which are collectively referred to as axion-like particles (ALPs). They are motivated by a spontaneously broken global $U(1)$ symmetry beyond the Standard Model (SM), such as those appearing in string theory [19–21]. The generalized ALP field a can include non-gravitational couplings arising from the following interaction Lagrangian

$$\mathcal{L}_{\text{int}} = \frac{\partial_\mu a}{f_{\text{lin}}}\bar{\psi}\gamma^\mu\gamma^5\psi + \frac{\partial_\mu a^2}{f_{\text{quad}}^2}\bar{\psi}\gamma^\mu\gamma^5\psi + \cdots, \tag{1}$$

where f_{lin} and f_{quad} are the effective linear and quadratic interaction scales in energy units, ψ is a fermion field in SM [22–24]. This allows ALPs to couple to atomic spins through a gradient interaction.

In most direct searches for dark matter, the density of dark matter in the solar system is assumed to be relatively uniform [25]. However, in addition to this conventional model of dark matter distribution, it is possible that the local dark matter density is highly nonuniform. This can occur as a result of the formation process of pseudo-scalar fields during cosmological inflation [26,27]. An example is the Kibble mechanism [28] which describes a cosmological phase transition during the cooling down of the early Universe. The phase transitions associated with symmetry breaking might induce local selections of broken symmetry and eventually result in separated domains with locally degenerate broken symmetry. Then, this can naturally lead to topological defects if the separation between domains are too far to communicate. The type of defect mainly depends on the property of the broken symmetry and the characteristics of the phase transition but can be classified based on their dimensionality: monopoles in 0D, strings in 1D, and domain walls (DWs) in 2D or higher dimensions. Among them, the DWs are objects formed from the discrete broken symmetry at the phase transition, and a network of such DWs may divide the universe into different sections. The size of DWs is assumed to be on the scale of $d \approx$

* Corresponding author.
E-mail address: yunshin@ibs.re.kr (Y.C. Shin).

https://doi.org/10.1016/j.dark.2022.101118
2212-6864/© 2022 Elsevier B.V. All rights reserved.

$1/m_a$ where d is the thickness of the DW and m_a is the mass of the ALP.

DWs may contribute to the dark matter in the universe. However, stable DWs of QCD axions would be cosmologically disastrous because they would store too much energy [29,30]. Nevertheless, ALP DW could exist up to the modern epoch in the post-inflation scenario since ALP fields are not restricted by the QCD phenomenology [31–36]. If they indeed exist, such ALP DW dark matter would have a highly nonuniform local density. Recently, a series of experiments have been proposed to attempt the direct detection of locally dense dark matter [22,37–46]. The Global Network of Optical Magnetometers for Exotic physics searches (GNOME) is the first experiment to look for localized dark matter governed by transient spin-dependent interactions between the dark matter and atomic spins [37,47–49]. At present, the GNOME consists of more than ten dedicated optical atomic magnetometers located at stations throughout the world as shown in Table 2. The magnetometric sensitivity of each GNOME sensor is roughly in a range of $\approx$ 100 fT/$\sqrt{\text{Hz}}$ over a bandwidth of $\approx$ 100 Hz. Each magnetometer is located within a multi-layered magnetic shield to reduce the influence of magnetic noise and perturbations. Even with the magnetic shielding, there is inevitably some level of transient signals and noise associated with the local environment. Therefore, each GNOME sensor uses auxiliary unshielded magnetometers and other sensors (such as accelerometers and gyroscopes) to measure relevant environmental conditions, enabling exclusion/vetoing of data with known systematic issues. The signals from the GNOME sensors are recorded with accurate timing provided by the global positioning system (GPS) using a custom GPS-disciplined data acquisition system and have a characteristic temporal resolution of $\lesssim$ 10 ms enabling resolution of events that propagate at the speed of light (or slower) across the Earth. Because of the broad geographical distribution of sensors, the GNOME is able to achieve good spatial resolution, acting as an exotic physics "telescope" with a baseline comparable to the diameter of the Earth. The details of this experiment are described in Refs. [37,48].

The common idea of these experiments is to distribute multiple sensors (optical magnetometers in the case of GNOME) across geographically separated locations on the Earth, and connect them as an array. The passage of the Earth through a localized dark matter object may cause an interaction at each sensor with a distinctive amplitude at a certain time depending on the spatiotemporal distribution of the dark matter.

However, the behavior of the signal amplitude and timing from a dark matter crossing event is a priori unpredictable since the local density of the dark matter cannot be determined from existing theories. Therefore, the analysis of such measurements needs to have a feasible model to determine whether the signal pattern actually corresponds to a possible dark matter crossing event. For example, if the distance between domains L is much larger than the Earth scale ($L \gg R_\oplus$), the boundaries of such DW would be approximated as a flat object with non-zero thickness [22]. Then the DW crossing events captured by the detector network can be described by a simple parametric template with the relative velocity and orientation of a flat DW. Nevertheless, the measurements may contain a large multidimensional array of information which could cause complications when accessing from a conventional data analysis scheme. In addition, the fact that the domain-wall crossing event would be unpredictable makes it almost impossible to define a 'signal-to-noise ratio' in these kinds of searches. For example, a geometrical situation (i.e., an orientation of DW) affects the signal strength as well as the physical property of ALPs.

Recent developments in machine learning (ML) techniques have shown the ability to extract a particular feature from multidimensional datasets in various fields, including physics [50,51]. One of these ML methods, stochastic optimization (or stochastic gradient descent) with adaptive momentum, enables one to fit a parametric template of crossing events in a time window of network data without massive calculations of every possible combination of templates [52–56]. This ML assisted fitting method can be used for discerning whether the network data can be well fit to any of the possible DW crossing events. This optimization can also be extended to search not only for DWs, but also for a variety of other locally dense dark matter objects [47,57–60].

In this paper, a data analysis method to discern DW crossing events from the GNOME data is presented. This analysis scheme utilizes a parametric template of DW crossing events, instead of scanning a possible parameter lattice (as was done in Refs. [49,61]). The feasible parameter range of the template and event detection threshold are optimized via a simulated dataset.

The expected signal amplitudes and timings of the DW crossing event are derived in Section 2. The procedure of the data analysis is presented in Section 3. The reliability and validity of the data analysis are studied based on binary classification as described in Section 4. The conclusion and prospects for future data analysis are described in Section 5.

2. Parametric template of the domain-wall crossing event

The ALP fields inside the DW between two neighboring vacua along the normal direction parametrized by z can be described as [22,61]

$$a(z) = \frac{4f_{\text{SB}}}{\sqrt{\hbar c}} \arctan\left(\exp\left(\frac{m_a c^2}{\hbar c} z\right)\right), \tag{2}$$

where f_{SB} is the symmetry-breaking scale of the ALP. The pseudoscalar linear coupling of the ALP field a to the Standard Model axial–vector current has the interaction Hamiltonian from Eq. (1) as

$$H_{\text{int}} = \frac{(\hbar c)^{3/2}}{f_{\text{int}}} \hat{S} \cdot \vec{\nabla} a, \tag{3}$$

where f_{int} is the effective interaction scale between the gradient of the ALP field $\vec{\nabla} a$ and unit fermionic spins $\hat{S}$. The effective interaction scale can be different for electrons, neutrons, and protons depending on the particular theoretical model, but here we consider only ALP-proton coupling. The GNOME magnetometers are placed inside multi-layer magnetic shield, which cancel the effects of electron spin couplings due to an induced magnetization of the shield [62]. All GNOME magnetometers use atoms whose nuclei have valence protons, and thus are primarily sensitive to ALP-proton interactions [63].

The ALP-DW interactions with atomic spins can be interpreted as a pseudo-magnetic field $\vec{B}$ acting on the atoms. It can be written in analogy with the Zeeman Hamiltonian form as

$$H = -\gamma \vec{S} \cdot \vec{B}, \tag{4}$$

where γ is the gyromagnetic ratio of atomic species employed in each magnetometer. While a conventional magnetic field is screened by the multi-layer magnetic shields, the ALP-proton coupling cannot be screened as they are treated as pseudo-magnetic fields, thereby being detected by the atomic magnetometer.

The expected strength of the pseudo-magnetic field B_s at the magnetometer labeled by s is derived as

$$B_s = \frac{4}{\mu_B} \frac{f_{\text{SB}}}{f_{\text{int}}} m_a c^2 \frac{\sigma_s}{g_{F,s}} \cos\psi_s, \tag{5}$$

where μ_B is the Bohr magneton, $\sigma_s/g_{F,s}$ is the estimated ratio between the effective proton spin polarization and the Landé g-factor for the magnetometer s, and ψ_s is the angle between

the ALP-field gradient and the sensitive direction of the magnetometer [61]. Here, $f_{\rm SB}$ can be expressed by the cosmological parameters through the domain-wall energy density $\rho_{\rm DW}$ and surface tension $\sigma_{\rm DW} \sim L\rho_{\rm DW}$. Using Eq. (2), one has a relationship as

$$L\rho_{\rm DW} \sim \sigma_{\rm DW} \sim \int dz \left|\frac{da}{dz}\right|^2 = \frac{8f_{\rm SB}^2 m_a c^3}{\hbar^2 c^2}. \tag{6}$$

If the domain walls take a portion of dark matter in our universe, then

$$f_{\rm SB} = \hbar c\sqrt{\frac{L\rho_{\rm DW}}{8m_a c^2}} \leq \hbar c\sqrt{\frac{L\rho_{\rm DM}}{8m_a c^2}}, \tag{7}$$

where $\rho_{\rm DM}$ is the energy density of dark matter.

The expected strength of the pseudo-magnetic field can be written with cosmological parameters L and $\rho_{\rm DW}$ as

$$\begin{aligned} B_s &= \frac{4}{\mu_B}\frac{\sigma_s}{g_{F,s}} f_{\rm SB}\sqrt{m_a c^2}\frac{\sqrt{m_a c^2}}{f_{\rm int}}\cos\psi_s \\ &= \frac{4\hbar c}{\sqrt{8}\mu_B}\frac{\sigma_s}{g_{F,s}}\sqrt{L\rho_{\rm DW}}\frac{\sqrt{m_a c^2}}{f_{\rm int}}\cos\psi_s. \end{aligned} \tag{8}$$

We assume that $L \approx 7.5 \times 10^{-5}$ ly (wall crosses the Earth approximately once a month with a relative velocity $|v_d| = 10^{-3}c$) and $\rho_{\rm DW} = 0.4$ GeV/cm^3 (local dark matter density [1]). Independently measured L and $\rho_{\rm DW}$ can change the effective interaction scale.

The expected amplitude of the effective magnetic field B_s depends not only on the ALP parameters (mass and effective interaction scale) but also on the properties of the magnetometer and geometric factors. In particular, the factor $\cos\psi_s$ that describes a relative crossing direction could suppress the magnetometer signal, even to zero in some cases.

If a DW crossing event takes place, the timings $t_{0,s}$ of the expected signal at each detector s are also unpredictable in a priori because they are determined by the relative positions between the magnetometers and DW. For simplicity, we assume that the DW is relatively flat with respect to Earth scale, and it travels at velocity $\vec{v}_d$, initially at $\vec{x}_d$ as the closest point in the wall to Earth. Then the encounter time can be estimated as

$$t_{0,s} = \frac{(\vec{x}_d - \vec{x}_s)\cdot\hat{x}_d}{|\vec{v}_d|}, \tag{9}$$

where $\vec{x}_s$ is the position of the magnetometer s. The signal duration τ is characterized by the DW thickness $d = 2\sqrt{2}\hbar c/m_a c^2$ and the relative speed $|\vec{v}_d|$

$$\tau = \frac{2\sqrt{2}\hbar c}{m_a c^2\,|\vec{v}_d|}, \tag{10}$$

independent of the sensor.

Fig. 1 shows the conceptual diagram of the magnetometer s on the Earth's surface at $\vec{x}_s = R_\oplus\hat{x}_s$ with the sensitive direction $\hat{\alpha}_s$ and the DW at its position $\vec{x}_d$. $R_\oplus$ denotes the Earth radius. Each of the position vectors can be represented using spherical coordinates centered at the Earth's center, aligned to the North pole (polar angle) and prime meridian (azimuthal angle). Then the DW crossing event parameters for generating the signal pattern (amplitude B_s, timing $t_{0,s}$, and duration τ) are determined by the six parameters listed in Table 1.

3. Method

The likelihood of a geographically correlated signal pattern in the data being produced by a DW is evaluated by comparing it to the predicted pattern from Eqs. (8), (9), and (10). If such an event occurs and generates signals distinguishable from noise, the relevant physical parameters of the correlated pattern can also be estimated. This process can be divided into several steps. (1) Prepare a test-time window to be analyzed containing data from the stations operating during that specific interval of time. Data pre-processing is applied to each point in the test-time window. (2) Generate the parameter space for variables used to optimize the parametric model for the data. (3) Perform the stochastic optimization for fitting the pattern to the data and evaluate the goodness of the fitting. Here, the parameter-estimation error will be used. (4) Characterize the test-time window based on the evaluated estimation error.

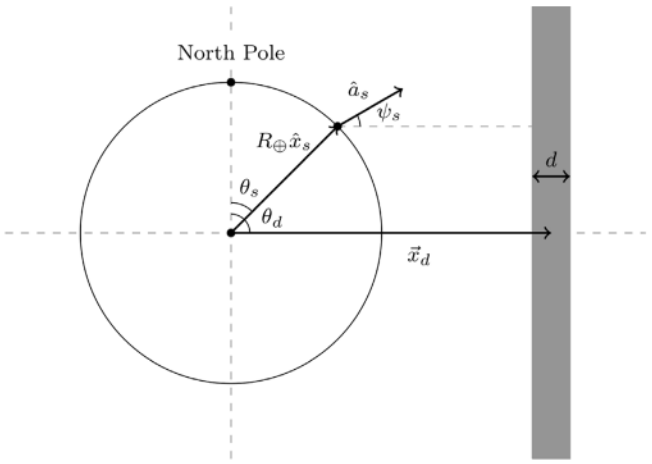

Fig. 1. Conceptual diagrams of the Earth (left circle) and DW (right rectangle) are represented to describe the geometrical parameters. For each magnetometer s, its position vector $\vec{x}_s = R_\oplus\hat{x}_s$ is described by the spherical coordinates $(R_\oplus, \theta_s, \phi_s)$ centered at the Earth center. The sensitive direction vector of the magnetometer s is denoted by $\hat{\alpha}_s$. A position of a DW, $\vec{x}_d$, is described in the same manner as $(|\vec{x}_d(t)|, \theta_d, \phi_d)$. d is the DW thickness and ψ_s is the angle between $\hat{\alpha}_s$ and $\vec{x}_d$.

A test-time window of data is defined by a set of discrete data points from each sensor in the network. For a given time interval and sampling rate, the test-time window is constructed from all data points from all available sensors during the time interval. The linear baseline of each sensor is removed by subtracting a linear fit to the data in advance. There are no additional filters or time binning in the pre-processing step.

Then the stochastic optimization iteratively updates the template parameters to fit a signal pattern to the test-time window. In order to build the stochastic optimization process to search for DW crossing events, the proper template parameters of the event should be determined. The parameters are updated based on the gradient with respect to them in the parameter space, towards the minimum value of the estimation error. Each parameter should be normalized to prevent a directional bias during every updating iteration. This normalization requires a definite boundary for each parameter.

Fig. 2 visualizes a test-time window and fitted pattern of network data. Arbitrary peaks are injected to mimic a fake DW signal pattern. The sign, amplitude, and timing of peaks are determined by the relative position and orientation between the network of magnetometers and DW. These signal properties are peculiar to each station. Different positions and sensitive axis of each station also make different signal patterns. Three time-series data (semi-transparent lines) in a synchronized time window are prepared as the test-time window. Correlated patterns in the given time windows are estimated from stochastic optimization (solid lines). The characteristics of the network would vary depending on the position, sensitive axis, and operation duty cycle of each station.

Table 1
Six parameters describing DW crossing event and their estimation range. The parameter boundaries are normalized to the unit interval, and the azimuthal angle indicating the DW has a periodic boundary condition.

Parameters	Symbols	Estimation ranges	Normalization maps $f(x) : x \mapsto f(x)$
Mass	$m_a c^2$	$[10^{-15}$ eV, 10^{-11} eV$]$	$(\log_{10}(x/\text{eV}) + 15)/4$
Interaction scale	f_{int}	$[10^4$ GeV, 10^8 GeV$]$	$(\log_{10}(x/\text{GeV}) - 4)/4$
Polar angle	θ_d	$[0, \pi]$	x/π
Azimuthal angle	ϕ_d	$[0, 2\pi)$	$x/2\pi$
Relative speed	$\lvert\vec{v}_d\rvert$	[100 km/s, 550 km/s]	$(x - 100$ km/s$)/450$ km/s
Relative position	$\lvert\vec{p}_d\rvert$	$[6.4 \times 10^3$ km, 12×10^3 km$]$	$(x - 6.4 \times 10^3$ km$)/5.6 \times 10^3$ km

3.1. Boundary, distribution, and normalization of the DW crossing event parameters

For ALP DW, the mass $m_a c^2$ and the effective interaction scale f_{int} are free parameters. So they are treated as unknown parameters with logarithmic-uniform distribution. The values of $m_a c^2$ and f_{int} will be estimated as a point in the distribution during the analysis. Their boundaries are set by referring to the prior GNOME analysis range as shown in Table 1 [61].

The direction parameters, polar and azimuthal angles of the DW, have a feasible range with a linear-uniform distributions, where the polar angle θ_d satisfies $\theta_d \in [0, \pi]$ and the azimuthal angle ϕ_d satisfies $\phi_d \in [0, 2\pi)$ with a periodic boundary condition.

The speed should be considered a random variable from the Maxwell–Boltzmann distribution if DWs are virialized in our galaxy. In addition, the escape velocity at a solar system galactic radius from the Milky Way's gravity is given by $v_e = 550.9$ km/s, so we assume that faster DWs cannot exist in the galaxy [64]. A slow DW could leave a signal pattern on the data, but long-term linear drifts in the magnetometers make such a DW to be difficult to characterize [48].

In the GNOME experiment, each GNOME data file contains one minute long information. As the GNOME data is a collection of the synchronized time series data from multiple stations, each synchronized data consists of 60 s long data from each station with HDF5 format which are uploaded to the GNOME server every minute.

Since GNOME data files are stored for every minute (regardless of the data acquisition rate), the minimum speed of the DW that allows the DW to pass the Earth within at most n-concatenated data files is

$$v_{\text{min}} = \frac{2R_\oplus}{n \text{ minutes}} \approx \frac{213}{n} \text{ km/s}. \tag{11}$$

In order to minimize the effect of drifts in the magnetometer data, n is chosen to be 2. Then the boundary of the relative speed of the DW is $|\vec{v}_d| \in [100$ km/s, 550 km/s$]$. A 7 km/s margin is given for the lower bound for the convenience of calculation. The speed follows the Maxwell–Boltzmann distribution with the scale parameter of $220/\sqrt{2}$ km/s and edges on both sides are cut to exclude slow and fast DWs [65].

The initial relative distance is also restricted by the same data length, $n = 2$. To contain all signal peaks (from the closest and farthest surfaces of the Earth) in a test-time window, the boundary is needed to be

$$\left|\vec{x}_d\right| \in [R_\oplus, \min(|\vec{v}_d|) \cdot n \text{ minutes}] \tag{12}$$

The initial relative distance is uniformly distributed in the range.

In the worst case (the slowest and farthest DW from the Earth), $|\vec{v}_d| = 107$ km/s and $\left|\vec{x}_d\right| = 12 \times 10^3$ km, the DW could not pass through the Earth completely within a given test-time window of $n = 2$. Instead, it will be contained in the one-minute overlapping neighboring test-time window with a new set of parameters $|\vec{v}_d| = 107$ km/s and $\left|\vec{x}_d\right| = 6.4 \times 10^3$ km. This DW or other combinations of slow and far DWs can be searched

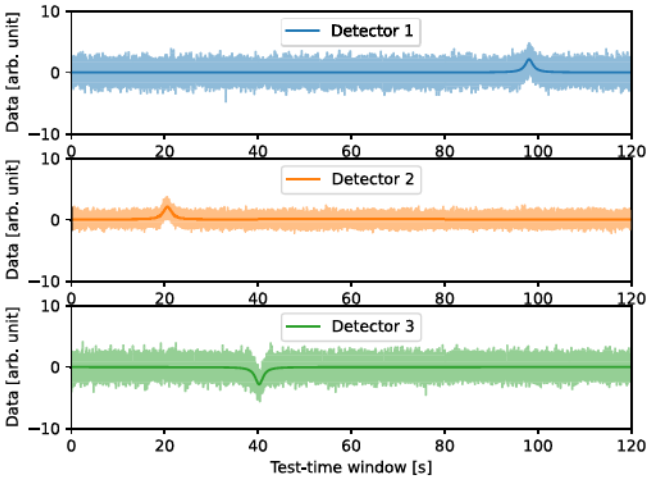

Fig. 2. An example of the test-time window for three sensors. Arbitrary peaks are injected into the data from the station as shaded fluctuations. Solid lines represent the fitted pattern from stochastic optimization. The signal properties, which are sign, amplitude, and timing, are determined by the relative position and orientation between the network of magnetometers and DW.

with $n/2$-minutes overlapping adjacent test-time windows by shifting $\left|\vec{x}_d\right|$. For a single test-time window search, the range of the DW crossing event for the relative speed and distance $|\vec{v}_d| \in [100$ km/s, 550 km/s$]$ and $\left|\vec{x}_d\right| \in [6.4 \times 10^3$ km, 12×10^3 km$]$ is sufficient.

The range and distribution of parameters describing the DW crossing events are represented in Table 1. In actual analysis, they are calculated in the normalized unit space through normalization maps. The normalization enables the parameters with different scales and units to have a uniform gradient scale during the stochastic optimization.

3.2. Stochastic optimization

Stochastic optimization is a stochastic version of the gradient descent optimization, which finds the minimal value of a given function based on Newton's method. A function to be optimized, the cost function, is traced by updating based on the gradient with respect to optimization parameters. In our case, the normalized optimization parameters are DW crossing event parameters listed in Table 1. However, since there are inevitably noise fluctuations in the data, and infinitely many forms of the signal pattern for a given DW crossing event, the shape of the cost function should reflect such fluctuations and non-linearity.

An ansatz of the cost function $\mathcal{E} = \frac{1}{N}\sum_s \mathcal{E}_s$ to estimate the DW crossing event for each magnetometer s is defined as

$$\mathcal{E}_s = \frac{1}{\mathcal{T}\sigma_s^2} \int_0^{\mathcal{T}} \left(\int_0^t \left(S_s(t') - \tilde{S}_s(t') \right) dt' \right)^2 dt, \tag{13}$$

where N is the number of magnetometers, $\mathcal{T}$ is the time interval of the test-time window, σ_s is the standard deviation of the time-series data at the magnetometer for a given time window, $S_s(t')$ is

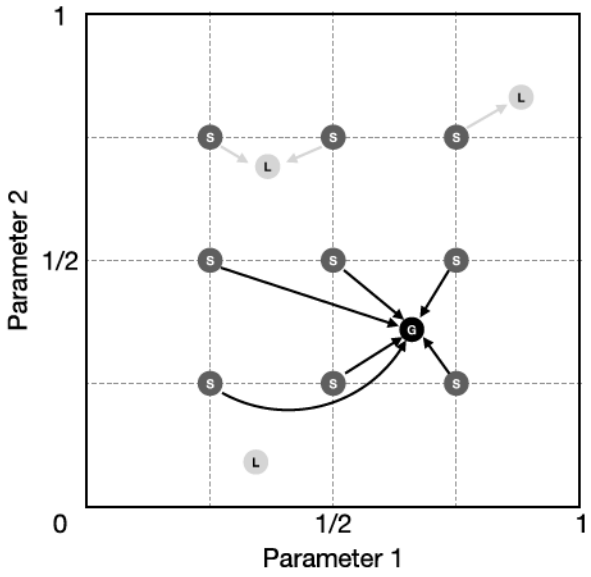

Fig. 3. A conceptual diagram of the grid estimation with various initial points in the normalized parameter space. In two normalized parameters estimation, there is one global minimum (G) that is the ground truth, but local minima (L) obstruct the estimation. The grid estimation uses multiple initial points (S) in the parameter space to find the most probable estimation solution.

the baseline-removed time-series data of the magnetometer s at time-series point t', and $\tilde{S}_s(t')$ is the expected signal pattern of the magnetometer s at time-series point t'. The time variables t and t' in $[0, \mathcal{T}]$ represent the common time interval for the sensors, digitized during the data acquisition.

The cumulative integration over the time-series point t' can reduce the local fluctuation of the Gaussian noise. $\mathcal{T}$ in the denominator normalizes the length of the interval, while the variance of the magnetometer σ_s^2 normalizes the Gaussian noise. Practically, the magnetometer data does not precisely follow a Gaussian distribution, but this pseudo-normalization factor $1/\sigma_s^2$ weights the contribution of each magnetometer to the cost functions by their intrinsic noise. Therefore, it makes the cost function have less dependence on the sensor characteristics, or even the number of sensors in general. The details of the cost function ansatz is described in Appendix B.

The remaining dependence on detector characteristics can be avoided by clustering distinct optimization processes from the multiple initial points, which generate $\tilde{S}_s$ on the estimating parameter space grid. Multiple initial points are defined as intersections of D grid lines on the parameter space. The optimization is conducted for each D^m initial points for m parameters. Fig. 3 shows a conceptual diagram of the grid estimation in the normalized parameter space with $D = 3$. Without loss of generality, the parameter space is represented as a two-dimensional space with two normalized parameters as an example for demonstration. Then it has $3^2 = 9$ initial grid points.

The optimization process is conducted for each initial point (S), and each of them move towards the local minima (L) or the global minimum (G) depending on the situation. They may not converge in a finite number of iterations to any of minima. Instead, the estimated parameter points (including minima) at the last iteration step cluster in the parameter space with a distance d_c. Then each of them forms a hypersphere with the radius d_c representing the cluster. The cost function and parameter values at clusters are averaged and compared to evaluate the minimal cost cluster. Clusters are placed far from each other, at least d_c, and each of clusters represents different estimated parameters. Then they have different cost function values. The minimal cost cluster is a representative estimated parameter point of the grid estimation. Two parameters determine the overall optimization performance of the grid estimation. The number of grid lines D balances computing time against precision. Also, a clustering distance d_c among optimized results is related to accuracy.

The ADAM (ADAptive Momentum) optimization is employed for the stochastic optimization process, which can cover a larger range of the evaluated gradients by using machine learning [56]. The learning rate is one of the representative hyper-parameters in the machine learning. If the learning rate is too large, the training is not stable, but if this value is too small, the training becomes too slow. Therefore, the learning rate is the parameter that needs be tuned before anything else. Gradient descent is an algorithm to optimize the learning rate. The idea of the gradient descent is to follow the negative gradient of an objective function in order to locate the minimum of the function. A problem with gradient descent is that it can bounce around the search space on optimization problems that have large amounts of curvature or noisy gradients, thereby getting stuck in flat spots in the search space that have no gradient.

To avoid this issue, one can optimize the weight parameters more easily by adding various techniques to this existing gradient descent method. One such technique is to use what is known as "momentum" in the gradient descent optimization algorithm, where momentum accounts for the history of past steps in calculating the next step in parameter space. Momentum provides a number of benefits. One can not only easily escape the local minima through momentum, but also accelerate the learning rate. ADAM optimizer is one of optimizers the most widely used in the modern deep machine learning. ADAM can be seen as an algorithm that adds momentum to the existing adaptive learning rate method. In addition, by supplementing existing algorithms, it has become the most effective algorithm so far. Therefore, the training scheme can have the effect of being robust to the learning rate when using ADAM optimization.

The expected signal amplitudes can vary from zero to more than a picoTesla amplitude depending on the parameters and geometric properties of the DW crossing event. The ADAM optimization can update the estimating parameters during the optimization, to fit the pattern to wide ranges of signal amplitudes with a universal process.

3.3. Performance evaluation

The stochastic optimization evaluates an estimation error given by the cost function value $\mathcal{E}$. If the estimation error is small enough for a given test-time window, a physical model (the DW crossing event in this case), and an appropriate optimization process, then this test-time window has a possibility to contain the physical event. However, noisy test-time widows can sometimes produce a small estimation error. On the other hand, a high value of $\mathcal{E}$ means it does not contain the physical event or the optimization process cannot find an appropriate solution. The decision about the presence of an event is characterized by a binary classification. We can define four different cases in terms of whether our data contains an event and whether the algorithm identifies it as true positive (TP), false negative (FN), false positive (FP), and true negative (TN).

The binary classification of the DW crossing event has been tested based on simulations with GNOME data. The active magnetometers of the GNOME from January 30th to April 30th, 2020 are listed in Table 2. Since each of the magnetometers in the GNOME do not always provide proper data due to local glitches, the data may not be continuous in time. Therefore, the available dataset from magnetometers for a given time interval may have

Table 2
Geographic and geometric information of the GNOME station magnetometers. The magnetometer position is based on the global positioning system (GPS), where the West and South directions have a negative sign. The sensitive direction is based on the horizontal coordinate system.

GNOME stations	Magnetometer position		Sensitive direction		σ/g_F
	Longitude [deg]	Latitude [deg]	Altitude [deg]	Azimuth [deg]	
Berkeley 1	−122.3	37.9	0	28	−0.39
Berkeley 2	−122.3	37.9	90	0	−0.39
Daejeon	127.4	36.4	90	0	−0.39
Hayward	−122.1	37.7	90	0	0.70
Krakow	19.9	50.0	0	45	0.50
Lewisburg	−76.9	41.0	90	0	0.70
Los Angeles	−118.4	34.1	0	270	0.50
Mainz	8.2	50.0	−90	0	0.50
Moxa	11.6	50.6	0	270	−0.39
Oberlin	−81.8	41.3	0	300	−0.49

different combinations of magnetometers. For each simulation, a random combination of magnetometers is chosen to cover the most general case. However, the number of magnetometers must be larger than 3 in order to at least estimate the direction of the DW. To prevent any real dark matter event appearing during the simulation, a two-minute time window was generated using the mixed date and time of each magnetometer station.

In total 400 simulations were conducted for the binary classification. Among them, 200 simulations have a randomly injected DW crossing event from the boundary and distribution of the estimating parameters, while the remaining 200 simulations have artificially injected random noise Lorentzian peaks (0 to 4 peaks at random timings, uniformly distributed between −50 and 50 pT amplitude, uniformly distributed between 0.12 and 12 seconds long full-width at half-maximum) to test the robustness of the algorithm against other patterns corresponding no DW crossing event. Both signal and noise peak shapes are assumed to be Lorentzian. The number of grid lines $D = 2$ and the clustering distance $d_c = 0.02$ were set during analysis. For each optimization at a single grid point, the DW parameters are estimated by 500 iterations steps of the ADAM optimization.

In order to characterize the binary classification, each simulation is identified as positive or negative depending on the decision criterion, which in this analysis is the estimation error $\mathcal{E}$. Therefore, the true positive rate (TPR) and false positive rate (FPR) are derived with respect to the estimation error. The decision criterion threshold $\mathcal{E}_{\rm th}$ is determined at 95% level of the TPR. The confidence interval of each rate p (TPR or FPR) in n observations is described by the continuity-corrected Wilson score interval with a critical value z [66–69]

$$p = \frac{2n\hat{p} + z^2}{2\left(n + z^2\right)} \pm \frac{1}{2\left(n + z^2\right)}\left(1 + z\sqrt{4n\hat{p}(1-\hat{p}) \mp 2\left(1 - 2\hat{p}\right) - \frac{1}{n} + z^2}\right). \tag{14}$$

For the 95% of confidence level, $z = 1.96$. The continuity-corrected Wilson score interval at the 95% confidence level would be applied to estimate the confidence intervals of TPR and FPR.

3.4. Parameter space optimization

The set of virtual observations, 400 simulations, is classified within the parameter space bounded by the estimation range listed in Table 1. In the presence of an event for a given test-time window, the optimization estimates the corresponding DW parameters simultaneously. The estimated parameters need to be accurately converged, but sometimes they do not converge to stable values within a reasonable iteration time, due to computing limitations. This can be handled by applying a finer grid estimation and larger iterations of the fitting.

More efficiently, the ALP field parameter space generated from the mass and effective interaction scale can be optimized without any increase in computing cost by using a two-step process. The first step is to calculate a distance representing the estimation error. In the virtual observations with DW crossing events, which are characterized by a certain set of DW parameters, the distance between the estimated parameters and the desired parameters in the ALP field parameter space can be different for distinct regions in the space. Since the optimization is conducted in the normalized space, the distance can be defined as a norm. The norm indicates the distance function. Since the space is well normalized, the metric can be well defined. Hence, the norm that determines the distance value can also be well defined. Then the distance indicates the error level of the estimated parameters from the desired parameters.

The second step is to get an optimal parameter space corresponding to the projected sensitivity limit of this analysis. The distances calculated in the first step are used in this optimization process. Let a well-estimated candidate be when the optimization gives a distance of less than 0.02 (in the normalized space). This distance and corresponding candidate are meaningful only if the analysis method detects an event, i.e., TP event cases. The accuracy is defined as a population ratio between the well-estimated candidates and the total TP events. The ALP field parameter space can be optimized to maximize the sum of the accuracy and the area (mass range × interaction scale range) of optimized space. The subspace of the ALP field parameter space is swept by the mass and effective interaction scale. Then the parameter space sensitive to this analysis method will be derived.

This analysis method would not be able to guarantee the identification of the domain-wall crossing events correctly beyond the optimized parameter space. Since, the detector sensitivity is as critical as the analysis method in the sensitive parameter space, any dark matter signal insensitive to detectors cannot be analyzed. For example, weak signals with a huge $f_{\rm int}$ would effectively show no signals in datasets. The parameter space optimization represents a sensitive region of ALP field parameters for a current network detector architecture.

4. Result

Fig. 4 represents the TPR (blue) and FPR (orange) of the DW crossing event classification from the simulations before (left) and after (right) the parameter space optimization. The corresponding areas under the receiver operating characteristic (AUROC) curves were measured to be 0.963 and 0.974 [70]. The performance of the classification was enhanced after the parameter space optimization. Before the parameter space optimization, the decision

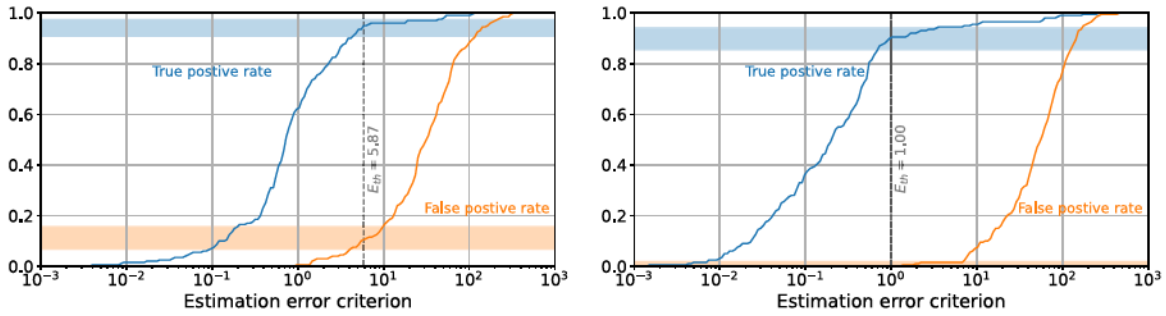

Fig. 4. Binary classifications of the analysis method with simulated DW crossing events before (left) and after (right) the parameter space optimization. TPR (blue strip on top of the region) and FPR (orange strip on bottom of the region) are represented with respect to the estimation error as confidence intervals of 95%. Before the parameter space optimization, the TPR and FPR are 95% and 10% at the decision criterion threshold $E_{\text{th}} = 5.87$. The TPR and FPR are changed to 91% and 0.0% at $E_{\text{th}} = 1.00$ after the parameter space optimization.

criterion threshold is derived to the value $E_{\text{th}} = 5.87$, which corresponds to the 95% of TPR from simulations. The FPR at the threshold was observed to be 10%. The corresponding confidence intervals with this classification algorithm were a TPR between 90% to 97% and an FPR in between 6% to 15%.

Based on the classification result, the parameter space searched as seen in Table 1 was optimized, as shown in Fig. 5. The 400 virtual observations were simulated within the optimized parameter subspace. The result shows a TPR as 91%. The TPR is in [85%, 94%] and the FPR is in [0.0%, 1.8%] at the 95% confidence level, when the threshold is set to 1.00 as shown in the right of Fig. 4. The optimized parameter space is a projected limit of the described algorithm for the GNOME setup listed in the Table 2 to search for DW crossing events, corresponding to $m_a c^2$ from 1.00×10^{-14} eV to 1.34×10^{-12} eV and f_{int} from 10^4 GeV to 4.65×10^5 GeV with a 2% acceptance error (a distance of 0.02 in the normalized unit space). For $f_{\text{int}} \leq 10^4$ GeV, the signal pattern would show the same pattern, but enlarged amplitude. They could be covered by extending the boundary of the effective interaction scale.

This result improves the parameter space analyzed with GNOME data as described in Ref. [61], but it is worth noting that the network status in Ref. [61] is dedicated to the Science Run 2, while present data analysis method is based on the Science Run 4 status. Also, this method is independent of $f_{\text{SB}}/f_{\text{int}}$, instead, it has to scan the continuous data for estimating the domain size L.

This study demonstrates a feasibility of new data analysis with machine learning scheme not only for the GNOME experiment but also for other dark matter searches. The projected parameter space can be further improved as more virtual observations are simulated with intensive computations with high-performance computing resources.

5. Conclusions

A new data analysis method based on machine-learning-assisted stochastic optimization has been presented to search for direct detection of localized dark matter using GNOME data. The identification and characterization of the ultra-light ALP DW events were evaluated using in simulated virtual observations. The identification showed TPR of at least 85% and FPR at most 2% at 95% confidence level, and characterization showed an accuracy of 88%. This accuracy can be improved by rescanning and complementary analyses.

This new method allows us to investigate signal patterns for localized dark matter using a geographically distributed network of sensors. The methodological application is not limited to DW signals, but can be extended to more general signals. Furthermore, it is also possible to employ a network of heterogeneous detectors if the pattern is theoretically predictable for each detector [60].

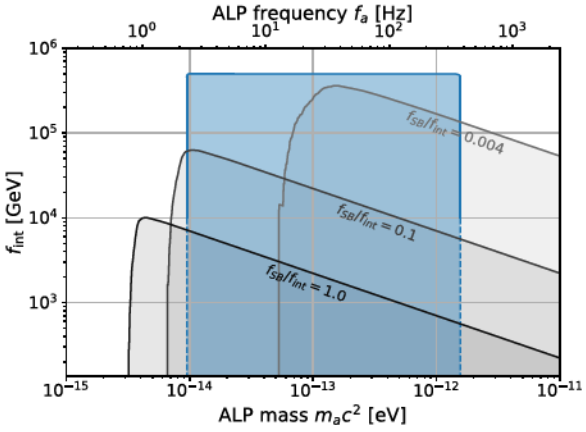

Fig. 5. The projected limit of the optimized parameter space identified and characterized by this method (blue). Previously analyzed regions with GNOME data and analysis method are presented for comparison (gray-scale) [61]. The acceptance error is 2% for the estimated parameter within $m_a c^2$ from 1.00×10^{-14} eV to 1.34×10^{-12} eV and f_{int} from 10^4 GeV to 4.65×10^5 GeV.

CRediT authorship contribution statement

Dongok Kim: Conceptualization, Methodology, Software, Writing – original draft. **Derek F. Jackson Kimball:** Methodology, Validation, Supervision. **Hector Masia-Roig:** Visualization, Investigation. **Joseph A. Smiga:** Visualization, Investigation. **Arne Wickenbrock:** Supervision. **Dmitry Budker:** Supervision. **Younggeun Kim:** Validation. **Yun Chang Shin:** Project administration, Funding acquisition, Writing – original draft, Writing – review & editing. **Yannis K. Semertzidis:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank to Vincent Dumont and Chris Pankow for early contribution to the data analysis, and to all the members of GNOME collaboration for helpful insights and discussions.

This work was supported by the Institute for Basic Science under grant No. IBS-R017-D1-2021-a00. The work of Derek F. Jackson Kimball was supported by the U.S. National Science Foundation under grant No. PHY-1707875 and PHY-2110388. The work of Dmitry Budker was supported by the European Research Council under the European Union's Horizon 2020 Research and Innovative Program under Grant agreement No. 695405, the Cluster of Excellence "Precision Physics, Fundamental Interactions, and Structure of Matter" (PRISMA+ EXC 2118), DFG Reinhart Koselleck (Project ID 390831469), Simons Foundation, USA, and Heising-Simons Foundation, USA.

Appendix A. Pseudocode of the data analysis algorithm

The pseudocode for the algorithm described in the main text is written in Algorithm 1.

Algorithm 1: Event parameters estimation

Data: Multiple time series $S_s(t)$ for $s \in S$
Result: Estimated error $\mathcal{E} = 1/|S| \sum_s \mathcal{E}_s$
for *each grid point g* **do**
  Optimize the domain-wall parameters $\vec{d}^g$;
  Evaluate estimation error $\mathcal{E}^g$;
end
for *each grid point g* **do**
  for *each grid point g′* **do**
    if $\sqrt{\left|\vec{d}^g - \vec{d}^{g'}\right|^2} < d_c$ **then**
      Clustering g' into cluster g;
    end
  end
  if *No clustering for any g′* **then**
    Discard the grid point g;
  else
    Evaluate representative estimation error $\mathcal{E}^g$ and corresponding $\vec{d}^g$;
  end
end
Find $\min_g(\mathcal{E}^g)$ and corresponding $\vec{d}^g$;

Appendix B. The cost function

The most straightforward ansatz function would be with the following form:

$$\mathcal{E}_s = \frac{1}{\mathcal{T}\sigma_s^2} \int_0^{\mathcal{T}} \left(S_s(t) - \tilde{S}_s(t)\right)^2 dt. \tag{B.1}$$

This equation is actually valid in most cases. In certain cases, however, especially when the scale of $\mathcal{T}$ is larger than that of τ, the integral value is barely affected by t_0 if $\tilde{S}_s(t)$ is fixed. This makes it difficult to see when the actual event occurred if the signal duration in a given time window is short.

If the rise time of an inferred event signal is longer than the duration, it is difficult to evaluate how similar the arising time of the inferred event signal with the data mainly because the integral value does not change even if the rise time is changed.

However, if the integration with respect to time is performed once more as shown in Eq. (13), the integral value would also increase in proportion to the deviation of the arising time of an inferred event signal from the real event time. This makes it possible to evaluate how the arising time of the inferred event signal is actually close to the real event time.

In the case of the GNOME, the model parameter vector $\vec{X}$ in 6-dimensional space represents an axion DW crossing event. For the given set of GNOME stations $\mathcal{S}$, the measurement data S_s from each atomic magnetometer s is defined as

$$\{S_s(t) \mid s \in \mathcal{S}\}. \tag{B.2}$$

Then the cost function should be minimized when the signal data $\tilde{S}_s(t;\vec{X})$ is equal to the actual measurement data $S_s(t)$ for any station $s \in \mathcal{S}$. Therefore, that the cost function of any station s can be expressed with a form of the mean squared error (mse) as follows

$$\mathcal{E}_s(\vec{X}) = \sum_s \int \left(S_s(t) - \tilde{S}_s(t;\vec{X})\right)^2 dt. \tag{B.3}$$

With the given number of GNOME stations in $\mathcal{S}$, an ansatz of the cost function to test the ALP DW signal is defined as

$$\begin{aligned}\mathcal{E}(\vec{X}) &= \frac{1}{N}\sum_s \mathcal{E}_s(\vec{X}) \\ &= \frac{1}{N}\sum_s k_s \int_0^{\mathcal{T}} \left(\int_0^t \left(S_s(t') - \tilde{S}_s(t';\vec{X})\right) dt'\right)^2 dt,\end{aligned} \tag{B.4}$$

where N is the number of stations in s, $\mathcal{T}$ is the size of the test time window, and k_s is a normalization coefficient depending on the measurement data at station s for a given test time window as follows,

$$k_s = \frac{1}{\mathcal{T}\,(\sigma_s\ s)^2}. \tag{B.5}$$

References

[1] J.I. Read, J. Phys. G Nucl. Part. Phys. 41 (2014) 063101.
[2] K.A. Hegyi, D.J. Olive, Phys. Lett. B. 126 (1983) 28.
[3] J. Preskill, M.B. Wise, F. Wilczek, Phys. Lett. B. 120 (1983) 127.
[4] P. Abbott, L. Sikivie, Phys. Lett. B. 120 (1983) 133.
[5] W. Dine, M. Fischler, Phys. Lett. B. 120 (1983) 137.
[6] E. Witten, Phys. Rev. D. 30 (1984) 272.
[7] L.M. Dodelson, S. Widrow, Phys. Rev. Lett. 72 (1994) 17.
[8] G. Jungman, M. Kamionkowski, K. Griest, Phys. Rep. 267 (1996) 195.
[9] W. Hu, R. Barkana, A. Gruzinov, Phys. Rev. Lett. 85 (2000) 1158.
[10] G. Kim, J.E. Carosi, Rev. Modern Phys. 82 (2010) 557.
[11] H.R. Peccei, R.D. Quinn, Phys. Rev. Lett. 38 (1977a) 1440.
[12] H.R. Peccei, R.D. Quinn, Phys. Rev. D 16 (1977b) 1791.
[13] S. Weinberg, Phys. Rev. Lett. 40 (1978) 223.
[14] F. Wilczek, Phys. Rev. Lett. 40 (1978) 279.
[15] J.E. Kim, Phys. Rev. Lett. 43 (1979) 103.
[16] M. Shifman, A. Vainshtein, V. Zakharov, Nucl. Phys. B. 166 (1980) 493.
[17] M. Dine, W. Fischler, M. Srednicki, Phys. Lett. B. 104 (1981) 199.
[18] A. Zhitnitsky, Sov. J. Nucl. Phys. 31 (1980) 260.
[19] E. Svrcek, P. Witten, J. High Energy Phys. 2006 (2006) 051.
[20] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, J. March-Russell, Phys. Rev. D. 81 (2010) 123530.
[21] A. Ringwald, Axions and axion-like particles, 2014.
[22] M. Pospelov, S. Pustelny, M.P. Ledbetter, D.F.J. Kimball, W. Gawlik, D. Budker, Phys. Rev. Lett. 110 (2013) 021803, 1205.6260.
[23] D.J.E. Marsh, Phys. Rep. 643 (2016) 1.
[24] K. Choi, S.H. Im, C.S. Shin, Ann. Rev. Nucl. Part. Sci. 71 (2021).
[25] J.L. Wechsler, R.H. Tinker, Ann. Rev. Astron. Astrophys. 56 (2018) 435, http://dx.doi.org/10.1146/annurev-astro-081817-051756.
[26] P. Sikivie, Phys. Rev. Lett. 48 (1982) 1156.
[27] A. Vilenkin, Phys. Rep. 121 (1985) 263.
[28] T.W.B. Kibble, J. Phys. A Math. Gen. 9 (1976) 1387, http://dx.doi.org/10.1088/0305-4470/9/8/029.
[29] Y.B. Zel'Dovich, I.Y. Kobzarev, L.B. Okun', Sov. J. Exp. Theor. Phys. 40 (1975) 1.
[30] J. Preskill, S.P. Trivedi, F. Wilczek, M.B. Wise, Nucl. Phys. B. 363 (1991) 207.
[31] S.E. Larsson, S. Sarkar, P.L. White, Phys. Rev. D. 55 (1997) 5129.
[32] A. Friedland, H. Murayama, M. Perelstein, Phys. Rev. D. 67 (2003) 043519, https://link.aps.org/doi/10.1103/PhysRevD.67.043519.
[33] T. Hiramatsu, M. Kawasaki, K. Saikawa, T. Sekiguchi, J. Cosmol. Astropart. Phys. 2013 (2013) 001.
[34] P.P. Avelino, Phys. Rev. D. 101 (2020) 023514.
[35] M. Ibe, S. Kobayashi, M. Suzuki, T.T. Yanagida, Phys. Rev. D. 101 (2020) 035029, https://link.aps.org/doi/10.1103/PhysRevD.101.035029.

[36] G.B. Gelmini, A. Simpson, E. Vitagliano, Phys. Rev. D. 104 (2021) L061301, https://link.aps.org/doi/10.1103/PhysRevD.104.L061301.
[37] S. Pustelny, D.F. Jackson Kimball, C. Pankow, M.P. Ledbetter, P. Wlodarczyk, P. Wcislo, M. Pospelov, J.R. Smith, J. Read, W. Gawlik, et al., Ann. Phys. 525 (2013) 659.
[38] D. Budker, P.W. Graham, M. Ledbetter, S. Rajendran, A.O. Sushkov, Phys. Rev. X. 4 (2014) 021030.
[39] V.V. Stadnik, Y.V. Flambaum, Phys. Rev. Lett. 114 (2015) 161301, https://link.aps.org/doi/10.1103/PhysRevLett.114.161301.
[40] V.V. Stadnik, Y.V. Flambaum, Phys. Rev. A. 93 (2016) 063630, https://link.aps.org/doi/10.1103/PhysRevA.93.063630.
[41] B.M. Roberts, G. Blewitt, C. Dailey, M. Murphy, M. Pospelov, A. Rollings, J. Sherman, W. Williams, A. Derevianko, Nat. Commun. 8 (2017) 1195, http://dx.doi.org/10.1038/s41467-017-01440-4.
[42] A. Pierce, K. Riles, Y. Zhao, Phys. Rev. Lett. 121 (2018) 061102.
[43] A. Garcon, J.W. Blanchard, G.P. Centers, N.L. Figueroa, P.W. Graham, D.F. Jackson Kimball, S. Rajendran, A.O. Sushkov, Y.V. Stadnik, A. Wickenbrock, et al., Sci. Adv. 5 (2019).
[44] Y.V. Grote, H. Stadnik, Phys. Rev. Res. 1 (2019) 033187, https://link.aps.org/doi/10.1103/PhysRevResearch.1.033187.
[45] T. McNally, R.L. Zelevinsky, Eur. Phys. J. D 74 (2020) 61.
[46] N.L. Figueroa, D. Budker, E.M. Rasel, Quantum Sci. Technol. 6 (2021) 034004.
[47] D.F. Jackson Kimball, D. Budker, J. Eby, M. Pospelov, S. Pustelny, T. Scholtes, Y.V. Stadnik, A. Weis, A. Wickenbrock, Phys. Rev. D. 97 (2018) 043002, https://link.aps.org/doi/10.1103/PhysRevD.97.043002.
[48] S. Afach, D. Budker, G. DeCamp, V. Dumont, Z. Grujić, H. Guo, D.J. Kimball, T. Kornack, V. Lebedev, W. Li, et al., Phys. Dark Univ. (ISSN: 2212-6864) 22 (2018) 162, http://www.sciencedirect.com/science/article/pii/S2212686418301031.
[49] H. Masia-Roig, J.A. Smiga, D. Budker, V. Dumont, Z. Grujic, D. Kim, D.F.J. Kimball, V. Lebedev, M. Monroy, S. Pustelny, et al., Phys. Dark Univ. (ISSN: 2212-6864) 28 (2020) 100494, http://www.sciencedirect.com/science/article/pii/S2212686419303760.
[50] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[51] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Rev. Modern Phys. 91 (2019) 045002, https://link.aps.org/doi/10.1103/RevModPhys.91.045002.
[52] N. Qian, Neural Netw. 12 (1999) 145.
[53] J. Duchi, E. Hazan, Y. Singer, J. Mach. Learn. Res. 12 (2011) 2121.
[54] T. Tieleman, G. Hinton, et al., COURSERA Neural Netw. Mach. Learn. 4 (2012) 26.
[55] M.D. Zeiler, Adadelta: an adaptive learning rate method, 2012.
[56] Kingma J., D.P. Ba, arXiv e-prints. (2014), 1412.6980.
[57] I.I. Kolb, E.W. Tkachev, Phys. Rev. Lett. 71 (1993) 3051, https://link.aps.org/doi/10.1103/PhysRevLett.71.3051.
[58] E. Braaten, A. Mohapatra, H. Zhang, Phys. Rev. Lett. 117 (2016) 121801.
[59] A. Arvanitaki, S. Dimopoulos, M. Galanis, L. Lehner, J.O. Thompson, K. Van Tilburg, Phys. Rev. D. 101 (2020) 083014.
[60] C. Dailey, C. Bradley, D.F. Jackson Kimball, I.A. Sulai, S. Pustelny, A. Wickenbrock, A. Derevianko, Nat. Astron. 5 (2021) 150.
[61] S. Afach, B.C. Buchler, D. Budker, C. Dailey, A. Derevianko, V. Dumont, N.L. Figueroa, I. Gerhardt, Z.D. Grujić, H. Guo, et al., Search for topological defect dark matter using the global network of optical magnetometers for exotic physics searches (gnome), 2021, 2102.13379.
[62] D.F. Jackson Kimball, J. Dudley, Y. Li, S. Thulasi, S. Pustelny, D. Budker, M. Zolotorev, Phys. Rev. D. 94 (2016) 082005, https://link.aps.org/doi/10.1103/PhysRevD.94.082005.
[63] D.F.J. Kimball, New J. Phys. 17 (2015) 073008.
[64] P.R. Kafle, S. Sharma, G.F. Lewis, J. Bland-Hawthorn, Astrophys. J. 794 (2014) 59, http://dx.doi.org/10.1088/0004-637x/794/1/59.
[65] M.S. Turner, Phys. Rev. D. 42 (1990) 3572, https://link.aps.org/doi/10.1103/PhysRevD.42.3572.
[66] G.N. Wilson, E.B. Lewis, Proc. Am. Acad. Arts Sci. (ISSN: 01999818) 48 (1912) 389, http://www.jstor.org/stable/20022840.
[67] F. Yates, Suppl. J. R. Stat. Soc. 1 (1934) 217, https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2983604.
[68] R.G. Newcombe, Stat. Med. 17 (1998) 857, http://dx.doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E.
[69] S. Wallis, J. Quant. Linguist. 20 (2013) 178, http://dx.doi.org/10.1080/09296174.2013.799918.
[70] E. Beck, J. Shultz, Arch. Pathol. Lab. Med. 110 (1986) 13.