Laboratory Constraints on the Neutron-Spin Coupling of feV-Scale Axions

Junyi Lee[*]

*Institute of Materials Research and Engineering (IMRE), Agency for Science,
Technology and Research (A*STAR), 2 Fusionopolis Way, Innovis #08-03,
Singapore 138634, Republic of Singapore*

Mariangela Lisanti

*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA
and Center for Computational Astrophysics, Flatiron Institute, New York, New York 10010, USA*

William A. Terrano

Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

Michael Romalis

Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

Ultralight axionlike particles can contribute to the dark matter near the Sun, leading to a distinct, stochastic signature in terrestrial experiments. We search for such particles through their neutron-spin coupling by reanalyzing approximately 40 days of data from a K-^{3}He comagnetometer with a new frequency-domain likelihood-based formalism that properly accounts for stochastic effects over all axion coherence times relative to the experimental time span. Assuming that axions make up all of the dark matter in the Sun's vicinity, we find a median 95% upper limit on the neutron-spin coupling of 2.4×10^{-10} GeV^{-1} for most axion masses from 0.4 to 4 feV, which is about 5 orders of magnitude more stringent than previous laboratory bounds in that mass range. Although several peaks in the experiment's magnetic power spectrum suggest the rejection of a white-noise null hypothesis, further analysis of their line shapes yields no positive evidence for a dark-matter axion.

Subject Areas: Astrophysics,
Atomic and Molecular Physics,
Cosmology

I. INTRODUCTION

Although the existence of dark matter is supported by astrophysical and cosmological observations over many length scales [1–4], its fundamental nature remains a mystery. Ultralight pseudo-scalar particles that are the pseudo-Nambu-Goldstone bosons of a spontaneously broken global symmetry arise generically in many theories beyond the standard model [5–8] and can potentially contribute to the local dark matter. These "axions" can couple to the fermionic sector and lead to a neutron-spin coupling. In this paper, we reanalyze data from a Princeton-based K-^{3}He comagnetometer experiment [9] to obtain leading laboratory constraints on the neutron-spin coupling of feV-scale axions.

One well-motivated example of an ultralight pseudo-scalar particle is the QCD axion. It was originally proposed to solve the strong *CP* problem within QCD [10–13], but is also an attractive dark-matter candidate [14–19]. While the particle mass and coupling of the QCD axion are fundamentally related, this is not the case for more generic axions where the coupling strength and mass are independent of each other. Indeed, axions generically couple to neutrons with the Lagrangian

$$\mathcal{L}_{\text{int}} = g_{\text{aNN}} \bar{N} \gamma^\mu \gamma_5 N \partial_\mu a, \tag{1}$$

where g_{aNN} parametrizes the coupling strength, and a, N refer to the axion and neutron, respectively. The interaction in Eq. (1) mediates a spin-dependent long-range force, which has been searched for by Refs. [9,20–23]. However, this signal is suppressed by two factors of the (small) coupling constant g_{aNN}.

An alternate detection strategy is to search for dark-matter axions that couple to the nuclear spin in a detector target [24–28]. In this case, the interaction Hamiltonian,

[*]lee_jun_yi@imre.a-star.edu.sg

which follows from the nonrelativistic limit of Eq. (1), is suppressed only by one factor of g_{aNN}:

$$H_{int} = g_{aNN} \boldsymbol{\nabla} a \cdot \boldsymbol{\sigma}_N, \qquad (2)$$

where $\boldsymbol{\sigma}_N$ is the neutron spin. Because light dark-matter axions near the Sun are sufficiently numerous to be treated as an oscillating classical field, interference effects between different axion waves can lead to stochastic signals in terrestrial experiments [29–31], which must be accounted for to properly recover or constrain a signal. This effect was first discussed by Refs. [29,30] for the axion-photon coupling that many haloscope experiments are based on [32–36], but was only recently addressed for the axion-nuclear-spin coupling [31]. For the latter case, the signal is proportional to the gradient of the axion field, and it also depends on the direction of $\boldsymbol{\sigma}_N$, which for a terrestrial experiment will oscillate with a period of a sidereal day as Earth rotates about its axis.

Experimentally, the interaction in Eq. (2) should lead to a measurable spin-dependent energy shift in the valence neutron of ^{3}He that rotates the macroscopic magnetization of spin-polarized ^{3}He atoms. In a K-^{3}He comagnetometer, this spin polarization is created along a circularly polarized pump beam, and the magnetization of the atoms is probed by a separate orthogonal probe beam. In this paper, we report novel laboratory constraints of the axion's neutron-spin coupling based on analyzing approximately 40 days of raw magnetization orientation data in a K-^{3}He comagnetometer that was originally used in an exotic long-range force experiment [9].

Our analysis fully accounts for the stochastic nature of dark-matter axions using a new frequency-domain likelihood-based formalism, which is complementary to the time-binned methodology introduced in the theoretical counterpart to this work [31]. The likelihood formalism in Ref. [31] properly accounts for all two-point correlations of the axion field in the time-domain data. As explicitly demonstrated in that work, ignoring these correlations can result in failure to discover a signal and/or result in unreliable upper limits. We explicitly show in Appendix A that both the frequency and time-binned approaches yield similar results when applied on the same subset of simulated and experimental data. The analysis in this paper, like that in Ref. [31], works for all axion coherence times independent of the experimental time span (including in the regime when the experimental measurement time is shorter or comparable to the axion coherence time), which is a feature not achievable in previous frameworks [37]. Crucially, given a measured experimental spectrum, our formalism sets limits on (or recovers, in the case of a discovery) g_{aNN} in a way that fully takes into account the stochasticity of the dark-matter axion signal. This is in contrast with a recent work that studied the stochastic axion line shape without providing a methodology for correctly

extracting g_{aNN} from an experimental measurement [38]. We rigorously validate our analysis by applying it to a Monte Carlo ensemble of stochastic axion signals to ensure that it indeed recovers or sets the correct limits on g_{aNN} (see Appendix A). There is therefore no need to apply any of the corrections in Ref. [37] to our limits.

This paper should be contrasted with a recent analysis [28,39] that used the published power spectra of the Princeton-based comagnetometer (not the raw underlying data) without properly treating the signal's stochasticity, compromising its ability to reliably recover or set upper limits. Our analysis of the raw data also allows us to examine and subsequently reject peaks in the spectrum that may be plausible axion candidates.

Assuming that the local dark-matter density is wholly due to axions, our median 95% upper limit on the axion's neutron-spin coupling is $g_{aNN} < 2.4 \times 10^{-10}$ GeV^{-1} for axion masses from 0.4 to 4 feV. This is about 5 orders of magnitude stronger than previous laboratory bounds in the same mass range. For example, the original long-range force experiment in Ref. [9] reported a 95% upper limit of $g_{aNN} < 9.1 \times 10^{-5}$ GeV^{-1}, and a recent terrestrial experiment searching for dark-matter axions in a similar mass range reported a tightest 90% upper limit of 4.1×10^{-5} GeV^{-1} [26]. Close to the largest axion mass we analyze (41 feV), another experiment recently reported a more comparable 95% upper limit of 3.2×10^{-9} GeV^{-1} at 53 feV [27], although we note that for the quoted limit the analysis there has significant frequency gaps (compared to the axion's linewidth) and also does not account for the axion's stochasticity. At the largest mass we analyze, the NASDUCK Collaboration recently reported a 95% upper limit of 10^{-6} GeV^{-1} [28], which is about 2 orders of magnitude weaker than our results. Our constraints are comparable with limits derived from stellar cooling arguments, although the latter limits are subject to significant astrophysical uncertainties. Our results therefore serve as useful complementary bounds.

The magnetic power spectrum of the original experiment in Ref. [9] was optimized only for a narrow bandwidth about 0.17 Hz, and peaks outside that narrow bandwidth were not seriously investigated. Our likelihood analysis of those peaks, which tests an axion signal hypothesis against a null white background hypothesis, yields a few peaks with significance above 5σ (after taking into account the look-elsewhere effect) that persist over the entire data-taking duration. Although these results suggest that the white background model should be rejected for these peaks, accepting an axion signal hypothesis should require additional tests. Comparison with the expected line shapes from Monte Carlo simulations suggest a nonaxion origin for all these peaks.

This paper is structured as follows. In Sec. II, we review features of the frequency-domain likelihood analysis, leaving detailed derivations to Appendix B. We then

describe the original experiment of Ref. [9] and its pertinent features to this work in Sec. III, before discussing the results of the analysis in Secs. IV–V and concluding in Sec. VI. We also include two additional appendixes. In Appendix A, we present the validation of our frequency-domain likelihood analysis with the method of Ref. [31], and in Appendix C, we motivate the frequency resolution at which we test for axion candidates.

II. FREQUENCY-DOMAIN ANALYSIS

A. Axion signal modeling

Weakly interacting astrophysical axions confined within a volume V can be treated as a free scalar field

$$a(x) = \sum_{\mathbf{p}} \frac{1}{\sqrt{2V\omega_{\mathbf{p}}}} (b_{\mathbf{p}} e^{-ip\cdot x} + b_{\mathbf{p}}^{\dagger} e^{ip\cdot x}), \quad (3)$$

where the sum is over all three-momenta $\mathbf{p}$, and $b_{\mathbf{p}}, b_{\mathbf{p}}^{\dagger}$ are the annihilation and creation operators of the mode with energy $\omega_{\mathbf{p}}$ and four-momentum p. If the local dark-matter density is due wholly to axions, then their occupancy numbers must be extremely large, and we may therefore take the classical limit where $b_{\mathbf{p}}^{\dagger} b_{\mathbf{p}} = N_{\mathbf{p}}$ is the mean occupation number of the mode with momentum $\mathbf{p}$. Doing so, we obtain the classical axion field

$$a(x) = \sum_{\mathbf{p}} \sqrt{\frac{2N_{\mathbf{p}}}{V\omega_{\mathbf{p}}}} \cos(p^0 t - \mathbf{p} \cdot \mathbf{x} + \phi_{\mathbf{p}}), \quad (4)$$

where $\phi_{\mathbf{p}}$ is an unknown phase between 0 and 2π that can be modeled as uniformly distributed over that interval. The sum over all modes $\mathbf{p}$ with distinct energies $\omega_{\mathbf{p}}$ leads to interference effects noticeable after a characteristic coherence time τ_a, which is dependent on the momentum distribution of the axions. For example, in the standard halo model, axions are assumed to follow a Maxwell-Boltzmann velocity distribution in the Galactic rest frame,

$$f(\mathbf{v})d^3v = \frac{1}{(2\pi\sigma_v^2)^{3/2}} \exp\left(-\frac{(\mathbf{v} + \mathbf{v}_E)^2}{2\sigma_v^2}\right) d^3v, \quad (5)$$

where $\mathbf{v}$ is the axion velocity in the laboratory frame, $\mathbf{v}_E$ is Earth's velocity [40] so that $\mathbf{v} + \mathbf{v}_E$ is the axion velocity in the Galactic rest frame, and $\sqrt{2}\sigma_v \sim 220$ km/s is the distribution's modal speed, which is typically taken to be the Sun's circular velocity in the Galactic rest frame [41–43]. In this case, the characteristic coherence time for an axion with mass m_a may be defined as $\tau_a = 1/(\sigma_v^2 m_a)$ [30].

The stochastic classical axion field in Eq. (4) naturally translates into a stochastic axion gradient, which at the location of the laboratory (which we define as $\mathbf{x} = 0$), can be written as [44]

$$\mathbf{\nabla}a(t) = \sum_{\mathbf{p}} \sqrt{\frac{2N_{\mathbf{p}}}{V\omega_{\mathbf{p}}}} \cos(p^0 t + \phi_{\mathbf{p}})\mathbf{p}. \quad (6)$$

Experimentally, the energy shift due to the interaction in Eq. (2) is measured as the energy shift in spin states of some basis, or equivalently, in spin states along some quantization axis. For the K-^{3}He comagnetometer introduced in Sec. I and illustrated in Fig. 1, we primarily measure the rotation of the atoms' magnetization in the plane of the pump-and-probe beams, which implies that we are mainly sensitive to energy shifts in states quantized along an axis orthogonal to both pump-and-probe beams. Defining this axis as $\hat{\mathbf{m}}$, the gradient of the axion field at discrete times $n\Delta t$ can be understood as a measurable stochastic anomalous magnetic field with a magnitude along $\hat{\mathbf{m}}$ that is given by

$$\beta_n = \frac{g_{\mathrm{aNN}}}{\mu} \mathbf{\nabla}a(n\Delta t) \cdot \hat{\mathbf{m}}(n\Delta t), \quad (7)$$
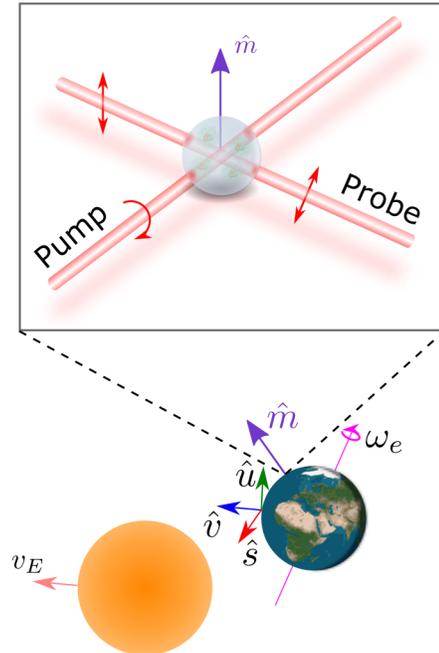

FIG. 1. Schematic showing the sensitive axis $\hat{\mathbf{m}}$ of the experiment as an outward normal from the surface of Earth that rotates at the sidereal frequency ω_e, leading to sidebands of the axion's power spectrum. The inset shows a schematic of the comagnetometer, which consists of spin-polarized atoms in a glass vapor cell that is optically pumped by resonant circularly polarized light and probed by off-resonant linearly polarized light in the horizontal plane. As described in the main text, this implies that the sensitive axis $\hat{\mathbf{m}}$ is oriented vertically, and by convention, we take it to be the outward normal from the surface of Earth. It is convenient, when deriving the covariance matrix of the axion's frequency spectrum, to use an orthonormal coordinate system $\{\hat{s}, \hat{u}, \hat{v}\}$ such that $\hat{v}$ is parallel to Earth's velocity v_E (see Appendix B 1), which is dominated by the Sun's circular velocity.

where Δt is the sampling interval, n is an integer, and μ has units of a magnetic dipole moment so that the experimentally measured energy shift from the interaction in Eq. (2) at time $n\Delta t$ is $\Delta E = \beta_n \mu$.

As the authors in Ref. [31] pointed out, because the summation over **p** in Eq. (6) can involve a great many terms and each $\phi_\mathbf{p}$ is, for each distinct value of **p**, an independent and identically distributed random variable, the set $\{\nabla a(n\Delta t)|n = 0, \ldots, N - 1\}$ is a Gaussian process by the central limit theorem. In particular, given a time series $\{0, \Delta t, \ldots, (N - 1)\Delta t\}$, the vector $[\nabla a(0), \ldots, \nabla a((N - 1)\Delta t)]$, and by extension the vector $[\beta_0, \ldots, \beta_{N-1}]$ that is experimentally measured, follows a multivariate normal distribution with zero mean (but with a covariance matrix that is not necessarily diagonal) due to the uniformly distributed random phase $\phi_\mathbf{p}$. It is therefore possible to use the statistical properties therein to correctly recover or set upper limits on the axion's coupling strength based on a time-series measurement $[\beta_0, \ldots, \beta_{N-1}]$ of the magneticlike anomalous field β, as demonstrated in Ref. [31]. However, it is frequently useful to perform the analysis in the frequency domain when there are other sources of frequency-dependent noise present. Because the discrete Fourier transform of a time series is a unitary transformation and the normal distribution is closed under linear transformations, one might expect that the Fourier transform of a time series $[\beta_0, \ldots, \beta_{N-1}]$ is still normally distributed. As we show by construction in Appendix B, this is indeed true. In the limit where the experimental time span T is much larger than the coherence time τ_a of the axion, an analytical form of the frequency-domain covariance matrix exists which allows for very efficient computation. Moreover, there is no need here, as required in the time-binned analysis, to recompute the size of the time bins for different axion masses. For brevity, we leave the full derivation of the frequency-domain covariance matrix to Appendix B and discuss only various physical implications of the frequency-domain approach here.

In the nonrelativistic limit of Eq. (6), the oscillation frequency of each axion mode with momentum **p** is $p^0 = m_a + m_a v^2/2$, where $v = |\mathbf{p}|/m_a$ is the axion speed. Furthermore, the amplitude of each mode is proportional to the mean occupation number of the mode $N_\mathbf{p}$, which is in turn proportional to the axion's speed distribution $f(v)dv$. Consequently, the axion's frequency line shape is essentially its "speed spectrum," with each frequency containing contributions from all momentum **p** with the same speed. This has been studied in the context of the axion's photon coupling [29], but one of the most striking differences of the neutron-spin coupling is that it introduces an additional frequency modulation due to the projection of ∇a on $\hat{\mathbf{m}}$, the experiment's fixed quantization axis, which rotates at Earth's angular sidereal frequency ω_e as the experiment moves about Earth's axis.

The measured anomalous field $\beta \propto \nabla a \cdot \hat{\mathbf{m}}$ therefore has sidebands spaced $\pm\omega_e$ apart from the frequency of ∇a.

If, in addition, $\hat{\mathbf{m}}(t)$ also has a dc component, i.e., the experiment's quantization axis has a nonzero projection along Earth's axis, then the frequency spectrum of β will also have power proportional to ∇a at its original frequency. This modulation of the axion's frequency line shape complicates its speed spectrum since each measured frequency can now also contain contributions with different axion speeds due to sampling from the two additional modulated spectrums. More precisely, the axion speeds that can contribute at any given measured frequency ω_k are

$$\nu_{k,n} = \sqrt{\frac{2[\omega_k - (m_a + n\omega_e)]}{m_a}} \quad (8)$$

for $n \in \{-1, 0, 1\}$ and $\omega_k - (m_a + n\omega_e) \geq 0$. Accordingly, frequencies that are separated by 1 or $2\omega_e$ can contain contributions from axions with the same speeds, which means that there can be nonzero correlations between these frequencies since the same phase $\phi_\mathbf{p}$ can be sampled twice in this case. An example of this is shown in Fig. 2. We emphasize that in the limit of infinite-frequency resolution (i.e., the regime where $T \gg \tau_a$) where there is no spectral leakage between neighboring frequency points, only frequencies that are spaced $\pm\omega_e$ and $\pm 2\omega_e$ apart can be correlated. When this limit is not valid, there will be additional correlations between neighboring frequency
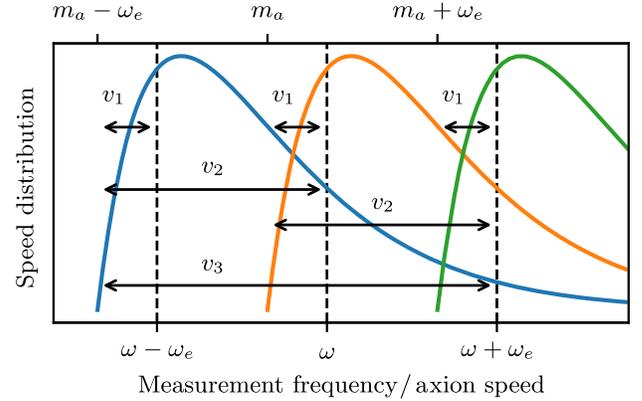

FIG. 2. Correlations between measurement frequencies ω and $\omega \pm \omega_e$. The original speed distribution (orange), which is zero for all frequencies below the axion mass m_a, is down (blue) and up (green) modulated by $\pm\omega_e$. Each frequency measurement (dashed lines) samples axions with varying speeds from both the original and modulated distributions (all denoted with double-headed horizontal arrows). For the measurement at $\omega - \omega_e$, only axions with speed v_1 can contribute, but axions with speeds v_1 and v_2 can contribute to the measurement at ω, while the measurement at $\omega + \omega_e$ has contributions from axions with speeds v_1, v_2, and v_3. $\omega - \omega_e$ is correlated with ω due to the contributions from axions with speed v_1, while ω is correlated with $\omega + \omega_e$ due to contributions from axions with speeds v_1 and v_2. Similarly, $\omega - \omega_e$ is correlated with $\omega + \omega_e$ because of axions with speed v_1.
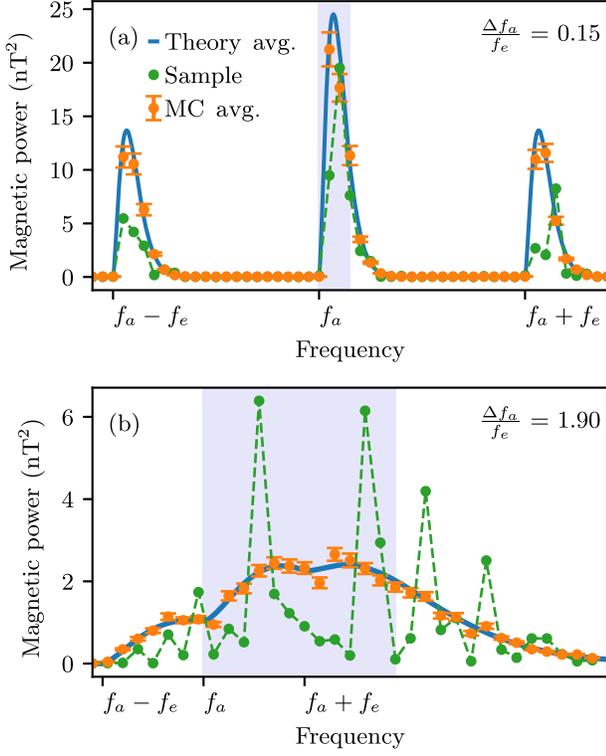
FIG. 3. Magnetic power spectrum for axion linewidths Δf_a where (a) $\Delta f_a/f_e < 1$ and (b) $\Delta f_a/f_e > 1$. $f_e = \omega_e/2\pi$ is Earth's sidereal rotation frequency. We assume in these plots that the experimental frequency resolution is sufficient to fully resolve the axion's line shape and that the experimental uncertainty at each point is negligible. The solid blue line is the mean theoretical power spectrum taken from Eq. (B52), while the orange points give the mean power from 100 Monte Carlo simulations. Orange error bars give the standard deviation of the mean power from the simulations. The green dots (with connecting lines to guide the eye) show an example of one particular realization of a stochastic axion power spectrum. The data here are generated for a coupling constant of $g_{aNN} = 1$ GeV^{-1}. Notice that for the same coupling constant, the peak axion power is lower for (b) compared to (a) due to the broadening of the axion peak for more massive axions. The shaded area denotes the frequency span $[f_a, f_a + \Delta f_a]$.

points, but this correlation is typically small compared to the correlation between frequencies spaced at intervals of ω_e.

The width of an axion's spectral line generally depends on its speed distribution. For example, in the standard halo model, the axion's coherence time τ_a is $\tau_a = 1/(\sigma_v^2 m_a)$ [30], and its linewidth Δf_a can be defined to be

$$\Delta f_a \equiv \frac{1}{\tau_a} = \sigma_v^2 m_a. \tag{9}$$

Physically, the axion's frequency linewidth is due to its kinetic energy, and consequently, more massive axions have larger linewidths because they have more kinetic

energy for the same speed. For less-massive axions where $\Delta f_a \ll f_e \equiv \omega_e/2\pi$, the modulation due to Earth's sidereal rotation splits the expected power spectrum from ∇a into three peaks. We illustrate this in Fig. 3(a) by plotting the expected magnetic power spectrum [Eq. (B52)] from an axion with linewidth $\Delta f_a = 0.15 f_e$ in blue. Orange points show the average power from 100 independent Monte Carlo simulations of axions with the same mass, while their error bars give the standard deviation of the mean power. Green dots (with dashed connecting lines to guide the eye) give a possible spectrum one might measure for an axion with the same mass. It should be emphasized that in Fig. 3 we assume a negligible experimental uncertainty, so the scatter of the green points is entirely due to the stochastic nature of the axion signal that one would observe in a particular experimental run. If the coherence time τ_a of the axion is sufficiently short so that many independent spectra may be measured during the experiment's data-taking run, the averaged power spectrum should (for negligible experimental noise) closely resemble the mean theoretical power spectrum (compare, for instance, the orange points with the blue theoretical line in Fig. 3). These conditions apply, for example, to many axion haloscopes [33,34,36] that search for axions with masses in the μeV range that have coherence times on the order of a fraction of a millisecond. However, for extremely light axions with very long coherence times, it may become impractical to obtain many independent spectra (as an example, the coherence time of a 0.1-feV axion is about 75 days long). In this case, because one particular measured spectrum might differ significantly from the analytical expected magnetic power (see Fig. 3 for an example), the correct g_{aNN} cannot be simply extracted by fitting Eq. (B52) to a measured spectrum. Rather, a more sophisticated analysis should be utilized to correctly extract g_{aNN} regardless of the stochastic variation of the measured spectrum.

At higher axion masses, $\Delta f_a \propto m_a$ increases and, as shown in Fig. 3(b), the three peaks recombine into a single peak when $\Delta f_a > f_e$. Another physical consequence of this broadening is that for the same coupling constant, the peak axion power is lower for a heavier axion than a lighter one since the total axion power must be conserved (for the same dark-matter density). This can be observed in Figs. 3(a) and 3(b), where both the heavier and lighter axion's spectrum are generated for a fixed coupling constant of g_{aNN}.

B. Likelihood procedure

The likelihood-based analysis used in this work applies to both the regime of low ($\Delta f_a < f_e$) axion mass depicted in Fig. 3(a), as well as in the high-mass regime of Fig. 3(b). We now briefly review the likelihood formalism [45] for recovering and/or setting upper limits on the coupling constant g_{aNN}. The measured experimental spectrum is $\{A_k, B_k | k = 0, ..., N - 1\}$, where

$$A_k = \frac{2}{N}\mathrm{Re}[\tilde{\beta}_k], \qquad B_k = -\frac{2}{N}\mathrm{Im}[\tilde{\beta}_k], \qquad (10)$$

and $\tilde{\beta}_k$ is the discrete Fourier transform of β_n from Eq. (7) given by

$$\tilde{\beta}_k = \sum_{n=0}^{N-1}\beta_n\mathrm{exp}^{-i\omega_k n\Delta t}, \qquad \omega_k \equiv \frac{2\pi k}{N\Delta t}. \qquad (11)$$

For a time series with gaps, A_k and B_k, as defined in Eq. (10), are determined by performing a linear least-squares fit of the time series to the form

$$A_k\cos\omega_k t + B_k\sin\omega_k t, \qquad (12)$$

where A_k, B_k here are the fit parameters. To compute the likelihood of obtaining a spectrum $\{A_k, B_k | k = 0, \ldots, N-1\}$, we need to have both a signal and background model. The spectrum for an axion with mass m_a can be shown to follow a multivariate normal distribution with zero mean and a nondiagonal covariance matrix $\Sigma_a(g_{\mathrm{aNN}}, m_a)$ that depends on g_{aNN} and m_a. We assume that the background is white with variance σ_b^2. Consequently, the measured signal is normally distributed with zero mean and variance $\Sigma = \Sigma_a + \sigma_b^2\mathbb{1}$, and the likelihood of measuring the spectrum $\mathbf{d} = \{A_k, B_k | k = 0, \ldots, N-1\}$ is therefore,

$$L(\mathbf{d}|g_{\mathrm{aNN}}, \sigma_b) = \frac{1}{\sqrt{(2\pi)^{2N}\det(\Sigma)}}\mathrm{exp}\left[-\frac{1}{2}\mathbf{d}^T\Sigma^{-1}\mathbf{d}\right]. \quad (13)$$

Although the experimental noise is frequency dependent, it is to a good approximation white over the bandwidth used for $\mathbf{d}$ (see Table I), except in certain localized cases that result in peaks in the recovered upper limits (see Fig. 5 and the discussion in Sec. IV).

To set a 95% upper limit, we define the test statistic

$$q(g) = \begin{cases} 2[\log L(\mathbf{d}|\hat{\mu}, \hat{\sigma}_b) - \max_{\sigma}\log L(\mathbf{d}|g, \sigma)], & \hat{\mu} \le g, \\ 0, & \hat{\mu} > g, \end{cases}$$
$$(14)$$

TABLE I. Bandwidth $\delta f = \delta f_{\mathrm{low}} + \delta f_{\mathrm{high}}$ spanning frequencies $[f_a - \delta f_{\mathrm{low}}, f_a + \delta f_{\mathrm{high}}]$ that is used to calculate the likelihood for an axion with mass f_a. Note that the bandwidth is asymmetrical about f_a since the axion power spectrum is itself asymmetrical due to the fact that the axion's kinetic energy can only be positive. $f_e \approx 11.6\ \mu\mathrm{Hz}$ is the frequency of Earth's rotation with respect to distant stars.

f_a (Hz)	δf_{low} (f_e)	δf_{high} (f_e)	δf (f_e)
(7, 10]	4	10	14
(5, 7]	4	8	12
(2, 5]	4	6	10
[0.01, 2]	4	4	8

where $\hat{\mu}$ and $\hat{\sigma}_b$ are the unconditional maximum likelihood estimators of g_{aNN} and σ_b, respectively,

$$\hat{\mu} = \arg\max_{\mu}\log L(\mathbf{d}|\mu, \sigma) \quad \forall\ \sigma,$$
$$\hat{\sigma}_b = \arg\max_{\sigma}\log L(\mathbf{d}|\mu, \sigma) \quad \forall\ \mu. \qquad (15)$$

By definition, $q(g) \ge 0$, with larger values of $q(g)$ indicating a greater probability that the hypothesized g is too large and is increasingly incompatible with the data. On the other hand, if the hypothesized g is smaller than the best-fit $\hat{\mu}$, we set $q(g) = 0$ because in finding an upper limit, we wish only to set g greater than $\hat{\mu}$, and hence, we set $q(g) = 0$ when $g < \hat{\mu}$ to indicate that the hypothesized g is not too large. We then find the upper limit by requiring that if the hypothesized g is indeed the true value, then the 95% upper limit g_{up} is such that 95% of the time, $q(g) \le q(g_{\mathrm{up}})$. The cumulative distribution function (CDF) of $q(g)$, assuming g is indeed the true value, can be asymptotically shown to be [45–47]

$$P(q(g) \le y) = \Phi(\sqrt{y}), \qquad (16)$$

where

$$\Phi(y) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{y}{\sqrt{2}}\right)\right] \qquad (17)$$

is the CDF of a standard normal distribution. Practically, the 95% upper limit is thus found by numerically solving for g_{up} such that $q(g_{\mathrm{up}}) \approx 2.7055$.

By definition, the test statistic $q(g)$ in Eq. (14) will always set an upper limit above the best-fit value $\hat{\mu}$. To obtain a double-sided confidence interval for a finite best-fit $\hat{\mu}$, we need to use a slightly different test statistic,

$$t(g) = 2[\log L(\mathbf{d}|\hat{\mu}, \hat{\sigma}_b) - \max_{\sigma}\log L(\mathbf{d}|g, \sigma)], \qquad (18)$$

which is not set to zero when $\hat{\mu} < g$. Rather, by definition, $t(g) \ge 0$ with larger values of $t(g)$ implying greater incompatibility of the hypothesized value of g with the best-fit $\hat{\mu}$ regardless of whether $g < \hat{\mu}$ or $g > \hat{\mu}$. Assuming that g is indeed the true value, the CDF of $t(g)$ is [45]

$$P(t(g) \le y) = 2\Phi(\sqrt{y}) - 1. \qquad (19)$$

Practically, the higher and lower end points of the confidence interval with confidence level CL may therefore be respectively obtained by solving for $g_{\mathrm{high}} > \hat{\mu}$ such that $P(t(g) < t(g_{\mathrm{high}})) = (1 - \mathrm{CL})/2$ and $g_{\mathrm{low}} < \hat{\mu}$ such that $P(t(g) < t(g_{\mathrm{low}})) = (1 - \mathrm{CL})/2$.

Similarly, to test the significance of the best-fit $\hat{\mu}$, we define the test statistic

$$q_0 = \begin{cases} 2[\log L(\mathbf{d}|\hat{\mu}, \hat{\sigma}_b) - \max_{\sigma} \log L(\mathbf{d}|0, \sigma)], & \hat{\mu} \geq 0, \\ 0, & \hat{\mu} < 0, \end{cases} \tag{20}$$

where we set $q_0 = 0$ when $\hat{\mu} < 0$ to restrict testing for $g_{aNN} \geq 0$ [48]. As before, larger values of q_0 indicate that the data are increasingly incompatible with the null hypothesis of there being no axion, which may be quantified by the asymptotic CDF of q_0 (assuming that the true value of g_{aNN} is indeed 0) [45–47],

$$P(q_0 \leq y) = \Phi(\sqrt{y}), \tag{21}$$

and the asymptotic probability density function for q_0,

$$f(q_0) = \frac{1}{2}\delta(q_0) + \frac{1}{2\sqrt{2\pi q_0}} e^{-q_0/2}, \tag{22}$$

where $\delta(q_0)$ is the Dirac delta function. The p value and significance Z of the best-fit $\hat{\mu}$ is thus simply

$$p = 1 - \Phi(\sqrt{q_0}) \quad \text{and} \quad Z = \sqrt{q_0}. \tag{23}$$

We now discuss some practical considerations when implementing the above formalism to actual experimental data. In the validation plots of Appendix A, we show that our analysis can correctly recover and set upper limits for g_{aNN} when testing for an axion at the correct mass (i.e., testing for an axion at its actual mass). However, since we do not know the actual mass of the axion, it is necessary to test for it over a large mass parameter space that can in principle span many decades. Given the very different axion line shapes for a light and heavy axion (see Fig. 3 for example), we expect that testing for an axion at a wrong mass (i.e., testing for an axion at a mass significantly different from its actual mass) should yield a wrong best-fit g_{aNN} and upper limit. However, it is not immediately obvious how small of a mass or frequency resolution one should use in testing for axions to ensure the recovery of a correct best-fit g_{aNN} from a real axion with *a priori* unknown mass. As we verify in Appendix C, testing with a spacing of approximately $\Delta f_a/2$ is sufficient to correctly recover or set upper limits on g_{aNN} from an axion signal within our mass range of interest.

The need to run the analysis over a large mass parameter space also complicates the interpretation of the significance in recovering a particular best-fit $\hat{\mu}$ with a corresponding discovery test statistic q_0. This is because the significance $Z = \sqrt{q_0}$ in Eq. (23) is only valid when testing at one mass and does not take into account the look-elsewhere effect when testing over a large (e.g., much larger than one) number of axion masses. To relate the global p value p_{global}, with a threshold in q_0, we observe that by definition, p_{global} is the probability that a white-noise background will

yield q_0 that is larger than some threshold q_{th} in any of the axion masses tested. For N independent tests, this is equivalent to the complement of the probability that all N tests yield $q_0 \leq q_{\text{th}}$:

$$p_{\text{global}} = 1 - (1 - p)^N. \tag{24}$$

For sufficiently small p so that $(1-p)^N \approx 1 - Np$, Eqs. (24) and (23) readily yield a relation between q_{th} and p_{global},

$$q_{\text{th}} = \left[\Phi^{-1}\left(1 - \frac{p_{\text{global}}}{N}\right)\right]^2. \tag{25}$$

However, as the authors in Ref. [29] noted, the number of independent tests N is frequently smaller than the total number of axion masses tested in the analysis since tests of closely separated axion masses rely on similar experimental spectra and are not independent. Physically, we expect axions that are separated by roughly a linewidth to be independent, with additional modifications due to the experiment's frequency resolution and the sidereal sidebands of the neutron-spin coupling. Consequently, given that the axion's linewidth is due to its kinetic energy, we expect that the axion masses

$$m_a^{(i)} = m_a^{(0)}(1 + \alpha v_0^2)^i \tag{26}$$

for $i = 0, \ldots, N-1$, to be independent, where $v_0 \sim 220$ km/s is the axion's modal speed [see Eq. (5) and the discussion there], and α is a parameter to be tuned via Monte Carlo simulations.

To tune α, we generate a large number of null Monte Carlo datasets and perform the likelihood analysis for axion masses between $f_{\min}$ and $f_{\max}$ using the same frequency spacing of $\Delta f_a/2$ that is used in the experimental analysis. We then extract $q_{\text{th}} = \max(q_0)$ from the analysis of each Monte Carlo dataset and obtain the empirical p_{global} from the distribution of q_{th} over all Monte Carlo datasets. For $N \gg 1$ and $\alpha v_0^2 \ll 1$ (with v_0 in $c = 1$ units), we obtain from Eq. (26)

$$N \approx \frac{1}{\alpha v_0^2} \log\left(\frac{f_{\max}}{f_{\min}}\right). \tag{27}$$

The tuned value of α can then be obtained by substituting Eq. (27) into Eq. (25) and fitting it to the q_{th} and p_{global} obtained from Monte Carlo simulations. We show an example of such a fit for axion masses centered at 0.562 Hz in the inset of Fig. 4 where the solid blue line shows p_{global} as a function of q_{th} as obtained from 100 000 null Monte Carlo datasets, and the dashed orange line is the fit of Eq. (25) to it.

For lack of computational resources, $[f_{\min}, f_{\max}]$ is normally such that $[f_{\min}, f_{\max}] \subset \mathcal{F}$, where $\mathcal{F}$ is the set

of all axion masses tested in the experimental analysis. The point, however, is that once α has been tuned, the number of independent axions $N_{\mathcal{F}}$ in $\mathcal{F}$ can then be obtained using Eq. (27) without resorting to Monte Carlo simulations.

As we discuss above and in Appendix C, we test axions with a spacing of $\Delta f_a/2$ to ensure that we can correctly recover or set upper limits on g_{aNN} from an axion signal in our mass range of interest. However, since $\Delta f_a \propto f_a$ from Eq. (9), the number of independent axions decreases at lower frequencies since the frequency spacing $\Delta f_a/2$ decreases at low frequencies but the experimental frequency resolution remains constant, which results in a significant overlap of the experimental spectra used in the likelihood analysis at each axion mass. $N_{\mathcal{F}}$ should therefore be obtained via an integration of Eq. (27) in logarithmic space

$$N_{\mathcal{F}} = \int_{\log \ \min(\mathcal{F})}^{\log \ \max(\mathcal{F})} \frac{1}{\alpha(\nu) v_0^2} \, d\nu. \qquad (28)$$

To obtain $N_{\mathcal{F}}$, we thus first perform the Monte Carlo procedure above with $f_{\max}/f_{\min} \approx 1.0008$ centered at 11 different frequencies between 0.01 and 10 Hz to obtain $\alpha(\nu)$ at 11 discrete points, which is shown as blue circles in Fig. 4. We then approximate $\alpha(\nu)$ from 0.01 to 10 Hz using an interpolating function (orange line in Fig. 4) and compute $N_{\mathcal{F}}$ through Eq. (28). For the frequency grid used in our experimental analysis, a q_0 of 30.1, 35.6, and 52.1 is estimated by substituting $N_{\mathcal{F}}$ into Eq. (25) to correspond to a one-sided global significance of 2σ, 3σ, and 5σ, respectively.

We conclude this section by noting that while the frequency bandwidth used in the calculation of the likelihood in Eq. (13) at each axion mass must be at least as large as the axion linewidth (including its up and down modulation of $\pm f_e$), its exact value is arbitrary. In general, one desires a bandwidth that is significantly wider than the
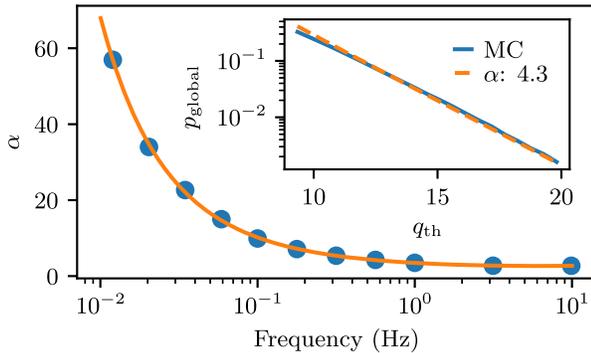

FIG. 4.   Main figure: blue circles give α as computed for that particular frequency. The orange line shows the interpolating function for $\alpha(\nu)$ obtained using the blue circles. Inset: solid blue line is obtained from 100 000 null Monte Carlo datasets for axion masses centered at 0.562 Hz, while the dashed orange line shows the fit of Eq. (27) to the blue line.

axion linewidth for proper estimation of the background noise, but is still sufficiently narrow such that the frequency-dependent experimental noise is nevertheless approximately white within the chosen bandwidth. Table I shows the bandwidth we use in our analysis.

III. EXPERIMENTAL SETUP

The data used in this paper come from an experiment that was originally designed to search for exotic long-range nuclear-spin-dependent forces [9], and the interested reader may find more details of the setup in Ref. [49]. In the following, we briefly review the experimental setup of that work and highlight aspects of it that pertain to this study.

At the heart of the experimental setup in Ref. [9] is a K-^{3}He comagnetometer. Although atomic magnetometers can also be used to measure anomalous magneticlike fields from an axion's gradient, they are susceptible to ordinary magnetic noise, which limits their usefulness in searching for weak anomalous fields. On the other hand, a K-^{3}He comagnetometer, which consists of spin-polarized K and ^{3}He atoms colocated within the same glass cell, can be operated in a way that makes it sensitive to anomalous fields while having a reduced response to ordinary magnetic fields. For the experiment under consideration here, a 2.4-cm-diameter spherical aluminosilicate glass cell with 12 amagats of ^{3}He, 46 torr of N_2 (for quenching excited K atoms during optical pumping), and a droplet of K metal is used. To produce a dense alkali vapor, the cell is placed in a fiberglass oven and electrically heated up to 160 °C using approximately 200-kHz ac currents that are well above the comagnetometer's bandwidth. The K is spin polarized by a circularly polarized pump beam from an amplified distributed feedback laser operating at K's $\mathcal{D}_1$ line, and ^{3}He is eventually spin polarized through alkali–noble-gas spin-exchange collisions with K. A 0.8-mW linearly polarized probe beam blue detuned 237 GHz from K's $\mathcal{D}_1$ line is directed through the cell perpendicular to the pump beam, and paramagnetic Faraday rotation of its polarization is used to measure the comagnetometer's signal, which consists of the projection of K spins along the probe beam. To reduce noise in the probe beam measurement, the probe beam's polarization is first modulated at 50 kHz using a photoelastic modulator together with a quarter-wave plate and is later demodulated by a lock-in amplifier after detection.

In the absence of a magneticlike field perpendicular to both the pump-and-probe beams, the K's magnetization is aligned with the pump beam with zero projection on the probe beam. However, in the presence of such a field, the K's magnetization experiences a torque and rotates into the probe beam, thereby creating an optically detectable signal. Equivalently, a magneticlike field along the axis that is perpendicular to both the pump-and-probe beams results in an energy shift for basis spin states quantized along that

sensitive axis [corresponding to $\hat{\mathbf{m}}$ in Eq. (7)], which results in a rotation of the spin ensemble into the probe beam. The magnitude of this rotation is proportional to the coherence time of the K spins. In the presence of Earth's magnetic field, this coherence time is limited by decoherence from K-K spin-exchange collisions. However, at sufficiently low magnetic fields and sufficiently high alkali density, this decoherence can be greatly suppressed [50,51] and significantly improved sensitivity to magneticlike fields can be achieved. To operate in this low-field regime, and to provide for magnetic shielding from the environment, the cell is placed in five layers of μ-metal shielding with a shielding factor of approximately 10^6. This allows the K to achieve high sensitivity to both anomalous and ordinary magnetic fields.

To suppress the comagnetometer's sensitivity to ordinary magnetic fields while retaining full sensitivity toward anomalous fields, a bias magnetic field opposite to the sum of the effective magnetic fields from the spin-polarized ^{3}He and K atoms is applied [52]. Because each spin species in a spherical cell experiences only a magnetic field equal to the sum of the applied bias field and the effective magnetic field of the other spin species (which is antiparallel to the applied bias field), each spin species experiences only a magnetic field that is equal in magnitude to its own effective magnetic field. Consequently, although the gyromagnetic ratio of ^{3}He is much smaller than K due to the much larger mass of the neutron, the resonance frequencies of ^{3}He and K are approximately equal in this regime due to the very different magnetic field that each species experiences. In this regime where the resonance frequencies of both spin species are well matched, both spin ensembles exhibit highly coupled and damped evolution [52]. In particular, the ^{3}He magnetization can be shown to adiabatically cancel out small changes in the magnetic field that K experiences so that the comagnetometer's signal, which is proportional to the projection of the K spins along the probe beam, is to first order insensitive to ordinary magnetic fields. On the other hand, a slight perturbation of the ^{3}He spins due to interactions with a neutron-coupling anomalous field results in a rotation of the ^{3}He magnetization that the K experiences, which results in a corresponding perturbation of the K spin along the probe beam that is optically detected. Meanwhile, an electron-coupling anomalous field couples only to the K spin. This rotates the K's magnetization which is then optically detected so that the K-^{3}He comagnetometer is, in the final analysis, sensitive to the difference between neutron-coupling (β^n) and electron-coupling (β^e) anomalous fields [52].

Assuming that there is no accidental cancellation (i.e., assuming that the axion does not couple to both electrons and neutrons to produce anomalous fields of roughly the same magnitude), we can therefore set limits on neutron- and electron-coupling magneticlike anomalous fields independently. More precisely, given limits on the anomalous magneticlike field β_{lim}, we set $\beta^n < \beta_{\text{lim}}$ and $\beta^e < \beta_{\text{lim}}$, where $\beta^n \propto 0.87 g_{\text{aNN}}/\mu_{\text{He}}$ and $\beta^e \propto g_{\text{aee}}/\mu_B$ [compare with Eq. (7)]. $\mu_{\text{He}} \approx 2.148 \times 10^{-26}$ J/T here is the magnetic dipole moment of ^{3}He [53], while μ_B is the Bohr magneton. The factor of 0.87 arises from the fact that the neutron contributes only about 87% of the nuclear spin of ^{3}He [54,55]. From the above, we note that one can obtain limits on g_{aee} given limits on g_{aNN} since $g_{\text{aee}} < 0.87 \mu_B g_{\text{aNN}}/\mu_{\text{He}}$.

For continual suppression of the comagnetometer's sensitivity to ordinary magnetic fields, the bias magnetic field needs to be periodically adjusted due to slow drifts in the ^{3}He magnetization. This is typically performed by an automated routine that minimizes the comagnetometer's response to a modulated magnetic field along the sensitive axis [52]. Besides zeroing out the comagnetometer's response to ordinary magnetic fields, this routine is also used to calibrate the comagnetometer's sensitivity to anomalous fields [49,52] and is regularly run after every 200 s of continuous data taking. However, since no data can be taken during the zeroing routine, this leads to time gaps between each 200-s record of data that prevents us from obtaining the experimental frequency spectrum by directly performing a fast Fourier transform on the time-series data. Rather, we perform a linear least-squares fit of the form in Eq. (12) to obtain the experiment's frequency spectrum. Because of memory constraints, this is done by first performing the fits over bundles of data that are approximately 24 hours long and storing the best-fit coefficients as well as their fit covariance matrices (assuming a white-noise background) at each fit frequency and for each bundle. The final experimental spectrum is then obtained by computing a coherent weighted average of the best-fit coefficients of each frequency over all the bundles. Before performing our fits, we also filter the data with an appropriate bandpass filter and downsample the original data that are sampled at 200 Hz to a frequency that is at least 4 times larger than the fit frequency. In total, we fit for this analysis approximately 17×10^6 frequencies from 0.01 to 10 Hz with a frequency resolution of $0.57 \, \mu$Hz $\approx 1/(40 \text{ days})$.

The original experiment in Ref. [9] consisted of two main data-taking campaigns: one in the spring of 2008 and another in the summer of 2008, with a gap of approximately 50 days in between. Absolute time was recorded in the original experiment as fractional sidereal days since J2000.0 (defined as January 1, 2000, 12 p.m. Terrestrial Time), which allows for the orientation of the experiment's sensitive axis to be calculated in terms of Galactic coordinates. Given that the experiment's sensitive axis was oriented vertically throughout both data-taking campaigns, and the coordinates of the experiment at Princeton are approximately 40.35 °N, 74.65 °W [56], the experiment's sensitive axis in Galactic coordinates can be calculated from the definition of a (Greenwich mean) fractional sidereal day and performing a coordinate transformation between the equatorial and Galactic coordinate systems,

taking into account precession of Earth's axis since J2000.0 [57].

Lastly, we note that throughout the course of the original experiment, the direction of the bias field is periodically flipped as a check on systematic effects. Physically, this causes the K spins to rotate in the opposite direction under the influence of the same anomalous (or magnetic) field pointing along the sensitive axis. Experimentally, this is measured as a sign flip in the lock-in signal, and we therefore take these field reversals into account by multiplying the comagnetometer's calibration to anomalous fields with the appropriate signs.

IV. RESULTS

We analyze our experimental data for roughly 8×10^6 axion frequencies between 0.01 to 10 Hz with a spacing of $\Delta f_a/2$ using the likelihood procedure outlined in Sec. II. At each axion frequency, a slice of the experimental frequency spectrum with a bandwidth δf (tabulated in Table I) is used to calculate the likelihood [Eq. (13)], which assumes a white-noise background. To verify the assumption of a white-noise background, we perform a Shapiro-Wilks test on the noise portion of the bandwidth δf where the axion's power is expected to be negligible. As can be observed in Fig. 3, this portion can be defined to be frequencies in δf that are not within $+2\Delta f_a$ of the axion

frequency f_a and its sidebands $f_a \pm f_e$. Based on the p values of the Shapiro-Wilks test, we exclude about 1% of the approximately 8×10^6 axion frequencies we test at a significance of 0.0013, corresponding to a one-sided significance of 3σ. If the noise in our data is completely Gaussian, we expect a rejection rate of only about 0.1%; the actual higher rejection rate of 1% indicates the presence of some non-Gaussian noise in the data.

Since the applicability of our white-noise model for frequencies that do not survive the Shapiro-Wilks test is questionable, we do not set limits on g_{aNN} at these frequencies, and they constitute small gaps in our constraints that are not visually discernible in Fig. 5 that shows, in orange lines, the remaining 99% of frequency values at which we set 95% upper limits on g_{aNN}. None of the frequencies excluded from Fig. 5 due to the Shapiro-Wilks test have q_0 greater than 52.1 (corresponding to a global significance above 5σ if the noise model is actually valid), except for a few frequencies that are nearly an integer multiple of 1 Hz. These peaks feature prominently in Fig. 5, but are unlikely to be true axion signals for reasons that we further elaborate in Sec. V. To validate our analysis procedure, we also independently calculate in Appendix A 2 the 95% limits on a subset of experimental data using the time-domain method of Ref. [31]. Our comparison of both approaches in Fig. 13 indicates that the methods agree well with each other.
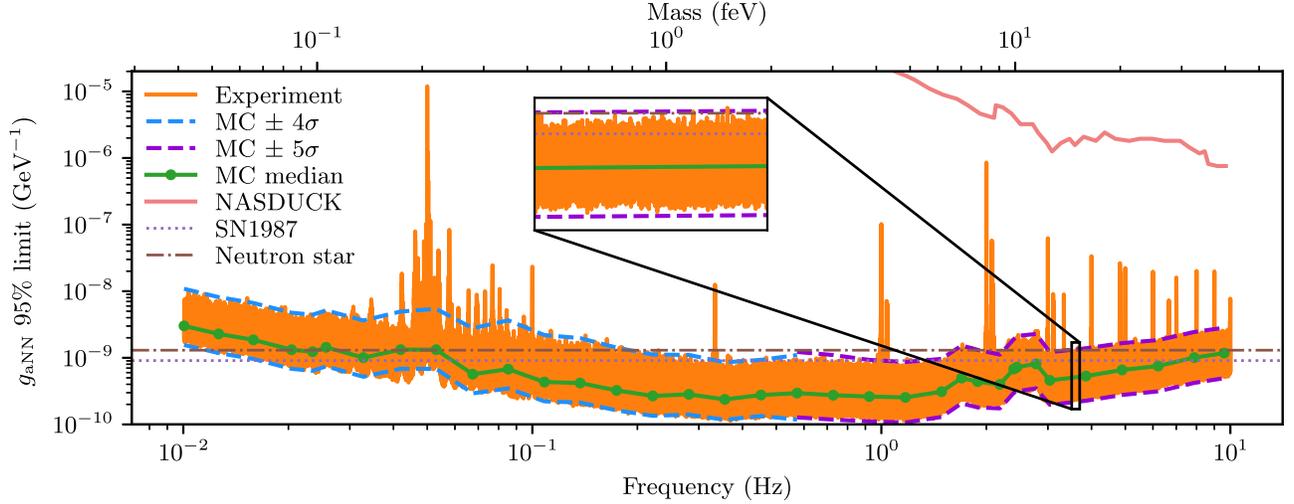

FIG. 5. Orange lines show the 95% upper limit on g_{aNN} obtained from our experimental data. The sensitivity of the experiment can be characterized by recovering the 95% upper limit over an ensemble of Monte Carlo data with no injected signal and plotting the median of those recovered limits (shown as green dots with connecting line to guide the eye). We achieve a median limit of 2.4×10^{-10} GeV^{-1} at 0.36 Hz, approximately 5 orders of magnitude stronger than previous laboratory bounds. Because of the large number of axion masses tested, there is a considerable spread in the recovered limits. This is, however, consistent with the $\pm 4\sigma$ and 5σ bands (illustrated by pairs of dashed blue and violet lines, respectively) obtained via Monte Carlo simulations. At higher axion frequencies, there is a greater density of axion masses on the log-scale plot, but as the magnifying inset shows, the recovered experimental limits are within the $\pm 5\sigma$ band as expected when viewed at the appropriate scale. The dotted SN1987 limit is a constraint from Ref. [58], which refines the usual one-pion-exchange approximation of the nucleon-nucleon bremsstrahlung process. The dash-dot neutron star limit is a 95% upper limit from Ref. [59] that analyzed cooling from five neutron stars. Limits from the NASDUCK Collaboration [28] are shown as a coral solid line.

Because of the large number of axion frequencies tested, a substantial statistical spread in the computed upper limits is to be expected. This expected statistical spread is illustrated by pairs of dashed blue and violet lines, which, respectively, give the $\pm 4\sigma$ and 5σ containment regions of 95% upper limits obtained by running our likelihood analysis on 10^7 Monte Carlo datasets at multiple frequencies. At each frequency point, the noise used in the Monte Carlo datasets is chosen to accurately reflect the experiment's noise spectrum. The green dots show the median 95% upper limit at those frequencies, and we achieve a low median limit of 2.4×10^{-10} GeV^{-1} at 0.36 Hz. Compared to previous laboratory results within the (0.4–40)-feV mass range, which come from the original long-range force experiment of this paper [9] and a more recent CASPEr-ZULF experiment that used low-field NMR techniques to search for astrophysical axions [26], our results represent an improvement of about 5 orders of magnitude. More importantly, our Monte-Carlo-validated experimental analysis correctly accounts for the stochastic nature of the interaction in Eq. (2), which is crucial for reliably recovering upper limits or best-fit values [31] but has heretofore not been carefully accounted for in most experimental analyses [25–27]. At higher frequencies between 1 and 10 Hz, our constraints are about 2 to 3 orders of magnitude stronger than what was recently reported by the NASDUCK Collaboration [28].

Our bounds surpass the upper limit from SN1987A [58], which is shown in Fig. 5 as a dotted line. This limit from SN1987A is, however, subject to significant uncertainties due to difficulties in correctly calculating the rate of axion production within the proto-neutron star, as well as complications arising from scattering and absorption within a dense plasma [58,60,61]. Moreover, if the neutrino emission of SN1987A came not from within the core of a proto-neutron star but from an accretion disk, axion production would not affect the neutrino emission, and in that case, constraints on the axion would be invalid [62]. Our constraints are also more stringent compared to the neutron star bound from Ref. [59] (shown as a brown dash-dot line in Fig. 5), which analyzed the cooling from five neutron stars. Like SN1987A, neutron star cooling arguments also have density-dependent coupling uncertainties [59,63]. In addition, magnetic field decays or other unknown heating mechanisms can plausibly lead to a relaxation of neutron star constraints [59,63]. It is worth noting that there is no universal consensus on the magnitude of the neutron star bound, and it can vary by about an order of magnitude depending on the details of the analysis [64]. Our results are therefore a useful complementary constraint to both the supernova and neutron star cooling limits.

One striking feature of Fig. 5 is the presence of multiple peaks in the recovered upper limits that are clearly above the background. This is due to the presence of various peaks in the experiment's power spectrum and can also be

seen in the distribution of the q_0 discovery test statistic, which we show as a blue line in Fig. 6, where we restrict the domain to $q_0 \geq 0.01$ for clarity. The expected distribution of q_0 [Eq. (22)] for the total number of axion masses we test is shown as a red line. At first glance, it might appear that there is an excess of q_0 for $q_0 > 10$ and an excessively long tail for $q_0 \gtrsim 50$. However, since not all of our tests are independent due to the finite experimental frequency resolution at low axion masses (see discussion at the end of Sec. II B), the number of independent excesses is actually smaller than what is suggested by Fig. 6. Moreover, the significance of obtaining a particular value of q_0 is no longer given by Eq. (23) due to the look-elsewhere effect. For example, a significance of 5σ corresponds to $q_0 = 52.1$ (see Sec. II B). It is also worth noting that the vast majority of axion masses we test have $q_0 < 0.01$ and are not shown in Fig. 6. Masses with $q_0 > 10$ comprise less than 0.5% of the total number of tests performed and there are in total only 63 masses with $q_0 > 52.1$.

Ordinarily, the presence of several masses with $q_0 > 52.1$ corresponding to a significance of more than 5σ after accounting for the look-elsewhere effect would warrant further experimental investigation. For example, the bias magnetic field of the comagnetometer could have been flipped to verify that the effect is not a purely electronic or optical effect, and orienting the experiment differently should, in principle, also give a different power spectrum from the axion that nevertheless returns the same best-fit g_{aNN} value from the likelihood analysis. However, since the
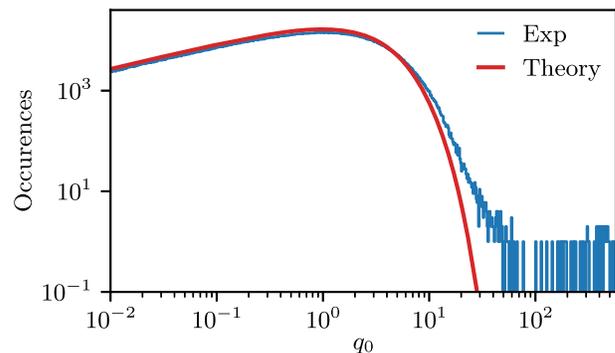

FIG. 6.   Distribution of the discovery test statistic is plotted in blue on a logarithmic scale from 10^{-2} onward. The red curve shows the expected distribution given the theoretical probability distribution function of q_0 for the total number of axion masses tested, assuming that each mass constitutes an independent test. As discussed in the main text, this is only approximately true. A vast majority of the axion masses tested have $q_0 < 0.01$ and are for clarity not shown. Masses with $q_0 > 10$ comprise less than 0.5% of all masses tested. We do not explicitly account for the look-elsewhere effect in this plot. As we note in the main text, a 5σ significance for rejecting the null hypothesis will correspond to a q_0 here of 52.1 after taking into account the look-elsewhere effect.

original experiment was not a dedicated search for astrophysical axions, and since we no longer have active control of the experiment, it is impossible for us to perform experimental checks to rule out spurious signals. Consequently, we are unable to make definitive claims about the origins of these peaks, but we discuss additional analyses below that suggest a nonaxion origin for them.

V. FURTHER ANALYSIS OF POSSIBLE AXION CANDIDATES

In this section, we further analyze possible axion candidates, which we here define as those axion masses with $q_0 > 52.1$. As discussed in Sec. III, data for the original experiment were taken over spring and summer of 2008, with a gap of approximately 50 days in between both campaigns during which systematic improvements particular to the original experiment were performed. A true axion signal should nevertheless persist in both datasets, and we expect that the recovered best-fit g_{aNN} from both analyses will not be significantly different from each other. This is true even though we use only an average Earth velocity $\langle \mathbf{v}_E \rangle$ in our analysis of both datasets instead of Earth's instantaneous velocity $\mathbf{v}_E$. As we show in Fig. 7(a), the expected discrepancy between the recovered best-fit g_{aNN} due to this approximation is negligible over the duration of our experiment. Consequently, any candidate that has significantly different best-fit g_{aNN} between the spring and summer datasets is likely to not be a true axion.

In Fig. 7(a), the central value of each data point gives the average best-fit g_{aNN} over 100 Monte Carlo datasets with a simulated data-taking interval of a week for a 6-Hz axion. Vertical lines indicate the 1σ containment interval of those best-fit values, which is a function of both the signal-to-noise ratio and the frequency resolution of the experiment relative to the axion linewidth. The latter contribution is due to the stochastic nature of the astrophysical axion signal where it is necessary to sample the axion signal over several coherence times to have a more reliable measure of g_{aNN}. Because of the directional nature of the neutron-spin coupling, our analysis is sensitive to both the magnitude and direction of $\mathbf{v}_E$. Since both of these quantities do not necessarily deviate symmetrically from their average values, this modulation of the best-fit g_{aNN} stemming from our use of $\langle \mathbf{v}_E \rangle$ rather than $\mathbf{v}_E$ is therefore also not always symmetric. This asymmetric modulation of the recovered best-fit g_{aNN} should be contrasted with the symmetric modulation of the expected raw axion power which we show in Fig. 7(b).

As an aside, it is interesting to note here that the annual peak-to-peak power modulation of thermalized axions in the standard halo model is only about 8% of its mean value, making this a relatively small effect that will be especially difficult to observe for low-frequency axions with coherence times on the order of a year.
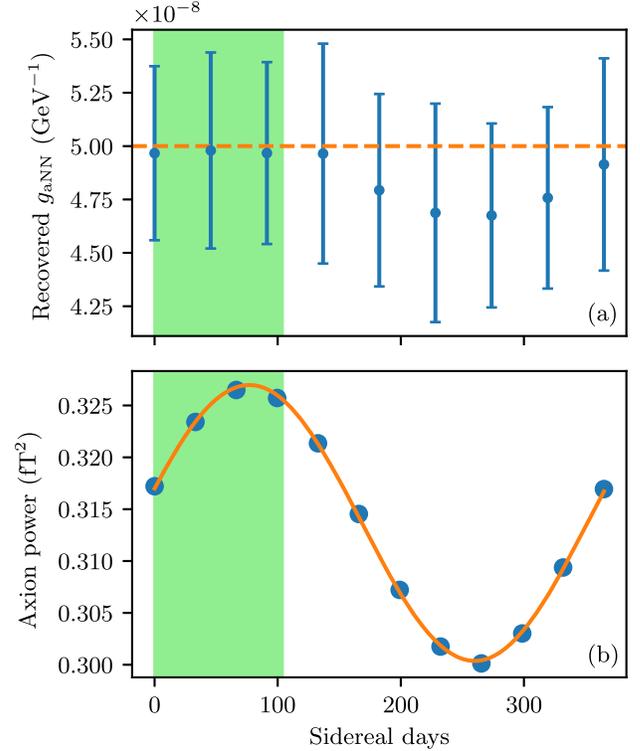

FIG. 7. (a) Expected annual modulation of the recovered best-fit g_{aNN} for a 6-Hz axion. This modulation of the best-fit g_{aNN} stems from deviation of both the magnitude and direction of $\mathbf{v}_E$ from $\langle \mathbf{v}_E \rangle$, where $\mathbf{v}_E$ is Earth's instantaneous velocity relative to the Galactic Center, and $\langle \mathbf{v}_E \rangle$ is its annual average that we use in our analysis. Because these deviations do not necessarily occur symmetrically, the annual modulation of the best-fit g_{aNN} is not always symmetric. This modulation of the recovered best-fit g_{aNN} should be contrasted with the expected annual modulation of the axion's power that we show in (b), where blue dots give the expected axion power over 12 months, and the orange line is a sinusoidal fit to the blue points. The dashed orange line in (a) indicates the injected g_{aNN}, which is larger than any of the recovered best-fit g_{aNN} in Fig. 8. The green shaded window indicates the time span of the experiment, which started data collection on 3022.63 sidereal days since J2000.0. As the results in (a) show, no significant deviation of the recovered g_{aNN} is expected over the time span of our experiment due to our use of $\langle \mathbf{v}_e \rangle$ rather than the instantaneous $\mathbf{v}_E$ in our analysis. Vertical lines in (a) give the 1σ containment interval for the best-fit g_{aNN} from 100 Monte Carlo datasets. The simulated data have a white-noise background of 1.6 fT/$\sqrt{\text{Hz}}$, which is comparable to the experiment's noise level.

In Fig. 7(a), we choose to simulate a 6-Hz axion because it is close in frequency to several of the axion candidates that we discuss below. Moreover, we inject the axion signal with a relatively large g_{aNN} of 5×10^{-8} GeV^{-1} [shown as a dashed orange line in Fig. 7(a)] and for a sufficiently long measurement time so that the 1σ containment interval will be comparable to the experimental confidence intervals for our best-fit g_{aNN}. Nevertheless, the expected modulation of

the recovered g_{aNN} in Fig. 7(a) is still negligible over the duration of our experiment, which is demarcated in the figure by the green shaded window. Consequently, we expect the 95% confidence intervals of the recovered best-fit g_{aNN} values from both the spring and summer datasets to overlap with each other if they are indeed due to a true axion signal.

In Fig. 8, we plot the best-fit g_{aNN} recovered from both the spring and summer datasets for all axion masses with $q_0 > 52.1$ in the full combined dataset. Vertical lines indicate their 95% confidence intervals obtained by analyzing the data in each season using the formalism in Sec. II B. Although there are numerous prominent peaks in Fig. 5 with frequencies below 2 Hz, their significance does not exceed 5σ because the width of their peaks is generally much broader than the expected axion linewidth at those frequencies. It therefore turns out that there are only 63 axion candidates above approximately 2 Hz that have

significance above 5σ after taking into account the look-elsewhere effect. Out of these 63 possible axion candidates with $q_0 > 52.1$, only 14 have overlapping 95% confidence intervals. We deem the other 49 axion candidates with nonoverlapping g_{aNN} confidence intervals as unlikely to be true axions due to inconsistencies from both the spring and summer datasets, and we do not perform any further analysis on them. For the remaining 14 axion candidates with overlapping 95% confidence intervals, we further perform a peak shape analysis on them that we describe below.

A unique feature of the signal line shape in the axion frequency range that we examine is the appearance of three distinct peaks spaced by Earth's sidereal rotation frequency. This allows us to discriminate between sidereal modulation of the axion signal and monochromatic signals that are likely due to terrestrial sources. Figure 9 shows an example of an experimental peak at 6.666 Hz where the likelihood
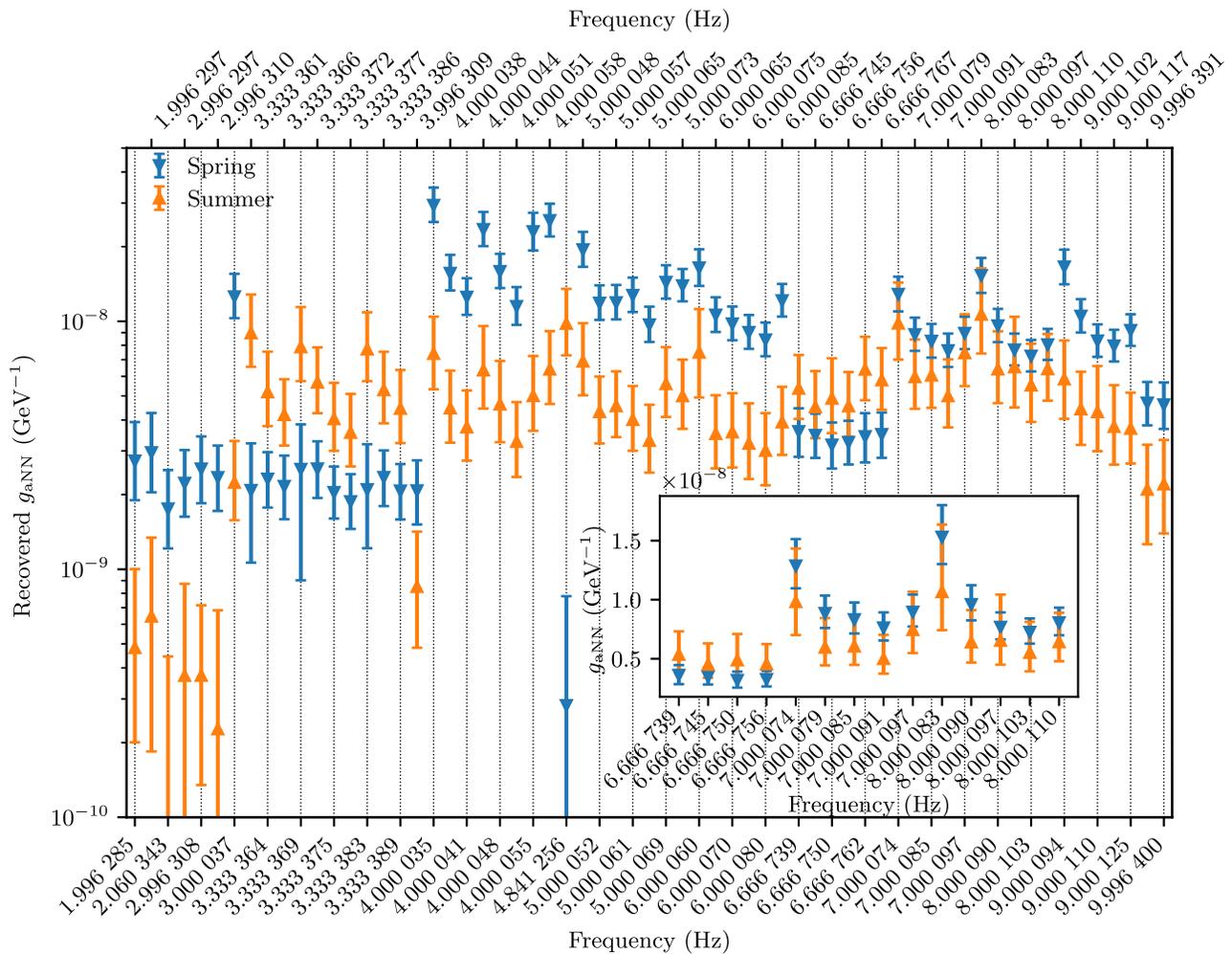

FIG. 8. Best-fit g_{aNN} recovered from both the spring and summer datasets for all axion masses with $q_0 > 52.1$ corresponding to a significance of more than 5σ after taking into account the look-elsewhere effect. Error bars show the 95% confidence intervals for the best-fit values. Of the 63 possible candidates shown here, only 14 have overlapping confidence intervals. These candidates are also shown in the inset.
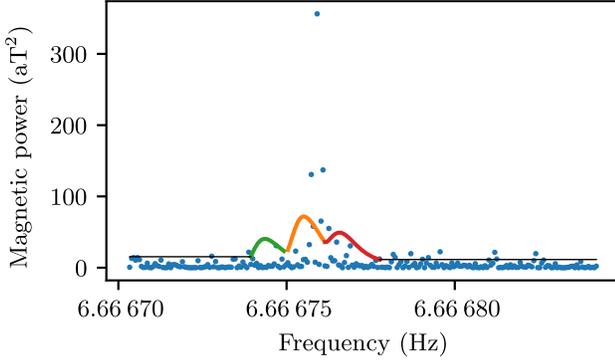
FIG. 9. Experimental magnetic power near 6.666 Hz (dots). The average magnetic power of a signal candidate identified by the likelihood analysis at 6.666 75 Hz is shown by the solid line. For the purposes of testing an axion candidate directly against the signal model (rather than against a white-noise null hypothesis as in our likelihood analysis), we split the signal power into three regions identified by colored lines to calculate the power under each peak. These peaks are due to sideband modulations of the axion's power due to Earth's rotation about its axis [see discussion around Eq. (8)]. The selection of these three regions is driven by a need to discriminate between true and false axion signals (see discussion in Sec. V of the text and Fig. 10).

analysis identifies several possible axion candidates with q_0 of around 200. However, because the likelihood analysis is effectively a hypothesis test for the axion model against the null hypothesis of a white-noise background, any relatively narrow deviation from a flat spectrum results in a high q_0 value favoring the axion hypothesis even if the line-shape match to a true axion signal is poor.

An obvious way to directly test an axion candidate against the signal model is to fit its line shape to the expected line shape of an axion. Nevertheless, fitting the line shape of the expected axion signal is in general difficult since it has large statistical fluctuations in each frequency bin. To reduce this uncertainty, one has to combine the power in many bins together, but the exact number of bins to combine depends on the axion linewidth, which is itself proportional to the axion frequency. In order to have a general line-shape analysis, we calculate the total power under each of the three expected axion peaks. When the peaks overlap, we separate the signal into three frequency bands as shown, for example, by the different colors in Fig. 9. We then calculate the ratio of the central peak power

to the average of the two side peak powers. From Monte Carlo simulations of axions in our mass range of interest, this ratio is generally expected to be slightly larger than 1 for a good axion candidate. Physically, this ratio depends on the projection of the experiment's sensitive axis [defined as $\hat{\mathbf{m}}$ in Eq. (7)] on the axis of Earth's rotation. An analytical result of this ratio can be obtained for the expected magnetic power spectrum [see Eq. (B52) and Table II], but there is some variability for a single measurement due to the stochasticity of dark-matter axions, which requires additional verification via Monte Carlo simulations. In Fig. 10, we plot a distribution of relative peak power ratios based on 1000 Monte Carlo simulations for a candidate signal at 6.666 75 Hz. We also show the experimental power ratios for the four signal candidates near 6.666 Hz that have overlapping g_{aNN} confidence intervals from both summer and spring datasets. One can see that all signal candidates are excluded with $>90\%$ probability. Similar analyses are performed for other candidate peaks near 7 and 8 Hz, and they are similarly inconsistent with the expected axion line shape.

Lastly, we note that the 6.666-Hz candidates are an integer multiple of the 3.333-Hz candidates which have nonoverlapping confidence intervals in Fig. 8, and that the 7- and 8-Hz candidates are integer multiples of a 1-Hz peak that is likely due to a clock signal in one of our electronics. Given the suspicious coincidences of these candidates and the fact that their line shapes do not agree well with a true axion's line shape, we do not in the final analysis deem any of them to be serious axion contenders.
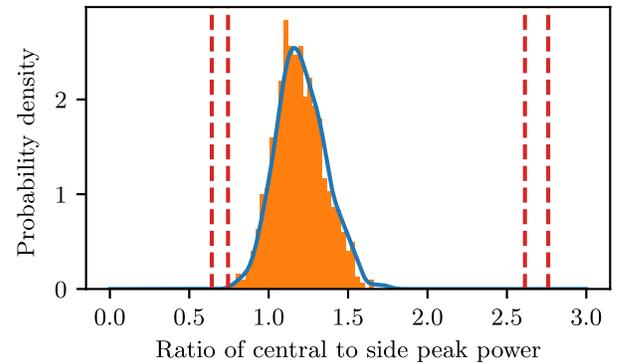

FIG. 10. A histogram of ratios of the power in the central axion peak (colored orange in Fig. 9) relative to the average power of the side peaks (colored green and red in Fig. 9) based on Monte Carlo simulations for a signal candidate at 6.666 75 Hz. From the Monte Carlo simulations, this ratio should be centered at approximately 1.2 for a true axion signal. The solid blue line shows a smoothed distribution of the ratio. Red lines indicate the experimental ratios for the four peaks near 6.666 Hz shown in the inset of Fig. 8. Given their distance from the Monte Carlo distribution, we can conclude with $>90\%$ confidence that these peaks are not true axions.

TABLE II. Fit parameters obtained via fitting Eq. (B21) to the exact components of the sensitive axis $\hat{\mathbf{m}}$.

C_i	Value	θ_i	Value	D_i	Value
C_s	0.670	θ_s	-1.550	D_s	-0.308
C_u	0.626	θ_u	-2.739	D_u	0.367
C_v	0.564	θ_v	-0.492	D_v	0.434

VI. CONCLUSION

Axions are well-motivated dark-matter candidates that arise in many theories beyond the standard model. In this paper, we reanalyze approximately 40 days of data from a K-^{3}He comagnetometer that was originally built for a long-range force experiment to search instead for dark-matter axions coupling to the neutron spin of ^{3}He.

Compared to long-range force experiments, searches for dark-matter axions produce energy shifts that are suppressed only by one factor of the small coupling constant g_{aNN} but their study is complicated by interference effects that result in stochastic experimental signatures. To correctly account for this stochastic signal, we develop a likelihood analysis to analyze the signal of the K-^{3}He comagnetometer in the frequency domain from 0.01 to 10 Hz (approximately 0.04 to 40 feV). Assuming that axions comprise all of the dark matter in the solar neighborhood, we are able to constrain $g_{aNN} < 2.4 \times 10^{-10}$ GeV^{-1} (median 95% confidence level) for axion masses between 0.4 to 4 feV. At higher masses (frequencies), the comagnetometer's sensitivity to anomalous fields is limited by the resonance frequency of ^{3}He (about 20 Hz in the original experiment). It is, in principle, possible to extend the analysis for lower frequencies, but the comagnetometer's sensitivity also deteriorates at lower frequencies. This is typically due to low-frequency noise from slow drifts in the pump-and-probe beams.

Our limits represent a significant 5-orders-of-magnitude improvement over previous laboratory bounds and serve as a useful verification of astrophysical constraints that have comparable limits but are subject to substantial uncertainties. Moreover, as we discuss in Sec. III, the K-^{3}He comagnetometer is also, barring an unlikely accidental cancellation, sensitive to anomalous fields coupling to the electron spin so that analogous constraints on the axion-electron coupling g_{aee} can be easily obtained from Fig. 5 and multiplying the limits by a simple rescaling factor of 0.87 μ_B/μ_{He} (see Sec. III for more details).

Peaks in the magnetic power spectrum of the original experiment resulted in several persistent possible axion candidates with significance greater than 5σ after taking into account the look-elsewhere effect. Their high significance suggests that the null hypothesis of a white-noise background should be rejected, but additional verification is required before acceptance of an axion hypothesis. Since we no longer have active control of the experiment, we are unable to perform detailed experimental checks on these candidates. However, analysis of their line shapes shows significant deviation from their expected values, which suggests a nonaxion origin for all of them.

ACKNOWLEDGMENTS

APPENDIX A: VALIDATION OF FREQUENCY-DOMAIN LIKELIHOOD ANALYSIS

We introduce our frequency-domain likelihood analysis in Sec. II and provide a full derivation of the frequency-domain covariance matrix in Appendix B. Here, we present our validation of the frequency-domain analysis by comparing it with the time-domain analysis recently published in Ref. [31]. We first compare both approaches on the same set of Monte Carlo data before testing them on the same subset of experimental data.

1. Validation on Monte Carlo data

In Fig. 11, we show the signal recovery plot of 100 Monte Carlo time-series datasets that are each 40 days long. The simulated datasets are injected with a constant white-noise background and a simulated axion signal of varying strengths spanning a few decades. The time-binned analysis
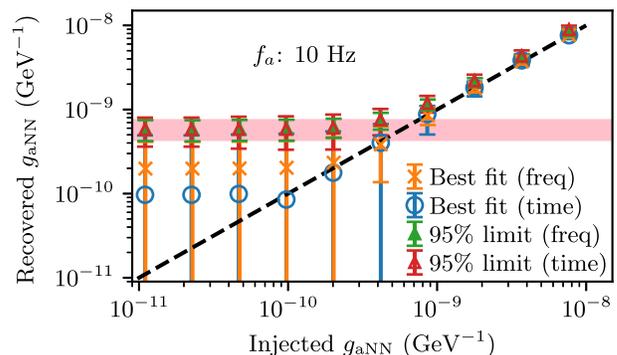

FIG. 11. Signal injection and recovery for an axion mass of 10 Hz. To create these plots, we generate 100 Monte Carlo time-series datasets sampled at 50 Hz and each lasting 40 days long with a white-noise background of 900 aT/$\sqrt{\text{Hz}}$. The time-binned analysis in Ref. [31] and the frequency-domain analysis in Sec. II are then both used to recover the best fit and 95% upper limit of the coupling constant. Error bars are the standard deviation of the recovered best fit and upper limits values over the 100 simulated datasets, while the markers denote the mean best fit and upper limits over the datasets. The null limit found by running the analysis on 100 null Monte Carlo data sets is shown as a pink band.

in Ref. [31] and the frequency-domain analysis are then independently applied to the same simulated data to recover the best-fit g_{aNN} (i.e., the unconditional maximum likelihood estimator of g_{aNN}) and 95% upper limits shown in Fig. 11. We emphasize that the Monte Carlo data here are generated in the time domain and are completely independent of the frequency-domain likelihood formalism. It is thus a nontrivial validation of the methodology presented in Appendix B.

The unconditional maximum likelihood estimator of g_{aNN} in Fig. 11 correctly recovers the injected value for a sufficiently strong injected signal. However, when the injected signal falls below the noise floor, the method is, as expected, unable to reliably recover the injected g_{aNN} as seen in the diverging error bars, which shows the standard deviation of the best-fit g_{aNN} recovered over all 100 datasets. This can also be seen in Fig. 12(a), which shows the test statistic q_0 obtained via the frequency-domain method as a function of the injected g_{aNN} for the 10-Hz
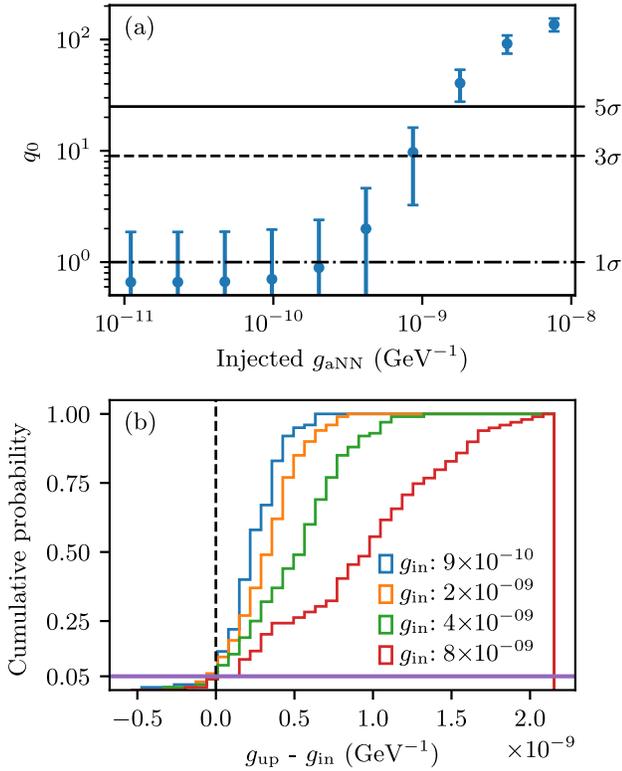

FIG. 12. (a) q_0 as a function of the injected g_{aNN} for the case of a 10-Hz axion. The marker denotes the average of the q_0 test statistic recovered from 100 independent Monte Carlo simulations, while the error bars denote the corresponding standard deviation. Horizontal lines with labels of 1σ, 3σ, and 5σ on the right indicate the significance of obtaining a particular value of q_0 (see Sec. II B for more details). (b) Empirical cumulative distribution of the recovered 95% upper limit g_{up} for the case of a 10-Hz axion where the injected g_{in} is above the null limit (see Fig. 11). As expected, the recovered limits are below the injected g_{in} for approximately 5% of the time.

axion. At low values of injected g_{aNN}, the significance of the recovered best-fit value is approximately 1σ, indicating that the null hypothesis of no axion should be preferred. However, at larger injected g_{aNN} values, the test statistic and significance of the recovered best-fit value increases accordingly before eventually saturating in the limit of high signal-to-noise ratio.

Although we cannot reliably recover the best-fit value when the injected signal disappears below the noise floor, we can nevertheless still reliably set upper limits. This can be seen for both the time- and frequency-domain analysis in Fig. 11 where both methods set the 95% upper limits in the low injected g_{aNN} regime with error bars that agree well with the null limit that is shown as a pink band. At higher injected g_{aNN} above the null limit, the logarithmic scale of Fig. 11 makes it difficult to discern if the upper limits are recovered correctly. To validate the upper limits in this regime, we plot, for the 10-Hz axion, the cumulative distribution functions of the recovered 95% upper limits g_{up} for all injected g_{in} above the null limit and compare it against the actual injected g_{in}. As Fig. 12(b) shows, the recovered upper limits are indeed only below g_{in} about 5% of the time, which indicates that the analysis is working as intended.

2. Validation on experimental data

Besides checking that both approaches work on the same Monte Carlo datasets, we also check that both methods give similar results on a subset of the actual experimental data. In Fig. 13, we use both the frequency-domain approach of this paper and the time-domain approach of Ref. [31] to independently calculate the 95% upper limit on g_{aNN} from the experimental data over a small frequency range in the neighborhood of 0.525 Hz where the experimental noise spectrum is relatively white. As the scatter plot and
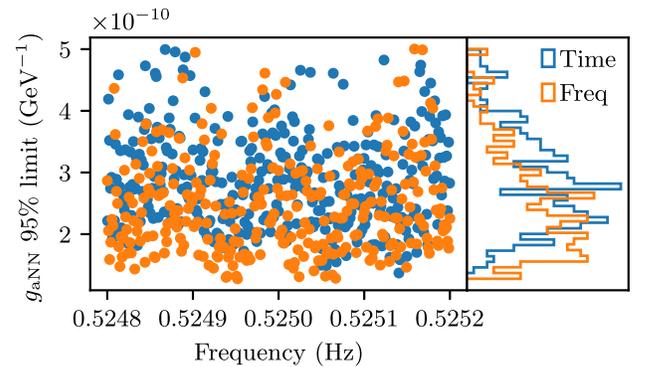

FIG. 13. Left: comparison of the 95% upper limit obtained via the frequency-domain approach of this paper (orange) and the time-domain approach of Ref. [31] (blue) over a small frequency range of the experimental data in the vicinity of 0.525 Hz. Note that the calculations are completely independent and that there is some misalignment of the frequency grids. Right: histogram of the obtained limits in the frequency range of the left panel.

histogram of Fig. 13 show, both methods agree with each other quite well.

APPENDIX B: STOCHASTIC PROPERTIES OF THE AXION GRADIENT'S FREQUENCY SPECTRUM

Throughout this paper, we claim that the experimental frequency spectrum $\mathbf{d} = \{A_k, B_k | k = 0, \ldots, N-1\}$ (see Sec. II), where A_k and B_k are as defined in Eq. (10), is normally distributed with zero mean and has a nondiagonal covariance matrix Σ_a. These stochastic properties are then used to calculate the likelihood of measuring a particular frequency spectrum. We now prove that $\mathbf{d}$ is indeed normally distributed with zero mean and provide the full derivation of Σ_a in this appendix.

Before obtaining the stochastic properties of A_k and B_k however, we note that the stochastic axion gradient, as written in Eq. (6), is a sum over all three-momentum $\mathbf{p}$ of the axion, which have different directions and magnitudes. However, since the axion's frequency spectrum can depend only on the magnitude of its momentum, it is clear that we need to first integrate out the angular degrees of freedom of $\mathbf{p}$ in Eq. (6) before we can obtain the stochastic properties of A_k and B_k.

1. Integrating out the angular dependence in p

We do this by first rewriting the sum over all $\mathbf{p}$ in Eq. (6) to a sum over infinitesimal domains $\Omega_{\mathbf{p}}$. Within each infinitesimal $\Omega_{\mathbf{p}}$ domain, we may safely assume that $\mathbf{p}$, $N_{\mathbf{p}}$, and $\omega_{\mathbf{p}}$ are approximately the same. However, because $\phi_{\mathbf{p}} \sim U(0, 2\pi)$ is a random number, it cannot be assumed to be constant within any infinitesimal domain and therefore requires an additional sum

$$
\begin{aligned}
\boldsymbol{\nabla} a(t) &= \sum_{\Omega_{\mathbf{p}}} \sqrt{\frac{2N_{\mathbf{p}}}{V\omega_{\mathbf{p}}}} \sum_{\mathbf{k} \in \Omega_{\mathbf{p}}} \cos(p^0 t + \phi_{\mathbf{k}}) \mathbf{p} \\
&= \sum_{\Omega_{\mathbf{p}}} \sqrt{\frac{2N_{\mathbf{p}}}{V\omega_{\mathbf{p}}}} \left(\cos p^0 t \sum_{\mathbf{k} \in \Omega_{\mathbf{p}}} \cos \phi_{\mathbf{k}} \right. \\
&\qquad \left. - \sin p^0 t \sum_{\mathbf{k} \in \Omega_{\mathbf{p}}} \sin \phi_{\mathbf{k}} \right) \mathbf{p} \\
&= \sum_{\Omega_{\mathbf{p}}} \sqrt{\frac{2N_{\mathbf{p}}}{V\omega_{\mathbf{p}}}} (x_{\mathbf{p}} \cos p^0 t - y_{\mathbf{p}} \sin p^0 t) \mathbf{p} \\
&= \sum_{\Omega_{\mathbf{p}}} \sqrt{\frac{N_{\mathbf{p}} M_{\mathbf{p}}}{V\omega_{\mathbf{p}}}} \alpha_{\mathbf{p}} \cos(p^0 t + \Phi_{\mathbf{p}}) \mathbf{p}.
\end{aligned}
\tag{B1}
$$

In going from the second to third line, we note that each $\cos \phi_{\mathbf{k}}$ and $\sin \phi_{\mathbf{k}}$ is identically and independently distributed. Consequently, by the central limit theorem, in the limit of $M_{\mathbf{p}} \to \infty$, where $M_{\mathbf{p}}$ is the number of modes in $\Omega_{\mathbf{p}}$,

$\sum_{\mathbf{k} \in \Omega_{\mathbf{p}}} \cos \phi_{\mathbf{k}} \to x_{\mathbf{p}} \sim N(M_{\mathbf{p}}\mathbb{E}[\cos \phi_{\mathbf{k}}], M_{\mathbf{p}}\mathrm{Var}[\cos \phi_{\mathbf{k}}]) = N(0, M_{\mathbf{p}}/2)$. Similarly, $\sum_{\mathbf{k} \in \Omega_{\mathbf{p}}} \sin \phi_{\mathbf{k}}$ converges to $y_{\mathbf{p}} \sim N(0, M_{\mathbf{p}}/2)$. Moreover, since $\mathbb{E}[\cos \phi_{\mathbf{k}} \sin \phi_{\mathbf{k}}] = \int_0^{2\pi} d\phi \times \cos \phi \sin \phi / 2\pi = 0$, the two sums are independent, and $x_{\mathbf{p}}$ and $y_{\mathbf{p}}$ are therefore independent, normally distributed variables with variance $M_{\mathbf{p}}/2$. Thus, the last line follows, where we rescale and perform a change of variables to Rayleigh distributed $\alpha_{\mathbf{p}} \sim R(1)$ and uniformly distributed $\Phi_{\mathbf{p}} \sim U(0, 2\pi)$. Furthermore, we observe that since $N_{\mathbf{p}}$ is the mean occupation number of each mode in $\Omega_{\mathbf{p}}$ and $M_{\mathbf{p}}$ is the number of modes in $\Omega_{\mathbf{p}}$, their product is the mean number of axions in $\Omega_{\mathbf{p}}$. If we assume that axions make up the entirety of the local dark-matter density $\rho_{\mathrm{DM}} \approx 0.3 \text{ GeV/cm}^3$ [65], and that they have a momentum distribution $f(\mathbf{p})d\mathbf{p}$ such that $\int d\mathbf{p} f(\mathbf{p}) = 1$, then we may write $N_{\mathbf{p}} M_{\mathbf{p}} = \rho_{\mathrm{DM}} V f(\mathbf{p})(\Delta p)^3 / \omega_{\mathbf{p}}$, which gives

$$
\begin{aligned}
\boldsymbol{\nabla} a(t) &= \sum_{\Omega_{\mathbf{p}}} \frac{\sqrt{\rho_{\mathrm{DM}} f(\mathbf{p})(\Delta p)^3}}{\omega_{\mathbf{p}}} \alpha_{\mathbf{p}} \cos(p^0 t + \Phi_{\mathbf{p}}) \mathbf{p} \\
&\approx \sum_{\Omega_{\mathbf{p}}} \sqrt{\rho_{\mathrm{DM}} f(\mathbf{v})(\Delta v)^3} \, \alpha_{\mathbf{v}} \\
&\qquad \times \cos \left[m_a \left(1 + \frac{1}{2} v^2 \right) t + \Phi_{\mathbf{v}} \right] \mathbf{v},
\end{aligned}
\tag{B2}
$$

where in the last line we take the nonrelativistic limit since halo axions are bound within the Galaxy and are nonrelativistic. In taking the nonrelativistic limit, we keep more terms in p^0 compared to $\omega_{\mathbf{p}}$ because here we are interested in the dispersion of the axion field which will give rise to qualitatively different effects. Also, we perform a change of variables from the momentum distribution function $f(\mathbf{p})$ to the velocity distribution function $f(\mathbf{v})$, which in the nonrelativistic limit is just a rescaling by the axion rest mass m_a.

The velocity distribution $f(\mathbf{v})$ can, in principle, be any probability distribution, but a prominent model is the standard halo model whereby $f(\mathbf{v})$ is given by the Maxwell-Boltzmann distribution in Eq. (5). For the sake of concreteness, we now assume that $f(\mathbf{v})$ is given by Eq. (5) and proceed to perform the angular integration in momentum space. To do so, we choose a Cartesian coordinate system $\{\hat{s}, \hat{u}, \hat{v}\}$ such that $\hat{v}$ is parallel to $\mathbf{v}_E$ and $\hat{s}, \hat{u}$ are two other orthonormal basis vectors (see Fig. 1). The ith component of $\boldsymbol{\nabla} a$ is then

$$
\begin{aligned}
(\boldsymbol{\nabla} a)^i(t) &= \sum_{jkl} \sqrt{\rho_{\mathrm{DM}} f_{jk} v_j^2 \sin \theta_k \Delta\phi \Delta\theta \Delta v} \, \alpha_{jkl} \\
&\qquad \times \cos(\omega_j t + \Phi_{jkl}) v^i,
\end{aligned}
\tag{B3}
$$

where we write the infinitesimal velocity volume Δv^3 in spherical coordinates and discretize the velocity space

into $N_j \times N_k \times N_l$ infinitesimal volumes at coordinates (v_j, θ_k, ϕ_l). The frequency ω_j is defined as $\omega_j \equiv m_a(1 + v_j^2/2)$, while $f_{jk} \equiv f(v_j, \theta_k, \phi_l) = f(v_j, \theta_k)$, and v^i is related to (v_j, θ_k, ϕ_l) by the standard transformations

$$
v^i = \begin{cases} v_j \sin\theta_k \cos\phi_l, & i = s, \\ v_j \sin\theta_k \sin\phi_l, & i = u, \\ v_j \cos\theta_k, & i = v. \end{cases} \tag{B4}
$$

Each infinitesimal volume labeled by jkl contributes for each jkl, an independent standard Rayleigh distributed random variable $\alpha_{jkl} \sim R(1)$, and an independent uniformly distributed random variable $\Phi_{jkl} \sim U(0, 2\pi)$ to the overall sum. We now rewrite Eq. (B3) as

$$
(\nabla a)^i(t) = \sum_{jk} \sqrt{\rho_{\mathrm{DM}} f_{jk} v_j^2 \sin\theta_k \Delta\theta \Delta v}\, v_j W_k^i
$$
$$
\times (\cos\omega_j t\, C_{jk}^i - \sin\omega_j t\, S_{jk}^i), \tag{B5}
$$

where

$$
C_{jk}^i = \begin{cases} \sqrt{\Delta\phi} \sum_l \alpha_{jkl} \cos\Phi_{jkl} \cos\phi_l, & i = s, \\ \sqrt{\Delta\phi} \sum_l \alpha_{jkl} \cos\Phi_{jkl} \sin\phi_l, & i = u, \\ \sqrt{\Delta\phi} \sum_l \alpha_{jkl} \cos\Phi_{jkl,} & i = v, \end{cases} \tag{B6}
$$

$$
S_{jk}^i = \begin{cases} \sqrt{\Delta\phi} \sum_l \alpha_{jkl} \sin\Phi_{jkl} \cos\phi_l, & i = s, \\ \sqrt{\Delta\phi} \sum_l \alpha_{jkl} \sin\Phi_{jkl} \sin\phi_l, & i = u, \\ \sqrt{\Delta\phi} \sum_l \alpha_{jkl} \sin\Phi_{jkl,} & i = v, \end{cases} \tag{B7}
$$

$$
W_k^i = \begin{cases} \sin\theta_k, & i = s, u, \\ \cos\theta_k, & i = v, \end{cases} \tag{B8}
$$

and in the spirit of Monte Carlo integration, we compute the azimuthal integral using the central limit theorem; i.e., we let $\phi_l \to \phi_l \sim U(0, 2\pi)$, $N_l \to \infty$, and $\Delta\phi = 2\pi/N_l \to 0$. Since each α_{jkl}, Φ_{jkl}, and ϕ_l are identically and independently distributed, the summation over l may, in the limit as $N_l \to \infty$, be evaluated using the central limit theorem to give

$$
C_{jk}^i = \begin{cases} \sqrt{\pi} x_{jk}^i, & i = s, u, \\ \sqrt{2\pi} x_{jk}^i, & i = v, \end{cases} \tag{B9}
$$

$$
S_{jk}^i = \begin{cases} \sqrt{\pi} y_{jk}^i, & i = s, u, \\ \sqrt{2\pi} y_{jk}^i, & i = v, \end{cases} \tag{B10}
$$

where each x_{jk}^i and y_{jk}^i is an independent, standard normal-distributed variable

$$
x_{jk}^i \sim N(0, 1), \qquad y_{jk}^i \sim N(0, 1). \tag{B11}
$$

It now remains to do the polar integral, which we perform by rewriting Eq. (B5) to become

$$
(\nabla a)^i(t) = \sum_j \sqrt{\rho_{\mathrm{DM}} f_j \Delta v}\, v_j^2 (D_j^i \cos\omega_j t - T_j^i \sin\omega_j t), \tag{B12}
$$

where we separate out the polar dependence in f_{jk} by defining

$$
f_j \equiv \frac{1}{(2\pi\sigma_v^2)^{3/2}} \exp\left[-\frac{v_j^2 + v_E^2}{2\sigma_v^2}\right], \tag{B13}
$$

and D_j^i, T_j^i are given by

$$
D_j^i = \sqrt{\Delta\theta} \sum_k \sqrt{\sin\theta_k e^{-\beta_j \cos\theta_k}}\, W_k^i C_{jk}^i, \tag{B14}
$$

$$
T_j^i = \sqrt{\Delta\theta} \sum_k \sqrt{\sin\theta_k e^{-\beta_j \cos\theta_k}}\, W_k^i S_{jk}^i, \tag{B15}
$$

with β_j is defined as $\beta_j \equiv v_j v_E/\sigma_v^2$. The polar integral may be evaluated using the central limit theorem as above, with $\Delta\theta = \pi/N_\theta$, $\theta_k \to \theta_k \sim U(0, \pi)$, and $N_\theta \to \infty$. By the central limit theorem, D_j^i and T_j^i are then (independent) normally distributed random variables,

$$
D_j^i, T_j^i \sim \begin{cases} N(0, 2\pi\psi_j), & i = s, u, \\ N(0, 4\pi\xi_j), & i = v, \end{cases} \tag{B16}
$$

where

$$
\psi_j = \begin{cases} \frac{2(\beta_j \cosh\beta_j - \sinh\beta_j)}{\beta_j^3}, & \beta_j \neq 0, \\ \frac{2}{3}, & \beta_j = 0, \end{cases} \tag{B17}
$$

$$
\xi_j = \begin{cases} \frac{(2+\beta_j^2)\sinh\beta_j - 2\beta_j \cosh\beta_j}{\beta_j^3}, & \beta_j \neq 0, \\ \frac{1}{3}, & \beta_j = 0. \end{cases} \tag{B18}
$$

Recognizing that D_j^i and T_j^i in Eq. (B12) are independent normally distributed variables with equal variances, we now rewrite Eq. (B12) in terms of a standard Rayleigh random variable $\alpha_{i,j} \sim R(1)$, and a uniformly distributed variable $\phi_{i,j} \sim U(0, 2\pi)$ to finally obtain an expression for the axion gradient without any angular dependence on the axion's velocity

$$(\boldsymbol{\nabla}a)^i(t) = \sum_j \sqrt{\rho_{\rm DM} f_j \Delta v} \, v_j^2 \sqrt{{\rm Var}(D_j^i)} \alpha_{i,j}$$
$$\times \cos(\omega_j t + \phi_{i,j})$$
$$= \sum_j \sqrt{\pi \rho_{\rm DM} f_j \Delta v} \, v_j^2 \epsilon_{i,j} \alpha_{i,j} \cos(\omega_j t + \phi_{i,j}),$$

$$(B19)$$

where we define

$$\epsilon_{i,j} \equiv \begin{cases} \sqrt{2\psi_j}, & i = u, s, \\ \sqrt{4\xi_j}, & i = v. \end{cases} \qquad (B20)$$

2. Distribution of A_k and B_k

Armed with Eq. (B19) for the axion's gradient that depends only on the magnitude of the axion's velocity, we are now ready to derive the stochastic properties of the experimental frequency spectrum, which as defined in Eqs. (10) and (11), is the discrete FFT of the anomalous field β_n in Eq. (7).

We begin by noting that to a good approximation, the time dependence of our experiment's sensitive axis $\hat{\mathbf{m}}$ over the approximately 100-day span can be approximated by

$$\hat{\mathbf{m}}_i(n\Delta t) \approx C_i \cos(\omega_e n\Delta t + \theta_i) + D_i, \qquad (B21)$$

where ω_e is $2\pi/(\text{sidereal day})$ and C_i, θ_i, and D_i are obtained from fitting Eq. (B21) to the actual $\hat{\mathbf{m}}_i(n\Delta t)$ during the span of the experiment, which can be calculated based on the orientation of the sensitive axis in the experiment (taken to be the outward normal to the surface of Earth), the experiment's location on Earth (40.35 °N, 74.65 °W), and the absolute time that is measured as fractional sidereal days since J2000.0. Figure 14 shows the $\{s, u, v\}$ components of the sensitive axis, as well as fits of Eq. (B21) to them, while Table II provides the values of C_i, D_i, and θ_i for $i = \{s, u, v\}$ obtained via the fits.
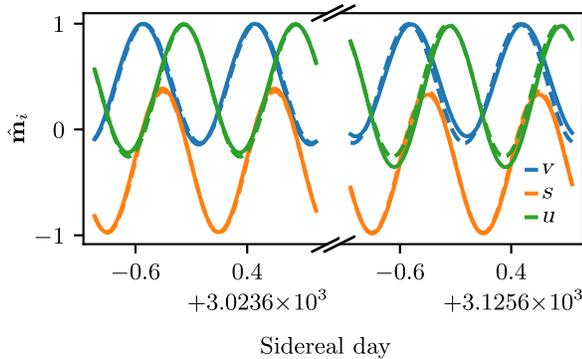

FIG. 14. Components s (orange), u (green), and v (blue) of the sensitive axis $\hat{\mathbf{m}}$ over the entire time span of the experiment. Solid lines give the exact results, while dashed lines show fits of Eq. (B21) to the exact components.

If the data are collected with a sampling interval Δt over a total interval $T \gg \Delta t$, such that the time series of the anomalous field β_n ranges from $n = 0, ..., N-1$, with $N \approx \Delta t/T$, then we may write

$$\frac{2}{N}\beta_k \approx \frac{2\sqrt{\pi\rho_{\rm DM}}g_{\rm aNN}}{T\mu_{\rm He}} \sum_{i=1}^{3} \sum_{j=1}^{M} \Delta t \sqrt{\Delta v f_j} \alpha_{i,j} \epsilon_{i,j} v_j^2$$
$$\times \sum_{n=0}^{N-1} e^{-i\Delta tn\omega_k} \left[\cos\left(\Delta tn\omega_j + \phi_{i,j}\right) D_i \right.$$
$$+ \frac{\cos\left(\Delta tn\omega_e - \Delta tn\omega_j - \phi_{i,j} + \theta_i\right)C_i}{2}$$
$$\left. + \frac{\cos\left(\Delta tn\omega_e + \Delta tn\omega_j + \phi_{i,j} + \theta_i\right)C_i}{2} \right]. \qquad (B22)$$

We note here that β_k is the discrete Fourier transform of β_n defined in Eq. (11), $\sum_i$ is over the three components of the $\{s, u, v\}$ basis, while $\sum_j$ is over the speed of the axion and $\sum_n$ is over the time series. M, which is the number of slices we discretize the speed domain of the axion into, is arbitrary, and we are therefore at liberty to take it as large as we like. The summation over n may be evaluated using a geometric sum to give

$$\sum_{n=0}^{N-1} \Delta t \cos(\omega n\Delta t + \phi) e^{-i\omega_k n\Delta t}$$
$$= \frac{\Delta t}{2} \left[\frac{1 - e^{i(\omega-\omega_k)\Delta tN}}{1 - e^{i(\omega-\omega_k)\Delta t}} e^{i\phi} + \frac{1 - e^{-i(\omega+\omega_k)\Delta tN}}{1 - e^{-i(\omega+\omega_k)\Delta t}} e^{-i\phi} \right].$$

$$(B23)$$

The terms in the bracket peak when $(\omega - \omega_k)\Delta t \approx 0$ or when $(\omega + \omega_k)\Delta t \approx 0$. For positive ω_k of interest, the first term dominates when $(\omega - \omega_k)\Delta t \approx 0$, and we may then write

$$\sum_{n=0}^{N-1} \Delta t \cos(\omega n\Delta t + \phi) e^{-i\omega_k n\Delta t}$$
$$\approx e^{i[(\omega-\omega_k)T/2+\phi]} \frac{\sin[(\omega - \omega_k)T/2]}{\omega - \omega_k}. \qquad (B24)$$

Practically, we therefore require that our experimental sampling frequency $f_s \equiv 1/\Delta t$ is sufficiently high such that $2\pi\Delta f_a/f_s \ll 1$, where Δf_a is the approximate full width at half maximum of the axion peak. For $\Delta f_a \sim 10^{-6}$ Hz, this is easily achieved and the approximation is thus well satisfied. Applying this approximation, we have from Eqs. (B22) and (10), after expanding the trigonometric functions in Eq. (B22) and collecting the coefficients of $\cos(\phi_{i,j})$ and $\sin(\phi_{i,j})$,

$$A_k = -\frac{2g_{\rm aNN}\sqrt{\pi\rho_{\rm DM}}}{T\mu_{\rm He}} \sum_{i=1}^{3} \sum_{j=1}^{M} \sqrt{\Delta v f_j} \, \alpha_{i,j} \epsilon_{i,j} v_j^2$$
$$\times [E_{i,j,k} \cos(\phi_{i,j}) + F_{i,j,k} \sin(\phi_{i,j})], \qquad (B25)$$

$$B_k = -\frac{2g_{\text{aNN}}\sqrt{\pi\rho_{\text{DM}}}}{T\mu_{\text{He}}}\sum_{i=1}^{3}\sum_{j=1}^{M}\sqrt{\Delta v f_j}\,\alpha_{i,j}\epsilon_{i,j}v_j^2$$

$$\times\,[F_{i,j,k}\cos(\phi_{i,j}) - E_{i,j,k}\sin(\phi_{i,j})],\qquad(\text{B26})$$

where $E_{i,j,k}$ is given by

$$E_{i,j,k}$$
$$= \frac{\sin\left(\frac{T\omega_e}{2} + \frac{T\omega_j}{2} - \frac{T\omega_k}{2}\right)\cos\left(\frac{T\omega_e}{2} + \frac{T\omega_j}{2} - \frac{T\omega_k}{2} + \theta_i\right)C_i}{2(\omega_e + \omega_j - \omega_k)}$$
$$+ \frac{\sin\left(\frac{T\omega_j}{2} - \frac{T\omega_k}{2}\right)\cos\left(\frac{T\omega_j}{2} - \frac{T\omega_k}{2}\right)D_i}{\omega_j - \omega_k}$$
$$- \frac{\sin\left(\frac{T\omega_e}{2} - \frac{T\omega_j}{2} + \frac{T\omega_k}{2}\right)\cos\left(\frac{T\omega_e}{2} - \frac{T\omega_j}{2} + \frac{T\omega_k}{2} + \theta_i\right)C_i}{2(-\omega_e + \omega_j - \omega_k)},$$
$$(\text{B27})$$

and $-F_{i,j,k}$ is

$$- F_{i,j,k}$$
$$= \frac{\sin\left(\frac{T\omega_e}{2} + \frac{T\omega_j}{2} - \frac{T\omega_k}{2}\right)\sin\left(\frac{T\omega_e}{2} + \frac{T\omega_j}{2} - \frac{T\omega_k}{2} + \theta_i\right)C_i}{2(\omega_e + \omega_j - \omega_k)}$$
$$+ \frac{\sin^2\left(\frac{T\omega_j}{2} - \frac{T\omega_k}{2}\right)D_i}{\omega_j - \omega_k}$$
$$+ \frac{\sin\left(\frac{T\omega_e}{2} - \frac{T\omega_j}{2} + \frac{T\omega_k}{2}\right)\sin\left(\frac{T\omega_e}{2} - \frac{T\omega_j}{2} + \frac{T\omega_k}{2} + \theta_i\right)C_i}{2(-\omega_e + \omega_j - \omega_k)}.$$
$$(\text{B28})$$

To see that A_k and B_k are normally distributed, we note that since $\alpha_{i,j} \sim R(1)$ and $\phi_{i,j} \sim U(0, 2\pi)$ are independent of each other for all i, j, we may rewrite $\alpha_{i,j}\cos(\phi_{i,j}) \equiv x_{i,j} \sim N(0,1)$ and $\alpha_{i,j}\sin(\phi_{i,j}) \equiv y_{i,j} \sim N(0,1)$. Using this substitution, we obtain

$$A_k = -\frac{2g_{\text{aNN}}\sqrt{\pi\rho_{\text{DM}}}}{T\mu_{\text{He}}}\sum_{i=1}^{3}\sum_{j=1}^{M}\sqrt{\Delta v f_j}\,\epsilon_{i,j}v_j^2$$

$$\times\,[E_{i,j,k}x_{i,j} + F_{i,j,k}y_{i,j}],\qquad(\text{B29})$$

$$B_k = -\frac{2g_{\text{aNN}}\sqrt{\pi\rho_{\text{DM}}}}{T\mu_{\text{He}}}\sum_{i=1}^{3}\sum_{j=1}^{M}\sqrt{\Delta v f_j}\,\epsilon_{i,j}v_j^2$$

$$\times\,[F_{i,j,k}x_{i,j} - E_{i,j,k}y_{i,j}].\qquad(\text{B30})$$

We thus show that A_k and B_k are normally distributed random variables with zero mean. However, A_k, B_k, A_r, B_r will in general have nonzero correlation with each other (note that k and r here index the fit frequency).

3. Covariance of A_k and B_k

We therefore desire to compute the covariance matrix that will in general consists of

$$\text{Cov}(A_k, A_r) = \mathbb{E}(A_k A_r),\qquad \text{Cov}(A_k, B_r) = \mathbb{E}(A_k B_r),$$
$$\text{Cov}(B_k, A_r) = \mathbb{E}(B_k A_r),\qquad \text{Cov}(B_k, B_r) = \mathbb{E}(B_k B_r).$$
$$(\text{B31})$$

It is important to note here that $x_{i,j}$ and $y_{i,j}$ in Eqs. (B29) and (B30) are independent of each other for all i, j, while $x_{i,j}$ and $x_{p,q}$ are independent for all i, j, p, q except when $i = p$ and $j = q$. Similarly, $y_{i,j}$ and $y_{p,q}$ are independent for all i, j, p, q except when $i = p$ and $j = q$. This may be expressed succinctly as

$$\mathbb{E}(x_{i,j}x_{p,q}) = \delta_{ip}\delta_{jq}, \mathbb{E}(y_{i,j}y_{p,q}) = \delta_{ip}\delta_{jq},$$
$$\mathbb{E}(x_{i,j}y_{p,q}) = 0.\qquad(\text{B32})$$

Using these identities, it is easy to show that the covariances are given by the following integrals:

$$\text{Cov}(A_k, A_r) = 4\pi\rho_{\text{DM}}\left(\frac{g_{\text{aNN}}}{T\mu_{\text{He}}}\right)^2\sum_i\int_0^\infty dv f(v)\epsilon_i^2(v)$$
$$\times\,v^4[E_{ik}(v)E_{ir}(v) + F_{ik}(v)F_{ir}(v)],\quad(\text{B33})$$

$$\text{Cov}(A_k, B_r) = 4\pi\rho_{\text{DM}}\left(\frac{g_{\text{aNN}}}{T\mu_{\text{He}}}\right)^2\sum_i\int_0^\infty dv f(v)\epsilon_i^2(v)$$
$$\times\,v^4[E_{ik}(v)F_{ir}(v) - F_{ik}(v)E_{ir}(v)],\quad(\text{B34})$$

$$\text{Cov}(B_k, A_r) = -\text{Cov}(A_k, B_r) = \text{Cov}(A_r, B_k),$$
$$\text{Cov}(B_k, B_r) = \text{Cov}(A_k, A_r),\qquad(\text{B35})$$

where we take the continuum limit by letting $M \to \infty$ and thus $\sum\Delta v \to \int dv$ and $v_j \to v$, which implies that $\omega_j \equiv m_a + m_a v_j^2/2 \to \omega(v) \equiv m_a + m_a v^2/2$ and $E_{ijk} \to E_{ik}(v)$, $F_{ijk} \to F_{ik}(v)$, $f_j \to f(v)$, $\epsilon_{i,j} \to \epsilon_i(v)$. Although closed-form expressions for Eqs. (B33)–(B35) may not exist, they can always be evaluated numerically.

4. Infinite-frequency resolution limit

In the limit that $T \to \infty$, the integrands in Eqs. (B33) and (B34) become increasingly oscillatory and difficult to evaluate numerically, but thankfully, closed-form solutions exist in that limit. To obtain those solutions, we note that in the continuum limit and as $T \to \infty$,

$$E_{ik}(v) \approx \frac{\pi}{m_a}\left[\frac{C_i\cos\theta_i}{2\nu_{k,-1}}\delta(v - \nu_{k,-1})\right.$$
$$\left. + \frac{C_i\cos\theta_i}{2\nu_{k,1}}\delta(v - \nu_{k,1}) + \frac{D_i}{\nu_{k,0}}\delta(v - \nu_{k,0})\right],\quad(\text{B36})$$

$$F_{ik}(v) \approx \frac{\pi C_i \sin\theta_i}{2m_a} \left[\frac{\delta(v - \nu_{k,1})}{\nu_{k,1}} - \frac{\delta(v - \nu_{k,-1})}{\nu_{k,-1}} \right], \qquad \text{(B37)}$$

since $\lim_{\epsilon \to 0} \sin(z/\epsilon)/z = \pi\delta(z)$. Physically, $\nu_{k,n}$ is the speed of an axion with mass m_a oscillating at a frequency ω_k, as measured from the nth sideband with $n \in \{-1, 0, 1\}$, and is given by Eq. (8). As explained in Sec. II, these sidebands originate from sidereal modulation of the experiment's sensitive axis as Earth rotates about its axis.

Now we like to make use of the Dirac delta functions in Eqs. (B36) and (B37) to obtain an analytical form of the covariance matrix. However, the integrals in Eqs. (B33) and (B34) contain factors like $E_{ik}E_{ir}$ and $E_{ik}F_{ir}$, which means that we end up with ill-defined terms containing two Dirac delta functions. To avoid this, we first take the continuum limit and integrate over v in Eqs. (B29) and (B30) before computing the covariance matrix elements from A_k and B_k. Although this approach avoids having to integrate over a product of Dirac delta functions, one must take the continuum limit carefully such that there is an integration measure of Δv and not $\sqrt{\Delta v}$. One way to do this is to note that a velocity grid $\{v_j\}$ has a corresponding (angular) frequency grid $\{\omega_j | \omega_j = m_a + m_a v_j^2/2\}$. The spacing of this frequency grid is $\Delta\omega \approx m_a v_j \Delta v$. If we choose this spacing to be $\Delta\omega = 2\pi/T$, then we have

$$\frac{1}{\sqrt{T}} \approx \sqrt{\frac{m_a v_j}{2\pi} \Delta v}. \qquad \text{(B38)}$$

Consequently, we may write the continuum limit of Eqs. (B29) and (B30) as

$$A_k = -\frac{g_{\text{aNN}}}{\mu_{\text{He}}} \sqrt{\frac{2m_a\rho_{\text{DM}}}{T}} \sum_{i=1}^{3} \int_0^\infty dv \sqrt{f(v)} \epsilon_i(v) v^{5/2}$$
$$\times [E_{ik}(v)x_i(v) + F_{ik}(v)y_i(v)], \qquad \text{(B39)}$$

$$B_k = -\frac{g_{\text{aNN}}}{\mu_{\text{He}}} \sqrt{\frac{2m_a\rho_{\text{DM}}}{T}} \sum_{i=1}^{3} \int_0^\infty dv \sqrt{f(v)} \epsilon_i(v) v^{5/2}$$
$$\times [F_{ik}(v)x_i(v) - E_{ik}(v)y_i(v)]. \qquad \text{(B40)}$$

Substituting in the approximations Eqs. (B36) and (B37), we obtain after integrating over v,

$$A_k \approx -\frac{\pi g_{\text{aNN}}}{m_a\mu_{\text{He}}} \sqrt{\frac{2m_a\rho_{\text{DM}}}{T}} \sum_{i=1}^{3} \left[\frac{C_i\cos\theta_i}{2} g_i(\nu_{k,-1})x(\nu_{k,-1}) \right.$$
$$+ \frac{C_i\cos\theta_i}{2} g_i(\nu_{k,1})x(\nu_{k,1}) + D_i g_i(\nu_{k,0})x(\nu_{k,0})$$
$$+ \frac{C_i\sin\theta_i}{2} g_i(\nu_{k,1})y(\nu_{k,1})$$
$$\left. - \frac{C_i\sin\theta_i}{2} g_i(\nu_{k,-1})y(\nu_{k,-1}) \right], \qquad \text{(B41)}$$

$$B_k \approx -\frac{\pi g_{\text{aNN}}}{m_a\mu_{\text{He}}} \sqrt{\frac{2m_a\rho_{\text{DM}}}{T}} \sum_{i=1}^{3} \left[-\frac{C_i\cos\theta_i}{2} g_i(\nu_{k,-1})y(\nu_{k,-1}) \right.$$
$$- \frac{C_i\cos\theta_i}{2} g_i(\nu_{k,1})y(\nu_{k,1}) - D_i g_i(\nu_{k,0})y(\nu_{k,0})$$
$$+ \frac{C_i\sin\theta_i}{2} g_i(\nu_{k,1})x(\nu_{k,1})$$
$$\left. - \frac{C_i\sin\theta_i}{2} g_i(\nu_{k,-1})x(\nu_{k,-1}) \right], \qquad \text{(B42)}$$

where for brevity we define

$$g_i(v) \equiv \sqrt{f(v)}\epsilon_i(v)v^{3/2}. \qquad \text{(B43)}$$

The covariance matrix may now be computed by remembering that $x(\nu)$ and $y(\nu)$ are independent standard normal variables for all ν, while $x(\nu_1)$ is independent from $x(\nu_2)$ for all $\nu_1 \neq \nu_2$, and $y(\nu_1)$ is independent from $y(\nu_2)$ for all $\nu_1 \neq \nu_2$. More precisely, we have

$$\text{Cov}[x(\nu_1)x(\nu_2)] = \delta_{\nu_1,\nu_2}, \qquad \text{Cov}[y(\nu_1)y(\nu_2)] = \delta_{\nu_1,\nu_2},$$
$$\text{Cov}[x(\nu_1)y(\nu_2)] = 0. \qquad \text{(B44)}$$

Using these properties, the variance of A_k and B_k can be computed to give

$$\text{Var}(A_k) = \text{Var}(B_k) = \left(\frac{g_{\text{aNN}}}{\mu_{\text{He}}}\right)^2 \frac{\pi\rho_{\text{DM}}\Delta f}{f_a} \sum_{i=1}^{3} \left[\frac{C_i^2}{4} g_i^2(\nu_{k,-1}) \right.$$
$$\left. + D_i^2 g_i^2(\nu_{k,0}) + \frac{C_i^2}{4} g_i^2(\nu_{k,1}) \right], \qquad \text{(B45)}$$

where we have $\Delta f \equiv 1/T$, and $f_a = m_a/(2\pi)$ is the axion mass frequency. Similarly, the covariance $\text{Cov}(A_kA_r)$ and $\text{Cov}(B_kB_r)$ for $k \neq r$ is

$$\text{Cov}(A_kA_r) = \left(\frac{g_{\text{aNN}}}{\mu_{\text{He}}}\right)^2 \frac{\pi\rho_{\text{DM}}\Delta f}{f_a} \sum_{i=1}^{3} \left(\chi_{i,k,-1}[\nu_{k,-1} = \nu_{r,1}] \right.$$
$$+ \sum_{m=-1}^{0} \eta_{i,k,m}[\nu_{k,m} = \nu_{r,m+1}]$$
$$+ \chi_{i,k,1}[\nu_{k,1} = \nu_{r,-1}]$$
$$\left. + \sum_{m=0}^{1} \eta_{i,k,m}[\nu_{k,m} = \nu_{r,m-1}] \right)$$
$$= \text{Cov}(B_kB_r), \qquad \text{(B46)}$$

and $\text{Cov}(A_kB_r)$ for all k, r is

$$\text{Cov}(A_k B_r) = \left(\frac{g_{\text{aNN}}}{\mu_{\text{He}}}\right)^2 \frac{\pi \rho_{\text{DM}} \Delta f}{f_a} \sum_{i=1}^{3} \left(\zeta_{i,k,-1}[\nu_{k,-1} = \nu_{r,1}]\right.$$

$$+ \sum_{m=-1}^{0} \kappa_{i,k,m}[\nu_{k,m} = \nu_{r,m+1}]$$

$$- \zeta_{i,k,1}[\nu_{k,1} = \nu_{r,-1}]$$

$$\left. - \sum_{m=0}^{1} \kappa_{i,k,m}[\nu_{k,m} = \nu_{r,m-1}]\right), \tag{B47}$$

where the square brackets [...] above denote the Iverson bracket, and we define

$$\chi_{i,k,m} \equiv \frac{C_i^2 \cos 2\theta_i}{4} g_i^2(\nu_{k,m}), \tag{B48}$$

$$\eta_{i,k,m} \equiv \frac{C_i D_i \cos \theta_i}{2} g_i^2(\nu_{k,m}), \tag{B49}$$

$$\zeta_{i,k,m} \equiv \frac{C_i^2 \sin 2\theta_i}{4} g_i^2(\nu_{k,m}), \tag{B50}$$

$$\kappa_{i,k,m} \equiv \frac{C_i D_i \sin \theta_i}{2} g_i^2(\nu_{k,m}). \tag{B51}$$

It is frequently useful to know the expected power spectrum of the axion, which in the continuum limit is defined as $\mathbb{E}[R^2] = \mathbb{E}[A^2(\omega) + B^2(\omega)]$. In the limit of infinite-frequency resolution, this can be, using techniques similar to above, evaluated to give

$$\mathbb{E}[R^2(\omega)] = \frac{2\pi \rho_{\text{DM}} \Delta f g_{\text{aNN}}^2}{f_a \mu_{\text{He}}^2}$$

$$\times \sum_{i=1}^{3} \left(\frac{C_i^2 \epsilon_i^2(\nu_{-1}(\omega)) \tilde{f}(\nu_{-1}(\omega)) \nu_{-1}^3(\omega)}{4}\right.$$

$$+ D_i^2 \epsilon_i^2(\nu_0(\omega)) \tilde{f}(\nu_0(\omega)) \nu_0^3(\omega)$$

$$\left. + \frac{C_i^2 \epsilon_i^2(\nu_1(\omega)) \tilde{f}(\nu_1(\omega)) \nu_1^3(\omega)}{4}\right), \tag{B52}$$

where

$$\nu_n(\omega) = \begin{cases} \sqrt{\dfrac{2[\omega - (m_a + n\omega_e)]}{m_a}}, & \omega - (m_a + n\omega_e) \geq 0, \\ 0, & \omega - (m_a + n\omega_e) < 0 \end{cases} \tag{B53}$$

for $n = -1, 0, 1$ is the continuum version of Eq. (8), and

$$\tilde{f}(v) \equiv \frac{1}{(2\pi\sigma_v^2)^{3/2}} \exp\left[-\frac{v^2 + v_E^2}{2\sigma_v^2}\right] \tag{B54}$$

is the continuum version of Eq. (B13).

APPENDIX C: FREQUENCY GRID SPACING

We claim in Sec. II B that it is necessary in our formalism to test for axions with a frequency grid of spacing approximately equal to $\Delta f_a/2$ to correctly recover or set upper limits on g_{aNN} from an axion signal within our mass range of interest. We now justify this claim below.

In Fig. 15(a), we show the recovered g_{aNN} from Monte Carlo datasets with injected axion signals of frequency f_a and varying g_{aNN}. At each value of injected g_{aNN}, we test for an axion with mass $f \neq f_a$ and attempt to recover g_{aNN} from 100 Monte Carlo datasets. Markers denote the average g_{aNN} recovered from all 100 Monte Carlo datasets, while vertical lines give the standard deviation of the recovered g_{aNN}. We show the results when the test frequency f is displaced from the actual axion frequency f_a by $\Delta f_a/4$ (blue) and $\Delta f/4$ (orange).

As the large standard deviation of the recovered best-fit values of g_{aNN} (orange triangle markers) in Fig. 15(a) show,
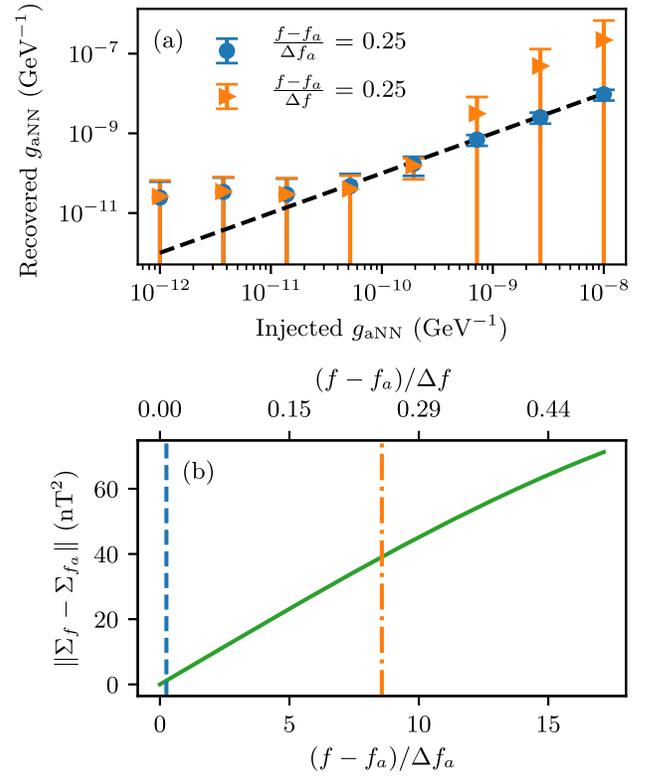

FIG. 15. (a) Signal injection and recovery plots for the case when the test frequency f does not equal the true axion frequency f_a. The blue marker shows the case when f is a quarter of the axion linewidth Δf_a away from the true axion frequency, while the orange marker shows the case when f is misaligned by a quarter of the frequency grid spacing Δf. We present here for $\Delta f_a \ll \Delta f$. (b) Frobenius norm of the difference between the test covariance matrix Σ_f and the true axion covariance matrix Σ_{f_a}. The blue dashed vertical line corresponds to the legend of the blue marker in (a) while the orange dash-dot vertical line corresponds to the legend of the orange marker in (a).

testing at the scale of the experimental resolution Δf can lead to wrong results even when the test frequency f is separated from the true axion frequency f_a by only a quarter of Δf if the axion linewidth Δf_a is sufficiently narrow compared to the frequency grid spacing ($\Delta f/\Delta f_a \approx 34$ in Fig. 15). This is because given a fixed frequency grid with spacing defined by the experimental frequency resolution, the expected power (and correlation) spectrum on that grid from a true axion signal changes smoothly as a function of the axion mass. Consequently, if there is sufficient discrepancy between the measured power (and correlation) spectrum coming from a real axion at frequency f_a compared to the expected power (and correlation) spectrum coming from an axion at test frequency $f \neq f_a$, the likelihood analysis will correctly conclude that there is no axion at test frequency f. One way to visualize this is to plot, as in Fig. 15(b), the Frobenius norm of the difference between the correlation matrix Σ_f of an axion at test frequency f and the correlation matrix Σ_{f_a} from the actual axion at frequency f_a (with both matrices defined on the same frequency grid and using the same value of g_{aNN}). As the orange dash-dot line in Fig. 15(b) shows, the test frequency f differs from f_a by only $\Delta f/4$ (upper x axis), but it differs from f_a by approximately $8\Delta f_a$ (bottom x axis), and the Frobenius norm of the difference in the correlation matrices is about 40 nT^2, which is sufficient for the likelihood analysis to conclude that there is no axion at test frequency f as seen by the diverging orange standard deviations of the recovered best-fit g_{aNN} in Fig. 15(a).

On the other hand, for a sufficiently small separation between f and f_a such that the difference between their respective correlation matrices is nearly zero, we expect that the likelihood analysis will be unable to differentiate between the two and will therefore recover g_{aNN} from an axion at frequency f_a as though it were at test frequency f. This is demonstrated by the blue circular markers in Fig. 15(a), which shows the likelihood analysis recovering the injected g_{aNN} from an actual axion at frequency f_a, while testing at frequency f when $f - f_a = \Delta f_a/4$ is sufficiently small such that the Frobenius norm of the difference in their correlation matrices is almost zero [see blue dashed line in Fig. 15(b)]. To put it another way, axion masses need to be tested at intervals of around $\Delta f_a/2$ so that if there were an axion signal, the farthest test frequency would be about $\Delta f_a/4$ away, and the analysis would correctly recover g_{aNN} from a real signal. Monte Carlo simulations across the frequency range of our analysis from 0.01 to 10 Hz indicates that a spacing of $\Delta f_a/2$ is adequate in that it recovers, within 1 standard deviation (taken over the ensemble of Monte Carlo simulations), the correct injected g_{aNN} value.

[1] E. Corbelli and P. Salucci, *The Extended Rotation Curve and the Dark Matter Halo of M33*, Mon. Not. R. Astron. Soc. **311**, 441 (2000).

[2] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, *A Direct Empirical Proof of the Existence of Dark Matter*, Astrophys. J. **648**, L109 (2006).

[3] M. J. Jee, H. C. Ford, G. D. Illingworth, R. L. White, T. J. Broadhurst, D. A. Coe, G. R. Meurer, A. van der Wel, N. Benitez, J. P. Blakeslee, R. J. Bouwens, L. D. Bradley, R. Demarco, N. L. Homeier, A. R. Martel, and S. Mei, *Discovery of a Ringlike Dark Matter Structure in the Core of the Galaxy Cluster Cl* $0024 + 17$, Astrophys. J. **661**, 728 (2007).

[4] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard *et al.*, *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*, Astrophys. J. Suppl. Ser. **192**, 14 (2011).

[5] M. B. Wise, H. Georgi, and S. L. Glashow, *SU(5) and the Invisible Axion*, Phys. Rev. Lett. **47**, 402 (1981).

[6] H. P. Nilles and S. Raby, *Supersymmetry and the Strong CP Problem*, Nucl. Phys. **B198**, 102 (1982).

[7] P. Svrcek and E. Witten, *Axions in String Theory*, J. High Energy Phys. 06 (2006) 051.

[8] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, *String Axiverse*, Phys. Rev. D **81**, 123530 (2010).

[9] G. Vasilakis, J. M. Brown, T. W. Kornack, and M. V. Romalis, *Limits on New Long Range Nuclear Spin-Dependent Forces Set with a K-^{3}He Comagnetometer*, Phys. Rev. Lett. **103**, 261801 (2009).

[10] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. **38**, 1440 (1977).

[11] R. D. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Pseudoparticles*, Phys. Rev. D **16**, 1791 (1977).

[12] S. Weinberg, *A New Light Boson?*, Phys. Rev. Lett. **40**, 223 (1978).

[13] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, Phys. Rev. Lett. **40**, 279 (1978).

[14] M. Dine and W. Fischler, *The Not-So-Harmless Axion*, Phys. Lett. **120B**, 137 (1983).

[15] L. Abbott and P. Sikivie, *A Cosmological Bound on the Invisible Axion*, Phys. Lett. **120B**, 133 (1983).

[16] J. Preskill, M. B. Wise, and F. Wilczek, *Cosmology of the Invisible Axion*, Phys. Lett. **120B**, 127 (1983).

[17] P. Sikivie, *Invisible Axion Search Methods*, Rev. Mod. Phys. **93**, 015004 (2021).

[18] O. Kwon, D. Lee, W. Chung, D. Ahn, H. S. Byun, F. Caspers, H. Choi, J. Choi, Y. Chong, H. Jeong, J. Jeong, J. E. Kim, J. Kim, C. Kutlu, J. Lee *et al.*, *First Results from an Axion Haloscope at CAPP around 10.7 μeV*, Phys. Rev. Lett. **126**, 191802 (2021).

[19] C. Bartram, T. Braine, R. Cervantes, N. Crisosto, N. Du, G. Leum, L. J. Rosenberg, G. Rybka, J. Yang, D. Bowring, A. S. Chou, R. Khatiwada, A. Sonnenschein, W. Wester, G. Carosi *et al.* (ADMX Collaboration), *Axion Dark Matter Experiment: Run 1B Analysis Details*, Phys. Rev. D **103**, 032002 (2021).

[20] C. D. Hoyle, D. J. Kapner, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, U. Schmidt, and H. E. Swanson, *Sub-Millimeter Tests of the Gravitational Inverse-Square Law*, Phys. Rev. D **70**, 042004 (2004).

[21] D. J. Kapner, T. S. Cook, E. G. Adelberger, J. H. Gundlach, B. R. Heckel, C. D. Hoyle, and H. E. Swanson, *Tests of the Gravitational Inverse-Square Law below the Dark-Energy Length Scale*, Phys. Rev. Lett. **98**, 021101 (2007).

[22] E. G. Adelberger, B. R. Heckel, S. Hoedl, C. D. Hoyle, D. J. Kapner, and A. Upadhye, *Particle-Physics Implications of a Recent Test of the Gravitational Inverse-Square Law*, Phys. Rev. Lett. **98**, 131104 (2007).

[23] A. G. Glenday, C. E. Cramer, D. F. Phillips, and R. L. Walsworth, *Limits on Anomalous Spin-Spin Couplings between Neutrons*, Phys. Rev. Lett. **101**, 261801 (2008).

[24] C. Abel, N. J. Ayres, G. Ban, G. Bison, K. Bodek, V. Bondar, M. Daum, M. Fairbairn, V. V. Flambaum, P. Geltenbort, K. Green, W. C. Griffith, M. van der Grinten, Z. D. Grujic, P. G. Harris *et al.*, *Search for Axionlike Dark Matter through Nuclear Spin Precession in Electric and Magnetic Fields*, Phys. Rev. X **7**, 041034 (2017).

[25] T. Wu, J. W. Blanchard, G. P. Centers, N. L. Figueroa, A. Garcon, P. W. Graham, Derek F. Jackson Kimball, S. Rajendran, Y. V. Stadnik, A. O. Sushkov, A. Wickenbrock, and D. Budker, *Search for Axionlike Dark Matter with a Liquid-State Nuclear Spin Comagnetometer*, Phys. Rev. Lett. **122**, 191302 (2019).

[26] A. Garcon, J. W. Blanchard, G. P. Centers, N. L. Figueroa, P. W. Graham, D. F. J. Kimball, S. Rajendran, A. O. Sushkov, Y. V. Stadnik, A. Wickenbrock, T. Wu, and D. Budker, *Constraints on Bosonic Dark Matter from Ultra-low-Field Nuclear Magnetic Resonance*, Sci. Adv. **5**, eaax4539 (2019).

[27] M. Jiang, H. Su, A. Garcon, X. Peng, and D. Budker, *Search for Axion-like Dark Matter with Spin-Based Amplifiers*, Nat. Phys. **17**, 1402 (2021).

[28] I. M. Bloch, G. Ronen, R. Shaham, O. Katz, T. Volansky, and O. Katz, *New Constraints on Axion-like Dark Matter Using a Floquet Quantum Detector*, Sci. Adv. **8**, eabl8919 (2022).

[29] J. W. Foster, N. L. Rodd, and B. R. Safdi, *Revealing the Dark Matter Halo with Axion Direct Detection*, Phys. Rev. D **97**, 123006 (2018).

[30] A. Derevianko, *Detecting Dark-Matter Waves with a Network of Precision-Measurement Tools*, Phys. Rev. A **97**, 042506 (2018).

[31] M. Lisanti, M. Moschella, and W. Terrano, *Stochastic Properties of Ultralight Scalar Field Gradients*, Phys. Rev. D **104**, 055037 (2021).

[32] C. P. Salemi, J. W. Foster, J. L. Ouellet, A. Gavin, K. M. W. Pappas, S. Cheng, K. A. Richardson, R. Henning, Y. Kahn, R. Nguyen, N. L. Rodd, B. R. Safdi, and L. Winslow, *Search for Low-Mass Axion Dark Matter with ABRACADABRA-10 cm*, Phys. Rev. Lett. **127**, 081801 (2021).

[33] T. Braine, R. Cervantes, N. Crisosto, N. Du, S. Kimes, L. J. Rosenberg, G. Rybka, J. Yang, D. Bowring, A. S. Chou, R. Khatiwada, A. Sonnenschein, W. Wester, G. Carosi, N. Woollett, L. D. Duffy *et al.* (ADMX Collaboration), *Extended Search for the Invisible Axion with the Axion Dark Matter Experiment*, Phys. Rev. Lett. **124**, 101303 (2020).

[34] L. Zhong, S. Al Kenany, K. M. Backes, B. M. Brubaker, S. B. Cahn, G. Carosi, Y. V. Gurevich, W. F. Kindel, S. K. Lamoreaux, K. W. Lehnert, S. M. Lewis, M. Malnou, R. H. Maruyama, D. A. Palken, N. M. Rapidis *et al.*, *Results from Phase 1 of the HAYSTAC Microwave Cavity Axion Experiment*, Phys. Rev. D **97**, 092001 (2018).

[35] A. V. Gramolin, D. Aybas, D. Johnson, J. Adam, and A. O. Sushkov, *Search for Axion-like Dark Matter with Ferromagnets*, Nat. Phys. **17**, 79 (2021).

[36] B. T. McAllister, G. Flower, E. N. Ivanov, M. Goryachev, J. Bourhill, and M. E. Tobar, *The ORGAN Experiment: An Axion Haloscope above 15 GHz*, Phys. Dark Universe **18**, 67 (2017).

[37] G. P. Centers, J. W. Blanchard, J. Conrad, N. L. Figueroa, A. Garcon, A. V. Gramolin, D. F. J. Kimball, M. Lawson, B. Pelssers, J. A. Smiga, A. O. Sushkov, A. Wickenbrock, D. Budker, and A. Derevianko, *Stochastic Fluctuations of Bosonic Dark Matter*, Nat. Commun. **12**, 7321 (2021).

[38] A. V. Gramolin, A. Wickenbrock, D. Aybas, H. Bekker, D. Budker, G. P. Centers, N. L. Figueroa, D. F. Jackson Kimball, and A. O. Sushkov, *Spectral Signatures of Axionlike Dark Matter*, Phys. Rev. D **105**, 035029 (2022).

[39] I. M. Bloch, Y. Hochberg, E. Kuflik, and T. Volansky, *Axion-like Relics: New Constraints from Old Comagnetometer Data*, J. High Energy Phys. 01 (2020) 167.

[40] In our analysis, we take the average Earth velocity over the time period of the experiment, including Earth's orbital velocity about the Sun, but ignoring Earth's rotation about its axis. This is shown to be a minor correction in Ref. [31].

[41] R. Catena and P. Ullio, *The Local Dark Matter Phase-Space Density and Impact on WIMP Direct Detection*, J. Cosmol. Astropart. Phys. 05 (2012) 005.

[42] J. Bovy, C. A. Prieto, T. C. Beers, D. Bizyaev, L. N. da Costa, K. Cunha, G. L. Ebelke, D. J. Eisenstein, P. M. Frinchaboy, A. E. G. Pérez, L. Girardi, F. R. Hearty, D. W. Hogg, J. Holtzman, M. A. G. Maia *et al.*, *The Milky Way Circular-Velocity Curve between 4 and 14 kpc from APOGEE Data*, Astrophys. J. **759**, 131 (2012).

[43] A.-C. Eilers, D. W. Hogg, H.-W. Rix, and M. K. Ness, *The Circular Velocity Curve of the Milky Way from 5 to 25 kpc*, Astrophys. J. **871**, 120 (2019).

[44] We absorb a phase shift of $\pi/2$ into the definition of $\phi_{\mathbf{p}}$, which we are at liberty to do since $\phi_{\mathbf{p}} \to \phi_{\mathbf{p}} - \pi/2$ is still uniformly distributed over 2π.

[45] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic Formulae for Likelihood-Based Tests of New Physics*, Eur. Phys. J. C **71**, 1554 (2011).

[46] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Ann. Math. Stat. **9**, 60 (1938).

[47] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large*, Trans. Am. Math. Soc. **54**, 426 (1943).

[48] Note that g_{aNN} appears in the likelihood Eq. (13) via the covariance matrix $\boldsymbol{\Sigma}$ as g_{aNN}^2 [see Eqs. (B33) and (B34)]. Consequently, it suffices to restrict testing to $g_{aNN} \geq 0$.

[49] G. Vasilakis, *Precision Measurements of Spin Interactions with High Density Atomic Vapors*, Ph. D. thesis, Princeton University, 2011.

[50] W. Happer and H. Tang, *Spin-Exchange Shift and Narrowing of Magnetic Resonance Lines in Optically Pumped Alkali Vapors*, Phys. Rev. Lett. **31,** 273 (1973).

[51] W. Happer and A. C. Tam, *Effect of Rapid Spin Exchange on the Magnetic-Resonance Spectrum of Alkali Vapors*, Phys. Rev. A **16,** 1877 (1977).

[52] T. W. Kornack and M. V. Romalis, *Dynamics of Two Overlapping Spin Ensembles Interacting by Spin Exchange*, Phys. Rev. Lett. **89,** 253002 (2002).

[53] J. L. Flowers, B. W. Petley, and M. G. Richards, *A Measurement of the Nuclear Magnetic Moment of the Helium-3 Atom in Terms of that of the Proton*, Metrologia **30,** 75 (1993).

[54] J. L. Friar, B. F. Gibson, G. L. Payne, A. M. Bernstein, and T. E. Chupp, *Neutron Polarization in Polarized ^{3}He Targets*, Phys. Rev. C **42,** 2310 (1990).

[55] J. J. Ethier and W. Melnitchouk, *Comparative Study of Nuclear Effects in Polarized Electron Scattering from ^{3}He*, Phys. Rev. C **88,** 054001 (2013).

[56] J. M. Brown, *A New Limit on Lorentz- and CPT-Violating Neutron Spin Interactions Using a Potassium-Helium Co-magnetometer*, Ph.D. thesis, Princeton University, 2011.

[57] C. McCabe, *The Earth's Velocity for Direct Detection Experiments*, J. Cosmol. Astropart. Phys. 02 (2014) 027.

[58] Pierluca Carenza, Tobias Fischer, Maurizio Giannotti, Gang Guo, Gabriel Martínez-Pinedo, and Alessandro Mirizzi, *Improved Axion Emissivity from a Supernova via Nucleon-Nucleon Bremsstrahlung*, J. Cosmol. Astropart. Phys. 10 (2019) 016.

[59] M. Buschmann, C. Dessert, J. W. Foster, A. J. Long, and B. R. Safdi, *Upper Limit on the QCD Axion Mass from Isolated Neutron Star Cooling*, Phys. Rev. Lett. **128,** 091102 (2022).

[60] G. G. Raffelt, *Astrophysical Axion Bounds*, in *Axions: Theory, Cosmology, and Experimental Searches*, edited by Markus Kuster, Georg Raffelt, and Berta Beltrán (Springer, Berlin, 2008), pp. 51–71.

[61] J. H. Chang, R. Essig, and S. D. McDermott, *Supernova 1987A Constraints on sub-GeV Dark Sectors, Millicharged Particles, the QCD Axion, and an Axion-like Particle*, J. High Energy Phys. 09 (2018) 051.

[62] N. Bar, K. Blum, and G. D'Amico, *Is There a Supernova Bound on Axions?*, Phys. Rev. D **101,** 123025 (2020).

[63] M. V. Beznogov, E. Rrapaj, D. Page, and S. Reddy, *Constraints on Axion-like Particles and Nucleon Pairing in Dense Matter from the Hot Neutron Star in HESS J1731-347*, Phys. Rev. C **98,** 035802 (2018).

[64] L. D. Luzio, M. Fedele, M. Giannotti, F. Mescia, and E. Nardi, *Stellar Evolution Confronts Axion Models*, J. Cosmol. Astropart. Phys. 02 (2022) 035.

[65] R. Catena and P. Ullio, *A Novel Determination of the Local Dark Matter Density*, J. Cosmol. Astropart. Phys. 08 (2010) 004.