# Improved analysis framework for axion dark matter searches

D. A. Palken,[1,2,*] B. M. Brubaker,[1,2,3] M. Malnou,[1,2] S. Al Kenany,[4] K. M. Backes,[3] S. B. Cahn,[3] Y. V. Gurevich,[3] S. K. Lamoreaux,[3] S. M. Lewis,[4] R. H. Maruyama,[3] N. M. Rapidis,[4] J. R. Root,[4] M. Simanovskaia,[4] T. M. Shokair,[4] Sukhman Singh,[3] D. H. Speller,[3] I. Urdinaran,[4] K. van Bibber,[4] L. Zhong,[3] and K. W. Lehnert[1,2]

[1]*JILA, National Institute of Standards and Technology and the University of Colorado, Boulder, Colorado 80309, USA*

[2]*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

[3]*Department of Physics, Yale University, New Haven, Connecticut 06511, USA*

[4]*Department of Nuclear Engineering, University of California Berkeley, California 94720, USA*

(Received 18 March 2020; accepted 20 May 2020; published 9 June 2020)

In experiments searching for axionic dark matter, the use of the standard threshold-based data analysis discards valuable information. We present a Bayesian analysis framework that builds on an existing processing protocol [B. M. Brubaker, L. Zhong, S. K. Lamoreaux, K. W. Lehnert, and K. A. van Bibber, Phys. Rev. D **96**, 123008 (2017)] to extract more information from the data of coherent axion detectors such as operating haloscopes. The analysis avoids logical subtleties that accompany the standard analysis framework and enables greater experimental flexibility on future data runs. Performing this analysis on the existing data from the HAYSTAC experiment, we find improved constraints on the axion-photon coupling $g_\gamma$ while also identifying the most promising regions of parameter space within the 23.15–24.0 $\mu$eV mass range. A comparison with the standard threshold analysis suggests a 36% improvement in scan rate from our analysis, demonstrating the utility of this framework for future axion haloscope analyses.

DOI: 10.1103/PhysRevD.101.123011

## I. INTRODUCTION

Dark matter axion detection experiments generically entail searching for a weak signal at an unknown frequency $\nu_a = m_a c^2/h$, where the axion mass $m_a$ is a free parameter whose possible values span many orders of magnitude [1]. A near-optimal search strategy [2] requires a resonant detector whose frequency is tuned slowly over a range much larger than its bandwidth, with the spectral scan rate being the figure of merit [3]. The technology enabling axion detection is most mature at rf and microwave frequencies [3–10], where haloscope [11] experiments, such as ADMX [12–21] and HAYSTAC [22–26], have achieved sensitivity to the QCD axion model band; ADMX is now sensitive to the Dine-Fischler-Srednicki-Zhitnitsky (DFSZ) benchmark coupling [18,19,27,28]. If DFSZ sensitivity is necessary to detect the axion, the scaling of a single-cavity haloscope's scan rate $R$ with $\nu_a$ as approximately $R \propto \nu_a^{-14/3}$ [25] becomes particularly foreboding. It implies that HAYSTAC and ADMX would each require roughly 20,000 years to scan the 1–10 GHz frequency decade, generously assuming noise at the quantum limit and 100% live time [3]. It is therefore important to introduce improved detection and analysis protocols applicable across the full landscape of axion detector platforms [2,29–37].

An optimization of the statistical analyses that haloscopes have historically used is readily available. The standard threshold-based confidence tests used in haloscope analyses transform continuous measurements of power into binary outcomes, discarding knowledge of what power was measured, and, with it, sensitivity to the axion. Because haloscopes are in practice statistics-limited experiments, for which additional data continues to enhance sensitivity, a more informative analysis will translate into tangible time savings during operation.

In this article, we introduce an analysis framework which simplifies the operational constraints placed on experimentalists while making use of more of the information content of coherent axion detection data. This framework builds on the existing HAYSTAC processing procedure [22] and may be readily adapted to other experiments. We begin in Sec. II by reviewing the standard threshold analysis, devoting particular attention to its discarding of valuable information and the difficulty of adhering to its rigid logical requirements. In Sec. III, we present our new framework, which we refer to as Bayesian power measured (BPM). We test the BPM framework in Sec. IV by reanalyzing the HAYSTAC phase 1 data, and find that the BPM analysis puts tighter constraints on the axion within the HAYSTAC window. In addition, it spotlights within the HAYSTAC dataset the

*daniel.palken@colorado.edu

 

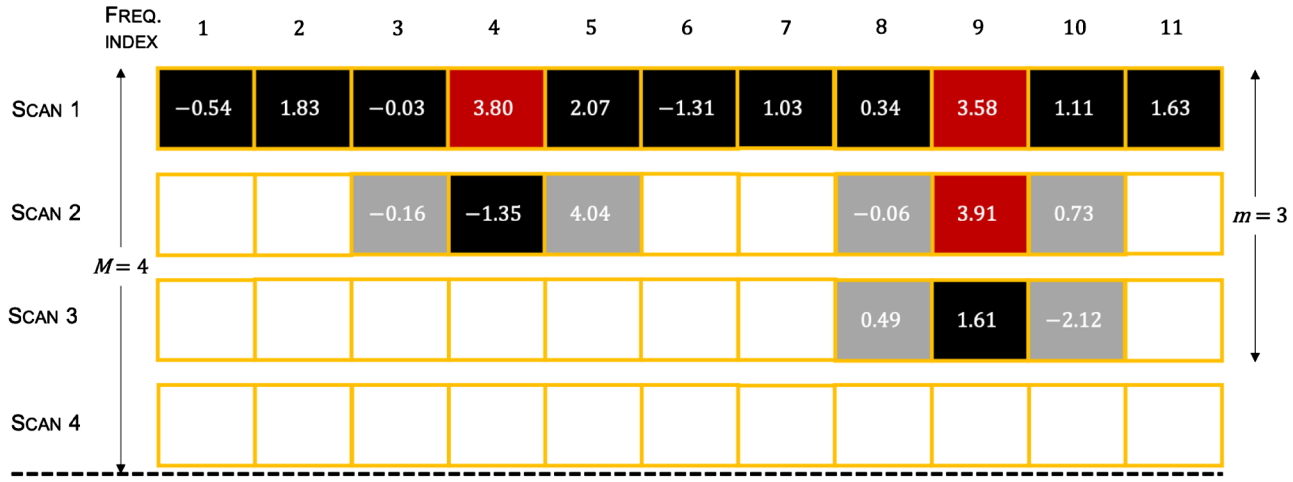


FIG. 1. Illustrative, fictitious grand spectrum data for a standard haloscope subject to a frequentist threshold (FT) analysis allowing $M = 4$ scans. The initial scan data from a typical haloscope [18,26] consists of a normalized power excess in each of $N \sim 10^5$–$10^6$ frequency bins, represented above by the boxes in the first row. At frequencies where the excess is above a predetermined threshold (red boxes), additional scans are conducted until a measurement below threshold is recorded. Only if all $M$ scans exceed threshold is failure to reject the axion hypothesis reported. If any scan comes in below threshold (black boxes), the axion hypothesis is rejected for the bin in accordance with Eq. (6) and further scanning need not be conducted (empty boxes). For the FT framework of Sec. II, both the threshold and the number of scans $M$ must be predetermined, and the exclusion therefore may take into account scans that were not performed. Here, scan 4 was not performed for any bins, but still impacts the final reported exclusion. Because the haloscope bandwidth is typically larger than the expected axion linewidth, bins adjacent to those exceeding threshold are automatically rescanned (gray boxes). However, the FT framework discards these data regardless of the power measured. Conversely, the BPM method discussed in Sec. III is able to use the information.

group of frequencies "least unlikely" to contain the axion. The most prominent among these, though still unlikely to contain an axion, nonetheless stands out relative to any other frequency within the scan range.

## II. FREQUENTIST THRESHOLD FRAMEWORK

The majority of existing axion haloscope exclusion has been obtained using a thresholding framework against which the alternative framework presented here is directly compared. The purpose of this section is to contextualize our analysis framework by explaining what has historically been meant by exclusion.[1]

Axion haloscopes place limits on the axion-photon coupling strength $g_\gamma{}^2$ by repeatedly measuring the power decaying out of a tunable resonant cavity and statistically evaluating its consistency with an axion-induced excess. A typical search involves averaging power spectra obtained from the fluctuations of the quadrature observables for an integration time of order minutes, before tuning the apparatus to a nearby frequency and repeating. A processing protocol [22] appropriately filters and combines these spectra.

This article starts where that data processing ends: with a grand spectrum of normalized power excesses predominantly from thermal fluctuations of the electromagnetic field. These normalized excesses $x_i^{(1)}$ are obtained by subtracting off the mean of the grand spectrum power distribution and dividing by its standard deviation. They are measured at frequencies $\nu_i$, with $i = 1, \ldots, N$ indexing the $N \gg 1$ frequency bins. The superscript (1) denotes data taken on an initial scan. Bins displaying large excesses on an initial scan are rescanned, potentially multiple times, yielding rescan spectra $x_i^{(j)}$ indexed as $j = 2, \ldots, M$, where many frequency bins $i$ will not be measured in rescans. Figure 1 shows how a realistic grand spectrum's frequency bins might be populated with measured power excess data. For bins not containing an axion, acquisitions of the order of a millionfold times longer than the inverse bin bandwidth yield, via the central limit theorem, normalized power excesses drawn from the standard Gaussian probability density function (PDF), obtained by setting $\mu = 0, \sigma = 1$ in

$$f_x(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{1}$$

This distribution is shaded green in Fig. 2. For a discussion of the consequences of practical departures from the Gaussian idealization, see the Appendix E.

[1]The story is complicated by the fact that not all exclusion historically uses the framework described in this section. For an example of an alternative framework, see Ref. [31]. A comparison to other, less common frameworks is beyond the scope of this work.

[2]Physically, $g_\gamma$ can take positive or negative values, but since only $g_\gamma^2$ contributes to the axion signal power, we use $g_\gamma$ to stand in for $|g_\gamma|$ throughout this article.

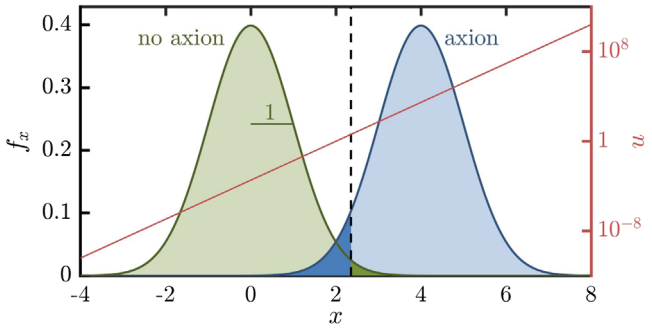


FIG. 2. Gaussian probability density functions $f_x(x)$ for a single-bin normalized power excess $x$ owing to noise alone (green) and noise plus-axion signal for a particular value of the axion-photon coupling $g_\gamma$ and haloscope sensitivity (blue). In the low-prior limit, the single-scan prior update $u(x)$ (red) is simply the ratio of the two distributions. The exponential dependence of $u$ upon $x$ captures the information content of the measurement. In threshold-based inference frameworks, conversely, an excess power threshold (black dashed line) is predetermined, and the only information analyzed is whether the measurement is above threshold. The light green (blue) region denotes to a true-negative (-positive) indication, while the dark green (blue) denotes a false-positive(-negative).

A bin $\nu_i$ that falls at the axion frequency $\nu_a$ will have its power excess drawn from what we term the axion distribution. There is in fact a family of such distributions parametrized by the axion-photon coupling $g_\gamma$, but it will suffice to consider just one. The axion distribution's standard deviation in the limit of weak axion signal power is approximately that of the no-axion distribution (i.e., 1) but its mean is offset to

$$\mu_{a,i}^{(j)} = g_{\gamma,i}^2 \eta_i^{(j)}. \tag{2}$$

The sensitivity parameter $\eta_i^{(j)}$ is the independently calibrated signal-to-noise ratio for an axion with coupling strength $g_\gamma = 1$ in the $i$th bin of the $j$th spectrum.[3] The PDF of the axion distribution for the $i$th bin (blue in Fig. 2) is thus Eq. (1) with $\mu = \mu_{a,i}^{(j)}$, $\sigma = 1$.

A power excess threshold $x_T^{(j)}$ is then set for each scan $j$ and this threshold is used to determine first whether or not to proceed with rescans, and second whether or not to exclude the axion hypothesis at each frequency. The logic of this exclusion is prescribed as follows. First, the experimentalist establishes a null hypothesis. In this case, the null hypothesis for each bin is that there *is* an axion at the bin's frequency.[4] Second, the experimentalist establishes and follows a procedure for acquiring data. This procedure may include conditional steps: e.g., "acquire more data in a given bin via a rescan if and only if the initial scan power exceeds the threshold set for that bin." All possible paths of action eventually terminate either in a rejection of the null or a failure to reject the null, forming a decision tree. Exclusion is set for all bins $i$ where the null is rejected at a confidence level of

$$E = 1 - F_n, \tag{3}$$

where $F_n$ is the bin-$i$ false-negative rate of the entire procedure: the total probability that the null would have been rejected, were it true. This approach is equivalent to defining a test statistic

$$t_s = \begin{cases} 1 & x^{(j)} \geq x_T^{(j)} \ \ \forall \ \ j \\ F_n(g_\gamma) & \text{otherwise} \end{cases} \tag{4}$$

from the total false-negative rate treated as an explicit function of $g_\gamma$. Couplings where $t_s$ goes below some predetermined target rate $F_{n,\text{target}}$ are rejected at $1 - F_{n,\text{target}}$ confidence level via Eq. (3).[5] We refer to this class of methods as frequentist threshold (FT) frameworks. FT frameworks answer the question, "assuming the axion exists, what is the probability of failing to observe it?"

A FT search procedure [26] is prepared to perform as many as $M$ scans at each frequency $\nu_i$, reporting exclusion subject to its predetermined false-negative rate. At a given $\nu_i$, no additional scans are performed once any scan's measured power excess $x_i^{(j)}$ fails to exceed its predetermined threshold $x_T^{(j)}$, the vertical, dashed line in Fig. 2. For the general case where each scan may use a different false-negative rate (dark blue region of Fig. 2), given from integrating Eq. (1) by

$$f_n^{(j)} = \int_{-\infty}^{x_T^{(j)}} f_x(x; \mu_a^{(j)}, 1) dx, \tag{5}$$

the total false-negative rate is

[3]The parameters $\eta_i^{(j)}$ account for the scaling of the haloscope sensitivity with integration time, experimental parameters, and the dark matter density $\rho_a$. By normalizing the signal-to-noise ratio to $g_\gamma = 1$ and treating the multiplicative coupling factor $g_\gamma^2$ as a parameter of the axion distribution, we are following the convention in the axion search literature of assuming a fixed value of $\rho_a$ and setting limits on $g_\gamma$. The $i$ index on the coupling factor indicates that we need not consider the same axion distribution in each bin.

[4]The axion hypothesis is chosen as the null so that it might be rejected, or excluded, later. Because of the unconventional choice to make the interesting (i.e., axion) hypothesis the null, the language of false negatives and positives will adopt the unconventional usage where a positive (negative) refers to the null being true (false).

[5]The piecewise definition of $t_s$, Eq. (4), yields exclusion only when at least one scan's power excess $x^{(j)}$ comes in below its threshold $x_T^{(j)}$. When all scans exceed threshold, $t_s = 1$ indicates the presence of (possibly axionic) excess power.

$$F_n = \sum_{j=1}^{M} f_n^{(j)} \prod_{k=1}^{j-1} (1 - f_n^{(k)}). \tag{6}$$

This total false-negative rate is what sets the confidence of the reported exclusion $E$ as in Eq. (3). For the special case where all $M$ scans have equal false-negative rates, $f_n^{(j)} = f_{n0}$, Eq. (6) reduces to

$$F_{n0} = 1 - (1 - f_{n0})^M. \tag{7}$$

In practice, the false-negative rates $f_n^{(j)}$ and power thresholds $x_T^{(j)}$ are set to convenient values, removing the frequency dependence of $\mu_a^{(j)} = g_{\gamma,i}^2 \eta_i^{(j)}$ by forcing $g_\gamma$ to compensate for the frequency dependence of the sensitivity. The initial scan sensitivities $\eta_i^{(1)}$ therefore determine the minimum coupling $g_{\gamma,i}$ that is excluded (or not) at each frequency. Subsequent ($j > 1$) scans, conversely, integrate to a sensitivity $\eta_i^{(j)}$ determined by the coupling $g_{\gamma,i}$ reached on the initial scan. The result, in the case where the axion hypothesis is rejected for all bins where it can be tested, is an exclusion plot [26] whose spectral structure reflects that of $\eta_i^{(1)}$.

The procedure for collecting data and reporting exclusion in these FT frameworks must be rigidly defined and laid out in advance. If, instead, the experimentalist is allowed to alter the decision tree that leads to negative or positive results while navigating that decision tree, the result becomes susceptible to bias. To this point, the total false-negative rate $F_n$, Eq. (6), directly responsible for setting the exclusion $E$, grows with the total number $M$ of scans in the predetermined experimental procedure. Suppose, then, that the experimentalist was willing to perform as many as $M$ scans at each frequency bin $i$ before failing to reject the axion hypothesis, but that in practice some number $m < M$ was all that was required to get at least one negative result in every bin. The correct exclusion to report, perhaps unintuitively, is that which takes into account all $M$ allowed scans, including even the $M - m$ unperformed scans. The number of such unperformed scans and their individual false-negative rates $f_n^{(j)}$, Eq. (5), must be known to properly report exclusion. If not written down in advance, these numbers are difficult to estimate without inserting bias. In practice it is extremely difficult to rigorously adhere to FT logic. For example, one would expect that some exclusion would be reported for a total number of scans $M$ surpassing $m$, the exact number performed, yet this never appears in the literature [12–15,17,23,38–41]. Furthermore, exclusion has been reported for a number of scans less than the number performed [22].

Though the preceding logic and inference is frequentist in nature, an axion haloscope analysis can alternatively be carried out using the language of Bayesian statistics. In Appendix A, we describe a threshold-based Bayesian analysis (denoted BT2) which is equivalent to the FT framework, given simple assumptions applicable to all dark matter axion experiments. That is, the two frameworks share operational procedures, and they ultimately output identical conditional probabilities. This correspondence implies that we can quantitatively interpret FT axion exclusion as decreased probability of the axion hypothesis.

Adopting a Bayesian perspective, we can now ask whether a more informative analysis is possible. We consider frameworks which operate by applying Bayes' theorem,

$$P(Y|Z) = \frac{P(Z|Y)P(Y)}{P(Z)}, \tag{8}$$

to an axion search data set [42,43]. Notationally, $P(B)$ denotes the probability of event $B$ being true, while $P(B|C)$ denotes the same, conditional upon event $C$ being true. In Eq. (8), $P(Y)$ is the Bayesian prior probability of event $Y$ being true, and it is updated by the occurrence of event $Z$ to the posterior probability $P(Y|Z)$. Appendix A describes two distinct Bayesian threshold (BT) frameworks in which $Z$ is taken to be a binary outcome, or set thereof. We show that a more informative analysis is indeed possible, but that thresholding always imposes some operational restrictions and needlessly discards valuable information. Nonetheless, the BT logic exemplifies how Bayes' theorem may be applied to axion detection. In doing so, it builds a bridge between the threshold methods of this section and the BPM framework of the next.

## III. BAYESIAN POWER-MEASURED FRAMEWORK

Bayes' theorem, Eq. (8), may be applied to haloscope data so as to preserve the information content of the measurement. The information at each frequency may be aggregated into a statement about the change in probability of an axion of arbitrary coupling strength $g_{\gamma,i}$ existing anywhere within the haloscope's scan range. In the common case where the data indicates the absence of an axion, the Bayesian language of updated probability maps onto the language of frequentist exclusion (Appendices A and C, and Ref. [44]). In this section, we discuss the BPM framework, and how its posteriors can be aggregated, before applying it to the HAYSTAC phase 1 data in Sec. IV. We address potential concerns about the inevitable non-Gaussianity of real haloscope data in Appendix E, about the seemingly subjective choice of priors in Appendix F, and about the information content of the grand spectrum data in Appendix G.

The BPM framework is readily motivated by the fact that both the FT (Sec. II) and BT (Appendix A) frameworks discard valuable information. Each power excess, measured on a continuum, is reduced to a simple "click" or "no-click"

response of an effective binary detector. A measurement that comes in just below threshold on a higher-order rescan after already exceeding threshold on previous scans is orders of magnitude more likely to indicate the presence of an axion than a typical, $0\sigma$ power excess. Yet the threshold frameworks treat these events identically.

The BPM framework uses Bayes' theorem to account for the precise effect of any initial measured power $x_i^{(1)}$ on the prior probability $P_{a,i}^{(0)}$ that an axion resides in bin $i$. After measurement, the prior is updated to the posterior, $P_{a,i}^{(1)} = u_i^{(1)} P_{a,i}^{(0)}$, where $u_i^{(1)}$ denotes the first scan's prior update. If subsequent scans $j > 1$ are performed, $P_{a,i}^{(j-1)}$ will be further updated to $P_{a,i}^{(j)} = u_i^{(j)} P_{a,i}^{(j-1)}$, and so forth. In the appropriate limit of infinitesimal priors (Appendix F), the single-scan prior update for bin $i$ and scan $j$ follows from Eq. (8), independent of one's choice of prior:

$$u_i^{(j)} \approx \frac{P(\mathcal{A}_i|x_i^{(j)})}{P(\mathcal{N}_i|x_i^{(j)})} = \exp\left[-\frac{(\mu_{a,i}^{(j)})^2}{2} + \mu_{a,i}^{(j)} x_i^{(j)}\right]. \quad (9)$$

The second equality uses the Gaussian probability densities of Eq. (1), and $\mathcal{A}_i$ or $\mathcal{N}_i$ denotes the event that an axion does or does not reside in bin $i$, respectively. The approximation in Eq. (9) identifies the prior update with the Bayes factor, which compares the likelihood of two hypotheses; it applies when the axion and other sources of excess power are sufficiently unlikely. The assumption underlying the validity of this approximation is made quantitative in Appendix F, and in Appendix E we show that this assumption is conservative with respect to axion exclusion. The BPM analysis can therefore alternatively be viewed as a Bayes factor analysis, but the interpretation of the Bayes factor as a conservative approximation to the prior update makes for a more intuitive and immediately useful end result. Equation (9) also indicates that the update to the probability of an axion's existence, plotted in red in Fig. 2, scales exponentially with the measured power excess $x_i^{(j)}$. Rather than assume a particular axion distribution, we now consider the prior update $u_i^{(j)}$ as a function of $\mu_{a,i}^{(j)}$. Formally, $u_i^{(j)}$ is maximized at $\mu_{a,i}^{(j)} = x_i^{(j)}$. Thus at each frequency where a positive excess $x_i^{(j)} > 0$ is measured, we can single out a maximum likelihood (ML) axion distribution parametrized by $\mu_{a,i}^{(j)} = x_i^{(j)}$.[6] The update to the prior probability in the ML axion scales sharply with $\mu_{a,i}^{(j)} \propto g_\gamma^2$ as $u_{i,\mathrm{ML}}^{(j)} = \exp[(\mu_{a,i}^{(j)})^2/2]$. Since the updates $u_i^{(j)}$ can be mapped back to their measured power excesses $x_i^{(j)}$ and sensitivities $\eta_i^{(j)}$, this use of Bayes' theorem preserves the information content of each individual measurement.

Applying Eq. (9) to all bins $x_i^{(j)}$ across all scans yields a spectrum of total updates $U_i$, defined as

$$U_i = \frac{P_{a,i}^{(M_i)}}{P_{a,i}^{(0)}} = \prod_{j=1}^{M_i} u_i^{(j)}, \quad (10)$$

where $M_i$ is the total number of scans performed on bin $i$. Unlike in the FT framework of Sec. II, scans that are not performed do not enter into the equation, and the number of scans need not be specified in advance. Subject to its equivalence to the Bayes factor (Appendices E and F), each total update $U_i$ is the change in probability that an axion of a specified coupling $g_\gamma$ resides in the bin. For example, $U_i = 0.1$ implies that it is 10% as probable that an axion resides in the bin after all $M_i$ scans were performed than it was prior to the experiment. A total prior update of $U_i = 0.1$ carries exactly the same meaning as 90% exclusion ($E = 0.9$) in the standard analysis framework, in the limit of vanishing false-positive rates (Appendix A).

The individual $P_{a,i}^{(j)}$, or $U_i$, regarded as functions of $g_\gamma$, can be spectrally combined into an aggregate prior update[7] function:

$$\mathcal{U}(g_\gamma) = \frac{\mathcal{P}'_a(g_\gamma)}{\mathcal{P}_a(g_\gamma)} = \frac{\sum_{i=1}^N P_{a,i}^{(M_i)}(g_\gamma)}{\sum_{i=1}^N P_{a,i}^{(0)}(g_\gamma)} \approx \frac{1}{N}\sum_{i=1}^N U_i(g_\gamma), \quad (11)$$

where $\mathcal{P}'_a(g_\gamma)$ and $\mathcal{P}_a(g_\gamma)$ are the aggregate posterior and prior probabilities, respectively, of that axion existing. The aggregate update $\mathcal{U}(g_\gamma)$ does for the entire haloscope run what the total update $U_i(g_\gamma)$ does for each bin: it expresses the update to the probability that an axion of coupling $g_\gamma$ resides anywhere within the scanned frequency window. The final equality of Eq. (11) holds in the limit of approximately uniform priors, $P_{a,i}^{(0)} \approx \mathcal{P}_a/N$, applicable when the total frequency scan range is small compared to any of the scan frequencies, $\nu_N - \nu_1 \ll \nu_1$, as is the case for HAYSTAC [26] or ADMX [18]. For low-frequency experiments such as ABRACADABRA [29] or DM Radio [32], logarithmically uniform priors, Eq. (F4), should be used [2].

[6] Conversely, measurement of a negative excess $x_i^{(j)} < 0$ reduces the probability of any axion relative to the no-axion hypothesis.

[7] In this text, we refer to four kinds of prior update. The *single-scan* updates $u_i^{(j)}$ of Eq. (9) are multiplicatively combined into the *total* updates of Eq. (10) at each frequency. The total updates are subsequently aggregated via Eq. (11) into the *aggregate* update. The *subaggregated* updates of Fig. 3(a) are likewise aggregated, but over 1% of the HAYSTAC search window apiece. All updates are functions of axion-photon coupling $g_\gamma$, and are written explicitly as such where relevant. The quantities as calculated are Bayes factors, identifiable as updates subject to Appendices E and F.

Regardless of the measured excess powers $x_i^{(j)}$ and sensitivities $\eta_i^{(j)}$, there will exist a coupling $g_{\gamma,\mathrm{low}}$ below which virtually nothing is learned about the presence or absence of an axion, $\mathcal{U}(g_\gamma \leq g_{\gamma,\mathrm{low}}) \approx 1$. Likewise, there will exist another coupling $g_{\gamma,\mathrm{high}}$ above which the probability of an axion existing vanishes $\mathcal{U}(g_\gamma \geq g_{\gamma,\mathrm{high}}) \approx 0$. Between these two extremes, $\mathcal{U}(g_\gamma)$ will depend strongly on the measured power spectra. In the context of a realistic search, for which the experimental overhead is high, $\mathcal{U} > 1$ suggests that at least some bins within the range should be rescanned until either a cause of excess power is found or $\mathcal{U}$ regresses to a lower value.

The aggregate prior update has several features that account for the look-elsewhere effect—the linear growth in expected number of excesses of a given, otherwise significant size with number of independent hypotheses tested (for dependent hypotheses, see Appendix D). Because the denominator of Eq. (11) also grows linearly with tests performed, these two linear factors cancel out. Second, the typical coupling excluded will move upwards (Appendix B) as more independent tests are conducted, reflective of the increasing difficulty in ruling out a special frequency when it has more imposters to hide among. Finally, the ability to incorporate nonuniform priors is a key feature for experiments sensitive over fractionally large spectral windows. Since 90% of all independent axion tests in such an experiment will occur in the highest decade, a standard frequentist trials factor approach [44] would set an artificially higher bar for discovery in lower decades than it would have if the highest decade were not scanned; logarithmic priors scale the height of that bar with the number of tests performed.

## IV. TEST USING HAYSTAC PHASE 1 DATASET

In order to test the BPM framework, we reanalyzed the full dataset from phase 1 of the HAYSTAC experiment, which includes at several frequencies as many as $m = 3$ scans. Features specific to this dataset are discussed in Ref. [26]. Notably, several frequency bins in the original analysis were discarded for having excess rf power that could with certainty be identified as at least partially nonaxionic in nature. We set $U_i = 1$ for these bins.

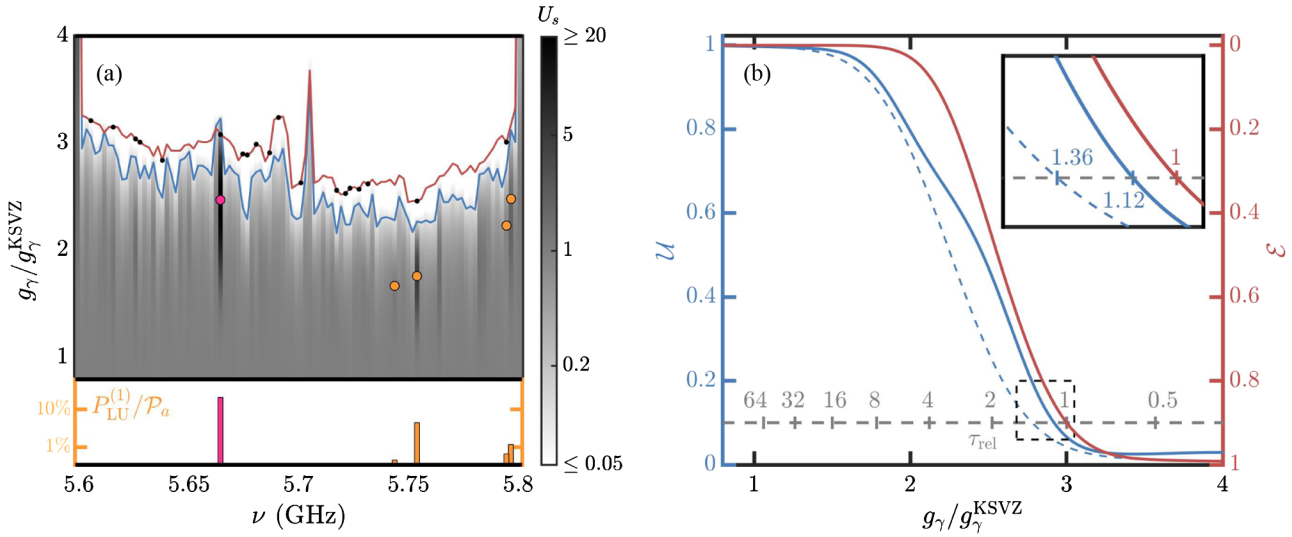


FIG. 3. (a) Reanalysis of HAYSTAC phase 1 data as a test of the Bayesian power measured (BPM) framework. The standard 90% exclusion line (red) achieved with the FT framework [26] is equivalent to the 10% prior update contour of the second Bayesian threshold (BT2) framework discussed in Appendix A. The 10% prior update contour (blue) achieved with the BPM analysis constrains more aggressive couplings at nearly all frequencies. The logarithmic color scale indicates the subaggregated prior updates $U_s$ throughout the full two-dimensional parameter space. Darker (lighter) shading corresponds to increased (decreased) probability of an axion of given frequency $\nu$ and axion-photon coupling $g_\gamma$ existing. Frequencies where rescans were performed are marked with a black dot on the threshold exclusion line. The five least unlikely (LU) axion candidates—those whose priors increased most—are marked at their maximum likelihood couplings with orange circles. In the lower panel, corresponding orange bars indicate on a logarithmic axis the total posterior probability $P_{\mathrm{LU}}^{(1)}$ within the BPM framework that an axion exists at each of the LU points normalized to the aggregate prior probability $\mathcal{P}_a$ that an axion resides anywhere in the window at that coupling. The LU candidate at $\nu_{\mathrm{LU}}^\star = 5.66417255$ GHz, marked in pink, is more than 2 times higher than the remainder of the top five combined. (b) Aggregate prior update $\mathcal{U}$ (equivalently, exclusion $\mathcal{E}$) taken as a function of coupling $g_\gamma$ over the entire HAYSTAC phase 1 frequency window for the FT and BT2 frameworks (solid red line), BPM framework (solid blue line), and BPM framework with several adjustments made for a clean comparison with the performance of the thresholding (dashed blue line; see main text). The dashed, gray line, zoomed in upon in the inset, indicates relative $\mathcal{U} = 10\%$ ($\mathcal{E} = 90\%$) scan time normalized to that achieved with the FT and BT2 frameworks [26]. Under typical experimental conditions, the exclusion achievable with the BPM framework would take an estimated 36% longer to attain using thresholding.

Otherwise, straightforward application of Eqs. (10) and (11) yield the results in Figs. 3(a) and 3(b), respectively.

The heat map of Fig. 3(a) shows the updates to the probability of the axion existing throughout the two-dimensional parameter space. At high couplings ($g_\gamma \geq g_{\gamma,\mathrm{high}} \approx 3g_\gamma^{\mathrm{KSVZ}}$),[8] the probability of an axion is greatly diminished from what it was prior to collecting data. At low couplings ($g_\gamma \leq g_{\gamma,\mathrm{low}} \approx g_\gamma^{\mathrm{KSVZ}}$), virtually no information can been gleaned from the data about the presence or absence of an axion. At intermediate couplings, $g_{\gamma,\mathrm{low}} < g_\gamma < g_{\gamma,\mathrm{high}}$, updates to the probability of an axion being present depend sharply on the specific measured powers $x_i^{(j)}$, and also on the somewhat smoother haloscope sensitivity profile $\eta_i^{(j)}$.

The FT framework has its 90% exclusion line plotted in red in Fig. 3(a). This corresponds directly[9] to the exclusion plotted in Ref. [26], and equivalently to the 10% prior update contour of the BT2 framework of Appendix A. The solid, blue line is the BPM framework's equivalent 10% prior update contour. At almost all frequencies, it constrains more pessimistic couplings than do the threshold frameworks.

At the few frequencies where the BPM framework indicates weaker exclusion than the threshold frameworks, it does so with good reason. Most of these "least unlikely" (LU) axion candidate frequencies had excess power above threshold on an initial scan and not far below threshold on a rescan. This can be seen from the spectral coincidence of several of these LU candidates, marked with orange dots at their ML couplings, with the black dots indicating the performance of at least one rescan. The LU candidates that were not rescanned corresponded to initial scan powers coming in just below threshold.

In the bottom panel of Fig. 3(a), the prior updates for the five LU candidates are plotted normalized to the aggregate prior probability that an axion resides in the HAYSTAC phase 1 window. As discussed in Ref. [2], it is not obvious how to distribute one's prior probability of an axion's existence with respect to coupling $g_\gamma$. However all five LU candidates have ML coupling strengths within a factor-of-1.5 range, wherein priors should not differ greatly. Therefore, the *relative* significance of the highest of these LU candidate prior updates (frequency $\nu_{\mathrm{LU}}^\star = 5.66417255$ GHz) is noteworthy: it is more than 2 times higher than the remaining LU candidates combined.

It should be emphasized that despite its exceptionally high prior update relative to the rest of the dataset, it is overwhelmingly unlikely that this LU candidate is an axion. A rough attempt at estimating priors (Appendix F) puts an optimistic probability of an axion in this bin still below part in 1,000, with more realistic estimates well below that. Second, as discussed in Appendix E, the BPM framework is inherently conservative with respect to exclusion in a real dataset, and consequently liberal in its identification of candidates. In particular, nonaxionic rf power excesses caused other spikes in the HAYSTAC phase 1 power spectra. Many of these were manually removed from the dataset on the grounds of supplementary evidence (e.g., failure to persist when the auxiliary, weakly coupled antenna was pulled out of the cavity) [26]. Since this LU candidate came in under threshold on a third and final scan, it was rejected subject to the FT framework. However, the more informative BPM framework suggests that significantly more than any of the other HAYSTAC phase 1 frequencies, it merits further interrogation. Generously precluding the possibility of nonaxionic excess power contamination, its probability of containing an axion at its maximum likelihood coupling is roughly the probability of 40 MHz of nearby, unscanned parameter space, or $P_{\mathrm{LU}}^{(1)}/\mathcal{P}_a = 20\%$ of the 200 MHz scan window, containing an axion at that coupling.

The frequency-dependent update (or exclusion) data of Fig. 3(a) has been aggregated in accordance with Eqs. (11) and (C1) for the BPM and FT frameworks, respectively, in Fig. 3(b). The solid, red line indicates aggregate updated probability (left axis) or exclusion of (right axis) an axion anywhere within the HAYSTAC phase 1 frequency window as a function of coupling. The solid, blue line is the updated probability for the BPM framework. Its upward bulge around $2.5g_\gamma^{\mathrm{KSVZ}}$ is largely due to the LU candidate already discussed. Had the knowledge from the BPM framework been in hand before the HAYSTAC phase 1 decommissioning [25], little additional scan time would have been required to either identify or promptly rule out an axion signal. Therefore, an exclusion curve more representative of the BPM framework for comparison purposes is given by the dashed, blue curve, where the prior update of the foremost least unlikely candidate is reset to unity. Other more minor differences between the dashed and solid blue curves due to particular features of the HAYSTAC phase 1 dataset not representative of typical haloscope data are discussed in Appendix H.

The figure of merit of an axion haloscope is the rate $R$ at which it can scan a given frequency window to a specified sensitivity [3]. The appropriate comparison to make between two analysis frameworks is therefore the relative amount of time $\tau_{\mathrm{rel}}$ taken to perform a given scan. The dashed, gray line in Fig. 3(b) represents the time that would have been required to scan any coupling $g_\gamma$ to the standard 90% exclusion (10% prior update) level using the FT (BT2) framework. The line derives from the $\tau_{\mathrm{rel}} \propto g_\gamma^{-4}$ relation [3] of scan times to couplings, and is normalized to the runtime of HAYSTAC phase 1. The inset of Fig. 3(b) shows all

[8]The benchmark Kim-Shifman-Vainshtein-Zakharov (KSVZ) [45,46] axion has coupling $g_\gamma^{\mathrm{KSVZ}} = 0.97$, making it a far more optimistic benchmark than the DFSZ axion, with $g_\gamma^{\mathrm{DFSZ}} = 0.36$.

[9]The correspondence comes with the caveat that the data here has been subaggregated into bins of width ≈2 MHz, narrower than would be of practical concern to any experimentalist considering a haloscope run at nearby frequencies.

three curves in the vicinity of this line. Had the BPM framework been in place to recommend further scanning at $\nu^{\star}_{\mathrm{LU}}$, its $\mathcal{U} = 10\%$ aggregate prior update would have taken $\approx$36% longer to achieve with a threshold analysis. In the context of detection efforts that take years to cover unsatisfactorily little parameter space, this is a significant enhancement available at no hardware or operational expense across a broad array of platforms.

## V. CONCLUSION

Haloscopes belong to a class of experiments that are statistics limited. As such, optimizing the extraction of the information content of their data is a high priority. Analyzing haloscope data with threshold frameworks, however, discards information pertinent to the presence or absence of an axion. The BPM analysis framework straightforwardly applies Bayes' theorem to incorporate the relevant information content of haloscope power spectra into a posterior probability of an axion existing. Taken in ratio with the prior probability of that axion, the updated probability is seen to be completely equivalent to the standard exclusion quoted in the literature. By applying the BPM framework to the phase 1 data of the HAYSTAC experiment, we show exclusion improved commensurate with what under typical experimental conditions would translate to a 36% improvement in scan time—within the sensitivity error bounds for the HAYSTAC phase 1 dataset, though independent of the contributing sources of error [23]. Additionally, the BPM framework better spotlights those regions of parameter space where the probability of an axion increases. For HAYSTAC's data, a single such frequency worthy of (but not demanding) further attention stands out.

The BPM framework may be straightforwardly applied to future data from ADMX [18], HAYSTAC [26], and other haloscopes and similar dark matter searches [2,29–37,47]. We expect that BPM will probe deeper couplings at no experimental cost while permitting experimentalists to operate nimbly, without the more strict constraints of preordained scan protocols. This latter advantage is especially important for narrow band, tunable searches, in which the question of how to optimize a scan protocol by responding in real time to acquired data is of paramount importance. Enhancements gained through operational efficiency will readily compound with those intrinsically available from BPM's informational advantage over threshold-based analyses.

## ACKNOWLEDGMENTS

We thank Tanya Roussy, Joshua Foster, and Lucas Sletten for helpful discussions. This work was supported by the National Science Foundation, under Grants No. PHY-1607223 and No. PHY-1734006, and by the Heising-Simons Foundation under Grant No. 2014-183.

## APPENDIX A: BAYESIAN THRESHOLD FRAMEWORKS

One of the central claims of this article is that in a haloscope analysis an apples-to-apples comparison may be made between the widely quoted frequentist exclusion $E$ of Sec. II and the Bayesian prior update $U_i$ of Sec. III. In general, these two are in fact not equivalent. The equivalence holds only between FT analyses with low false-positive rates and Bayesian analyses using low priors. In this Appendix, we consider two Bayesian threshold analyses. The first framework, BT1, is more informative than the second, BT2. In the dual limit of infinitesimal priors/low false-positive rates, BT2 reveals the Bayesian prior update to be completely equivalent to the frequentist exclusion. Together, BT1 and BT2 motivate the BPM framework as a more informative, less operationally constrained limit of haloscope analyses.

The first of the threshold-based Bayesian analyses we consider, BT1, updates its priors after each successive scan. We are interested in the conditional probability that there is an axion in the $i$th bin, given that the measured power in the $j$th scan either did ($x_i^{(j)} \geq x_T^{(j)}$) or did not ($x_i^{(j)} < x_T^{(j)}$) exceed threshold. For simplicity of notation, we will drop all frequency indices $i$ for the remainder of this and other Appendices where not necessary, with it understood that results can ultimately be aggregated via Eq. (11). We denote the event of a measured power exceeding (not exceeding) threshold as a binary detector going click (no-click), and we denote the event that the measured power excess came from the axion distribution (no-axion distribution) as $\mathcal{A}$ ($\mathcal{N}$). The posterior probability of an axion given a click is

$$P(\mathcal{A}|\mathrm{click}) = \frac{P(\mathrm{click}|\mathcal{A})P(\mathcal{A})}{P(\mathrm{click}|\mathcal{A})P(\mathcal{A}) + P(\mathrm{click}|\mathcal{N})P(\mathcal{N})}, \quad \text{(A1)}$$

where the denominator is equivalent to $P(\mathrm{click})$. Identifying $P(\mathrm{click}|\mathcal{A})$ as the single-scan true-positive rate $1 - f_n^{(j)}$ (Fig. 2, light blue) and $P(\mathrm{click}|\mathcal{N})$ as the single-scan false-positive rate $f_p^{(j)}$ (Fig. 2, dark green), abbreviating the prior (posterior) probability that there is an axion in the bin as $P_a^{(0)} = P(\mathcal{A})$ [$P_{a,\mathrm{cl}}^{(1)} = P(\mathcal{A}|\mathrm{click})$], and noting that $P(\mathcal{N}) = 1 - P_a^{(0)}$, Eq. (A1) simplifies to

$$u_{\mathrm{cl}}^{(1)} = \frac{P_{a,\mathrm{cl}}^{(1)}}{P_a^{(0)}} \approx \frac{1 - f_n^{(1)}}{f_p^{(1)}} \approx \frac{1}{f_p^{(1)}}. \quad \text{(A2)}$$

The first approximation holds in the appropriate limit of infinitesimal priors, here $P_a^{(0)} \ll f_p^{(1)}$, and the second holds in the limit of low single-scan false-negative rates, $f_n^{(1)} \ll 1$.

In the event of a no-click, the posterior probability $P_{a,\mathrm{no}}^{(1)} = P(\mathcal{A}|\mathrm{no-click})$ is given by an expression

analogous to Eq. (A1) containing the true- $(1-f_p^{(j)})$; Fig. 2, light green) and false- $(f_n^{(j)})$; Fig. 2, dark blue) negative rates. The expression simplifies as

$$u_{\rm no}^{(1)} = \frac{P_{a,\rm no}^{(1)}}{P_a^{(0)}} \approx \frac{f_n^{(1)}}{1-f_p^{(1)}} \approx f_n^{(1)}, \tag{A3}$$

where the first approximation holds for low priors $P_a^{(0)} \ll 1$, and the second holds for low single-scan false-positive rates, $f_p^{(1)} \ll 1$. Whereas single-scan false-negative rates in haloscope searches are typically close to 5%, false-positive rates are often kept much lower, to minimize the need for time-expensive rescans.

Applied to a sequence of $M$ scans, the BT1 analysis will be seen to provide a crucial operational freedom, shared by the BPM framework, to the experimentalist. Since the ordering of multiple updates is inconsequential, a series of $M$ identical $(f_n^{(j)} = f_{n0},\ f_p^{(j)} = f_{p0})$ scans has $M+1$ possible outcomes, corresponding to observing $c \in \{0, 1, \ldots, M\}$ clicks and $M-c$ no-clicks. Using the simplest forms of Eqs. (A2) and (A3), the final prior update, obtained from Eq. (10), is

$$U_c^{\rm BT1} \approx \frac{(f_{n0})^{M-c}}{(f_{p0})^c}. \tag{A4}$$

The prior is upgraded by $1/f_{p0}$ for every positive result, and downgraded by $f_{n0}$ for every negative result.

In the BT1 framework, the threshold and false-negative rate for each scan must be set in advance, but the number of scans $M$ need not be. This crucial operational freedom (which is shared by the BPM framework) is a consequence of the fact that the expected value of the prior update before performing a given scan is unity, and so the decision to perform another scan or not can be made on the fly, without biasing the outcome. In contrast, in the FT framework of Sec. II, adding another scan midway through a multiscan protocol and recalculating the false-negative rate for the entire protocol, the experimentalist can manipulate the expected value of the exclusion. This freedom has nothing to do with the use of frequentist versus Bayesian inference. We will see that Bayesian frameworks can be equally restrictive.

Whereas the BT1 framework offers $M+1$ possible outcomes for $M$ scans and provides interscan operational freedom to the experimentalist, the second BT framework (BT2) that we consider is designed to be uninformative and restrictive by comparison; its outcomes, tellingly, will map precisely onto those of the FT framework of Sec. II. The standard FT analysis permits two possible results at each frequency, illustrated for a simple case in Fig. 4: a rejection of the null (negative result) or a failure to reject it (positive result). For an $M$-scan protocol, the negative result is defined as the case where at least one scan no-clicks; the positive result occurs only when all $M$ scans click. The multiscan false-negative rate, $F_n$ from Eq. (6), sets the exclusion via Eq. (3). False-positive rates do not enter directly into the FT framework's reported exclusion.[10]

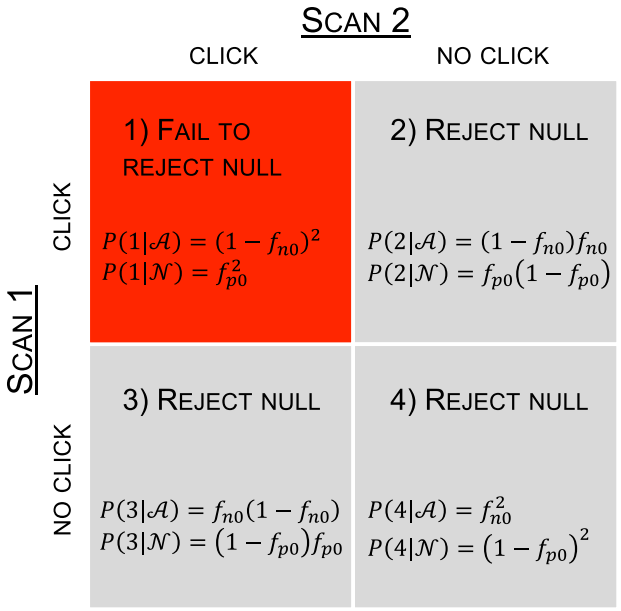


FIG. 4. The four possible outcomes of an $M = 2$-identical-scan threshold protocol. The conditional probabilities of each outcome (1)–(4) subject to the axion ($\mathcal{A}$) and no-axion ($\mathcal{N}$) hypotheses are given in terms of true and false-positive and -negative rates (Fig. 2). In both the FT framework of Sec. II and the BT2 framework of Appendix A, outcomes (2)–(4) (gray) are all treated identically as the negative result: in the FT (BT2) framework, the null is rejected (the prior is reduced). In this case, the exclusion $E = 1 - F_n$ is trivially equivalent to the prior update $U_{\rm neg}^{(\rm BT2)} \approx F_n$ in the dual limit of low false-positive rates and priors. Only region (1) (red) fails to reject the null (FT), or increases the prior (BT2).

The BT2 framework applies Bayes' theorem, Eq. (8), to this binary result landscape, obtaining a posterior either for a positive result of $P_{a,\rm pos}^{(1)}$ or for a negative result of $P_{a,\rm neg}^{(1)}$. For the positive case,

$$P(\mathcal{A}|\rm pos) = \frac{P(\rm pos|\mathcal{A})P(\mathcal{A})}{P(\rm pos|\mathcal{A})P(\mathcal{A}) + P(\rm pos|\mathcal{N})P(\mathcal{N})} \tag{A5}$$

yields

[10]More completely, false-positive rates only enter indirectly, as the global false-positive rate—the chance of at least one bin achieving a positive result (i.e., a click on all $M$ scans)—must be set sufficiently low to account for the look-elsewhere effect. Choosing a higher false-positive rate will achieve a deeper exclusion if no positive results are recorded, but will increase the probability of such an event.

$$U_{\text{pos}}^{\text{BT2}} = \frac{P_{a,\text{pos}}^{(1)}}{P_a^{(0)}} \approx \frac{1 - F_n}{F_p} \approx \frac{1}{F_p}, \tag{A6}$$

where the total false-negative rate $F_n$ is that of Eq. (6) and

$$F_p = \prod_{j=1}^{M} f_p^{(j)} \tag{A7}$$

is the total false-positive rate for the $M$ scans. The first (second) approximation in Eq. (A6) is valid for infinitesimal priors, $P_a^{(0)} \ll F_p$ (low total false-negative rate, $F_n \ll 1$). Note that BT2's positive prior update agrees in the case of identical scans (the general cases agree as well) precisely with the $c = M$ case of Eq. (A4) for BT1. The agreement is due to the fact that they derive from the same event—a series of $M$ clicks. Since false-positive rates are very small, $U_{\text{pos}}^{\text{BT2}}$ is a very large number, corresponding to just how unlikely a positive result would be to observe, without an axion present.

The negative result yields an update that will agree precisely with the standard FT exclusion $E$. From the equation for $P(\mathcal{A}|\text{neg})$ analogous to Eq. (A5), we obtain

$$U_{\text{neg}}^{\text{BT2}} = \frac{P_{a,\text{neg}}^{(1)}}{P_a^{(0)}} \approx \frac{F_n}{1 - F_p} \approx F_n, \tag{A8}$$

where the two approximations assume low priors $P_a^{(0)} \ll 1$ and low $M$-scan false-positive rates $F_p \ll 1$. With these conditions met, the BT2 negative prior update is the complement of the frequentist exclusion, $U_{\text{neg}}^{\text{BT2}} \approx 1 - E$. In the limit relevant for haloscope searches, the frequentist exclusion is therefore equivalent to the Bayesian prior update, an equivalence used in the context of other searches for new physics [44]. Whether this update is obtained through one of the BT frameworks or the BPM framework is immaterial.

Finally, a pair of differences between the two BT frameworks discussed in this Appendix hints at two important operational advantages of the BPM framework. The BT1 framework, by allowing $M + 1$ possible results with differing outcomes (or up to $2^M$ for scans with nonidentical false-negative and -positive rates), has an informational advantage over the BT2 framework. The BT2 framework deliberately blinds the experimentalist to which scan(s) no-clicked, combining $M$ of the $M + 1$ possible outcomes into a single negative result. For an experimentalist intent on discovering or excluding the axion, there can be no benefit to discarding this information. Taking the next logical step, the power-measured information from each scan need not be discarded. The BPM framework, unlike the BT frameworks, uses this information.

TABLE I. Comparison of the Bayesian threshold frameworks (BT1 and BT2) of Appendix A with the Bayesian power measured (BPM) framework of Sec. III for a protocol with $M$ possibly nonidentical scans performed. More informative frameworks map the continuum of possible measurements onto a larger number of reported outcomes. The more informative frameworks also permit the experimentalist greater freedom to alter the scan protocol without biasing the outcome. Interscan changes to upcoming scans are allowed within all but the BT2 framework [equivalently, the frequentist threshold (FT) framework of Sec. II], while the BPM framework even permits intrascan adjustments as information compiles.

| | BT2 (FT) | BT1 | BPM |
|---|---|---|---|
| Possible outcomes per bin | 2 | $2^M$ | $\infty$ |
| Interscan freedom | No | Yes | Yes |
| Intrascan freedom | No | No | Yes |

The BT1 framework also proves operationally superior to BT2. Whereas the BT2 framework locks in a commitment to $M$ scans, if necessary, BT1 allows the experimentalist to reinsert himself into the decision-making *between* scans without biasing the outcome. This interscan freedom is improved to intrascan freedom for BPM, as indicated in Table I, which summarizes the informativeness and operational constraints imposed by each framework. Under BPM, the expected prior update for the next iota of power measured always being unity protects the experimentalist from inserting bias. To the degree that the data processing allows it, probability can be tracked in real time and used to inform scan-protocol decisions on the fly. This capability provides an opportunity for optimizing a BPM haloscope search algorithm, a promising direction for future analysis beyond the scope of this article.

## APPENDIX B: OUTCOME DISTRIBUTIONS FOR BAYESIAN ANALYSES

The reanalysis of the HAYSTAC phase 1 dataset in Sec. IV indicates that the BPM framework achieves superior exclusion to thresholding in the case of one real dataset. In this Appendix, we demonstrate that this result is typical by treating the prior updates in the BPM framework and the BT1 and BT2 frameworks discussed in Appendix A as random variables. We make an apples-to-apples comparison between the aggregate prior update probability distributions obtained from an $M = 2$-identical-scan protocol in these three frameworks. In particular, while the mean aggregate prior update $\langle \mathcal{U} \rangle = 1$ for any unbiased analysis, we will see that the *median* aggregate prior update in the BPM framework is a factor-of-2 smaller than in either threshold framework. From this reduction in the median aggregate prior update we predict a typical scan rate enhancement of 30% for BPM relative to thresholding, consistent with the enhancement observed in the HAYSTAC phase 1 dataset. En route to this final result,

we obtain analytic expressions for the total prior update probability distributions in each framework. Comparing the three frameworks at this level likewise elucidates the difference between the BPM and threshold exclusion lines plotted in Fig. 3(a).

Throughout this Appendix, we assume an ideal, axionless haloscope dataset: the grand spectrum excesses are independent, identically distributed random variables $x_i^{(j)} \sim N(0,1)$. We compare prior update probability distributions for an $M = 2$-identical-scan protocol in which an initial scan is performed, followed by a rescan in each bin whose measured power excess exceeds a predetermined threshold.[11] For definiteness, we assume scan parameters similar to those used in HAYSTAC phase 1. For the BT1 and BT2 frameworks, we will first derive analytic expressions for the prior updates in a more general $M$-identical scan protocol, and then specify to $M = 2$.

Starting with the least informative BT2 framework, the total prior update in a single bin has probability mass function (PMF):

$$\begin{aligned} P(U^{\mathrm{BT2}} = U^{\mathrm{BT2}}_{\mathrm{pos}}) &= (f_{p0})^M \\ P(U^{\mathrm{BT2}} = U^{\mathrm{BT2}}_{\mathrm{neg}}) &= 1 - (f_{p0})^M. \end{aligned} \tag{B1}$$

In any scan of $N$ independent frequency bins, $k = 0, \ldots, N$ of those will realize a positive result, where $k$ is binomial distributed $k \sim B(N, (f_{p0})^M)$. The PMF for the aggregate prior update is then

$$\begin{aligned} &P\left(\mathcal{U}^{\mathrm{BT2}} = \frac{U^{\mathrm{BT2}}_{\mathrm{pos}} - U^{\mathrm{BT2}}_{\mathrm{neg}}}{N} k + U^{\mathrm{BT2}}_{\mathrm{neg}}\right) \\ &\quad = \binom{N}{k} [(f_{p0})^M]^k [1 - (f_{p0})^M]^{N-k}. \end{aligned} \tag{B2}$$

Next we consider the BT1 framework, whose total prior update $U_c^{\mathrm{BT1}}$, Eq. (A2), is written in terms of the single-scan updates $u_{\mathrm{cl}}$ and $u_{\mathrm{no}}$ for a click, Eq. (A2), and for a no-click, Eq. (A3), respectively. Each click occurs with probability $f_{p0}$, hence the probability of obtaining exactly $c = \{0, 1, \ldots, M\}$ clicks is

$$P_c = (f_{p0})^c (1 - f_{p0})^{1-\delta_{Mc}}, \tag{B3}$$

where $\delta_{ab}$ is the Kronecker delta function. The total prior update PDF is therefore

$$P(U^{\mathrm{BT1}} = [u_{\mathrm{cl}}]^c [u_{\mathrm{no}}]^{1-\delta_{Mc}}) = P_c. \tag{B4}$$

For the $M$-identical-scan protocol over $N$ independent bins, the aggregate prior update [Eq. (11)] is parametrized by the numbers $n_c$ of bins that click $c$ times, which are multinomial distributed with probabilities $P_c$ given by Eq. (B3). The aggregate prior update is

$$P\left(\mathcal{U}^{\mathrm{BT1}} = \frac{1}{N}\sum_{c=0}^{M} n_c \frac{(u_{\mathrm{cl}})^c}{(u_{\mathrm{no}})^{\delta_{Mc}-1}}\right) = \frac{N!}{\prod_{c'=0}^{M} n_{c'}!} \prod_{c''=0}^{M} (P_{c''})^{n_{c''}} \tag{B5}$$

if $\sum_{c=0}^{M} n_c = N$, and 0 otherwise.

For the BPM framework, we restrict ourselves to $M = 2$ identical scans, where the second occurs conditionally on the first exceeding threshold, and derive only the expression for the total prior update. To derive the PDF for the two-scan BPM prior update, we consider two mutually exclusive cases.

First, with probability $1 - f_{p0}$, the initial scan does not exceed threshold ($x < x_T$; a true-negative). In that case, the prior update PDF is

$$\begin{aligned} f_{u,\mathrm{tn}}(u_{\mathrm{tn}}) &= f_x(x)\left(\frac{du}{dx}\right)^{-1} \\ &= \begin{cases} (1 - f_{p0})^{-1} f_u(u) & 0 \le u_{\mathrm{tn}} < u_T \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{B6}$$

where

$$f_u(u) = \begin{cases} \left(\frac{\exp[-\mu_a^2/8]}{\mu_a\sqrt{2\pi}}\right)\left(\frac{\exp[-(\log u)^2/2\mu_a^2]}{u^{3/2}}\right) & 0 \le u \\ 0 & u < 0 \end{cases} \tag{B7}$$

and $u_T = \exp[-\mu_a^2/2 + \mu_a x_T]$ is the single-scan prior update achieved at threshold ($x = x_T$). The second line of Eq. (B6) follows from Eqs. (1) and (9). The normalization factor $(1 - f_{p0})$ accounts for the nonunit probability of drawing from this distribution in the initial scan.

Next we consider the case where the first scan exceeds threshold, ($x \ge x_T$; a false positive), which occurs with probability $f_{p0}$. The total prior update for the initial false positive is the product of two single-scan prior updates $U_{\mathrm{fp}} = u_{\mathrm{fp}} u$. The PDF for the first scan prior update $u_{\mathrm{fp}}$ is

$$f_{u,\mathrm{fp}}(u_{\mathrm{fp}}) = \begin{cases} (f_{p0})^{-1} f_u(u_{\mathrm{fp}}) & u_T \le u_{\mathrm{fp}} \\ 0 & u_{\mathrm{fp}} < u_T, \end{cases} \tag{B8}$$

and that for the second scan prior update $u$, which is not restricted to being above or below threshold, is given by Eq. (B7).

[11] As discussed in Appendix A, BPM provides freedoms to deviate from such a rigid protocol, and can therefore outperform the projections here.

The PDF for $U_{\rm fp}$ is then

$$f_{U,\rm fp}(U_{\rm fp}) = \int_{-\infty}^{\infty} f_u(u) f_{u,\rm fp}\left(\frac{U_{\rm fp}}{u}\right)\frac{du}{|u|} = \begin{cases} \frac{\exp\left[-(\mu_a/2)^2-(\log\left[U_{\rm fp}/2\mu_a\right])^2\right]}{f_{p0}\mu_a U_{\rm fp}^{3/2}4\sqrt{\pi}}\,\mathrm{erfc}\left(\frac{\log\left[u_T^2/U_{\rm fp}\right]}{2\mu_a}\right) & 0 \le U \\ 0 & U < 0, \end{cases} \tag{B9}$$

where erfc is the complementary error function.

The total prior update $U$ for the procedure is a probability-weighted sum of PDFs:

$$f_U(U) = (1-f_{p0}) f_{u,\rm tn}(U) + f_{p0} f_{U,\rm fp}(U). \tag{B10}$$

Written out fully,

$$f_U(U) = \begin{cases} \frac{\exp\left[-(\mu_a/2)^2-(\log\left[U/2\mu_a\right])^2\right]}{\mu_a U^{3/2}4\sqrt{\pi}}\,\mathrm{erfc}\left(\frac{\log\left[u_T^2/U\right]}{2\mu_a}\right) + \frac{\exp\left[-\mu_a^2/8-(\log U)^2/2\mu_a^2\right]}{\mu_a U^{3/2}\sqrt{2\pi}} & 0 \le U < u_T \\ \frac{\exp\left[-(\mu_a/2)^2-(\log\left[U/2\mu_a\right])^2\right]}{\mu_a U^{3/2}4\sqrt{\pi}}\,\mathrm{erfc}\left(\frac{\log\left[u_T^2/U\right]}{2\mu_a}\right) & u_T \le U \\ 0 & U < 0 \end{cases} \tag{B11}$$

is the two-identical-scan BPM total prior update PDF: that is, the probability density for obtaining a prior update $U$ in a single axionless bin using the standard initial and conditional-rescan procedure. The total updates for larger numbers of scans and the aggregate updates for multiple bins are too cumbersome to write out in closed form, but may readily be simulated.

Equations (B11), (B4), and (B1) are respectively the total prior update probability distributions for the BPM framework and the two BT frameworks discussed in Appendix A. These three equations are plotted as blue, green, and red lines, respectively, alongside simulated total prior updates for the three frameworks (blue, green, and red dots) for a two-scan protocol without an axion present in Fig. 5(a).

The simulations are performed by drawing Gaussian random variables according to Eq. (1), with $\mu = 0$, $\sigma = 1$, and performing a second draw as well if $x \ge x_T$. Prior updates are then applied according to Eqs. (9), (10), (A4), (A6), and (A8). The simulations match the analytic expressions, and together reveal that the vast majority of bins will receive a downward update under all three frameworks. In particular, the total prior updates for BT1 and BT2 almost always take the precise values $f_n$ and $1-(1-f_n)^2$, respectively. The price we pay for these reliably low typical updates are rare false positives with $U \gg 1$ (far off the right side of the plot). The common $U < 1$ and rare $U \gg 1$ outcomes together ensure $\langle U\rangle = 1$, the condition for an unbiased analysis. In contrast, moderate positive and negative prior updates occur much more frequently in the BPM framework, and $\langle U\rangle = 1$ is enforced by a long tail of positive prior updates. The fact that the median of the BPM distribution is lower than that of either threshold distribution is noteworthy, as we will see presently in our discussion of the aggregate prior update.

Figure 5(b) shows the aggregate prior updates $\mathcal{U}$ for the identical two-scan protocol when there are $2^{15}$ independent bins, not far from the actual number in HAYSTAC [22]. In all cases, the total updates from all bins are averaged according to Eq. (11) to obtain the aggregate prior update. All three methods deliver an unbiased aggregate prior update with mean $\langle\mathcal{U}\rangle = 1$, yet the BPM framework typically delivers a stronger exclusion in the absence of an axion, i.e., its *median* aggregate prior update is smaller than in either threshold framework by a factor of 2. The improvement is a consequence of the informational advantage of the BPM framework.

The factor-of-2 typical improvement in aggregate exclusion illustrated for typical HAYSTAC parameters in Fig. 5(b) yields a scan rate enhancement consistent with the 36% speedup inferred from real data in Fig. 3(b), absent an axion.[12] If the time for the BT2 framework to achieve the targeted 90% exclusion across the scan window (factor-of-10 reduction in aggregate probability) is $T_{90}$, the additional

[12]With an axion present, additional simulations (not shown) reveal its presence in a manner consistent with the interpretation of the global false-positive rate used in frequentist hypothesis testing.

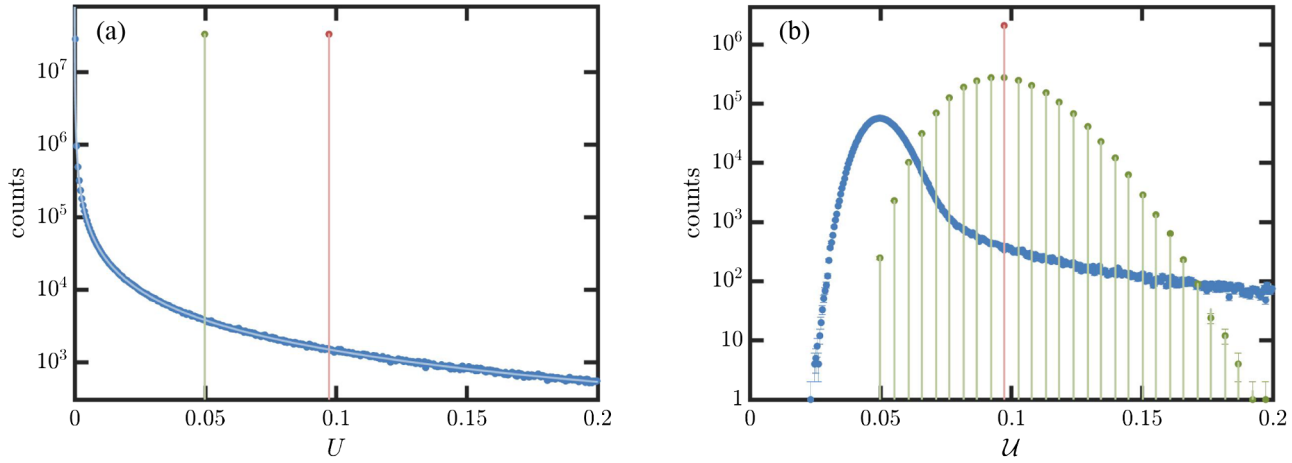


FIG. 5. Outcome probability distributions for prior updates for an axion that would produce a mean excess $\mu_a = 5.1$, assuming there is in reality no axion present and using a standard $M = 2$-identical-scan protocol with scan parameters similar to those of HAYSTAC: threshold $x_T = 3.455$ ($f_{n0} = 5\%$, $f_{p0} = 0.03\%$). (a) Histograms of simulated total prior updates $U$, Eq. (10), for the BPM (blue dots), BT1 (green dots), and BT2 (red dots) frameworks. $2^{25}$ independent bins were simulated, and error bars estimated from binomial statistics are smaller than the data points. The simulated probability distributions are in excellent agreement with the analytic predictions of Eqs. (B11), (B4), and (B1) (blue, green, and red lines). The total updates described in (a) are combined via Eq. (11) to form the aggregate prior updates $\mathcal{U}$ described in (b). The aggregate prior updates are numerically generated over $2^{21}$ independent trials of $N = 2^{15}$ independent bins for each of the three analysis frameworks. For the frameworks for which we derive analytic predictions [green and red lines for BT1, Eq. (B5), and BT2, Eq. (B2), respectively], the agreement with the simulations within binomial error bars confirms our expectations. For all three frameworks, the (relatively few) bins which fall outside of the plotted windows do indeed balance the scales so as to produce the unbiased expectation values $\langle U \rangle = \langle \mathcal{U} \rangle = 1$ (Appendix D). The BPM framework displays by far the lowest median outcomes in all cases, indicating that it will typically achieve superior exclusion when no axion is present.

factor of 2 should take approximately $T_{95} = T_{90} \log_{10}(2)$, roughly 30% longer.

## APPENDIX C: AGGREGATE EXCLUSION

The aggregation of the frequency-dependant Bayesian prior updates, $U_i(g_\gamma) \to \mathcal{U}(g_\gamma)$ via Eq. (11) has a useful frequentist counterpart for the FT framework discussed in Sec. II. In this Appendix, we motivate and define the aggregate exclusion $\mathcal{E}(g_\gamma)$ as the analog of $\mathcal{U}(g_\gamma)$, extending the apples-to-apples comparison of FT to BPM in Appendix A to aggregated results. Additionally, we show that $\mathcal{E}(g_\gamma)$ provides a more conservative, yet more faithful assessment of the exclusion achieved by a haloscope than does the standard practice of averaging the couplings at which a given exclusion $E$ is achieved.

The aggregate exclusion $\mathcal{E}(g_\gamma)$ answers the same question that the exclusion $E_i(g_\gamma)$, here written explicitly as dependent upon coupling $g_\gamma$ and frequency bin $i$, does, but for the entire frequency range. The exclusion $E_i(g_\gamma)$ answers the question: "supposing an axion of coupling $g_\gamma$ occupies bin $i$, what is the probability of observing a predefined positive result there?" We therefore define the aggregate exclusion $\mathcal{E}(g_\gamma)$ as the chance of observing a predefined, aggregate positive result (defined below), assuming an axion to exist within exactly one bin within the range. In keeping with the uniform priors assumption of Appendix F, the likelihood of an axion to be at any given bin is equal for all bins, and for simplicity the powers in bins are assumed otherwise uncorrelated (Appendix D). Since it is not known which bin the axion is in, the aggregate positive result is defined to mean at least one positive result across the $N$ frequency bins.[13] In the logic of Sec. II, the known sensitivities $\eta_i^{(j)}$ and predetermined power thresholds $x_T^{(j)}$ together preselect couplings $g_{\gamma,i}$ to test at a uniform, desired false-negative rate $F_n$. This logic can be partially inverted: assuming the same $\eta_i^{(j)}$ and $x_T^{(j)}$, now use a spectrally uniform coupling $g_\gamma$ to preselect frequency-dependent total false-negative rates $F_{n,i}(g_\gamma)$ at which to test. This framing allows for the testing of a uniform coupling $g_\gamma$ across the entire range to a confidence level of

$$\mathcal{E}(g_\gamma) = 1 - (1 - F_p)^N \frac{1}{N} \sum_{i=1}^{N} F_{n,i}(g_\gamma)$$
$$\approx \frac{1}{N} \sum_{i=1}^{N} E_i(g_\gamma), \tag{C1}$$

where the approximation is valid when the total false-positive rate, Eq. (A7), satisfies $F_p \ll 1/N$. The confidence

[13] In other words, in the case where any nonzero number of bins click, we fail to reject the hypothesis that there is an axion somewhere within the window.

tested at, shown for the HAYSTAC phase 1 sensitivity profile by the red curve in Fig. 3(b), increases monotonically with the coupling $g_\gamma$. The coupling coincident with a desired percentile (e.g., 90%) can be determined from the sensitivity profile $\eta_i^{(j)}$ alone, and a standard confidence test performed by comparing the measured power excesses $x_i^{(j)}$ against the predetermined thresholds $x_T^{(j)}$, with a positive result defined as a click on all $M$ scans in at least one bin. Defined as such, the aggregate exclusion $\mathcal{E}(g_\gamma)$ bears exactly the same relationship to the aggregate prior update $\mathcal{U}(g_\gamma)$ that $E_i(g_\gamma)$ bears to $U_i(g_\gamma)$, making it central to comparing analysis frameworks.

The aggregate exclusion, by answering the question of frequentist confidence testing, proves a more natural figure of merit for FT exclusion over a given scan window than the standard practice [26] of quoting the average coupling $g_\gamma^{\mathrm{avg}}$ excluded to a given confidence $E$. If the aggregate coupling $\mathcal{G}_\gamma$ is defined as that for which identical aggregate exclusion $\mathcal{E}(\mathcal{G}_\gamma) = E$ is achieved, then a comparison will reliably reveal $\mathcal{G}_\gamma > g_\gamma^{\mathrm{avg}}$.[14] The standard metric, $g_\gamma^{\mathrm{avg}}$, is not somehow achieving deeper sensitivity, it is averaging in a space inconvenient to the natural logic of frequentism, and biased relative to it. More directly, an experiment that achieves 90% exclusion at a frequency-averaged coupling of $g_\gamma^{\mathrm{KSVZ}}$ has *not* rejected the hypothesis that a KSVZ axion lies within its scan window with 90% confidence.

## APPENDIX D: CORRELATIONS

In this Appendix, we argue that correlations between grand spectrum powers at different frequencies do not bias the BPM framework's outputs, and are furthermore desirable. The main text and other Appendices treat the grand spectrum power excesses as uncorrelated for simplicity. In practice, this assumption is violated, as grand spectrum bin spacing is intentionally made spectrally smaller than the axion linewidth and the frequency window of the digital high-pass SG filter applied to the data. The effect of estimating power excesses based on the axion line shape (step 8 in Appendix G) is to positively correlate nearby bins, whereas the effect of the SG filtering (step 4) is to negatively correlate bins over a somewhat longer frequency scale [22]. These are the only two processing steps that will correlate adjacent bins.[15] The data processing introduces no interscan correlations, as it handles separate scans independently. The single-scan prior updates $u_i^{(j)}(g_\gamma)$ can therefore be treated as unconditionally unbiased ($\langle u_i^{(j)} \rangle = 1$), scan-independent, frequency-dependent random variables.

Given these properties of the single-scan prior updates $u_i^{(j)}(g_\gamma)$, the result that the aggregate prior update $\mathcal{U}(g_\gamma)$ is unbiased then follows from the fact that aggregation entails multiplying across scans and summing across bins. The total prior updates $U_i(g_\gamma)$ are defined in Eq. (10) as the product of grand spectrum prior updates from $M_i$ scans. Using the fact that the expectation value of the product of *independent* random variables is the product of the expectation values, here

$$\langle U_i(g_\gamma) \rangle = \left\langle \prod_{j=1}^{M} u_i^{(j)}(g_\gamma) \right\rangle = \prod_{j=1}^{M} \langle u_i^{(j)}(g_\gamma) \rangle = 1, \quad \text{(D1)}$$

the total prior updates are seen to be unbiased, i.e., have unit expectation. Aggregation then occurs through summation of the posteriors and priors via Eq. (11). The expectation value of the sum of *dependent* random variables is the sum of the expectation values, here

$$\langle \mathcal{U}(g_\gamma) \rangle = \left\langle \sum_{i=1}^{N} U_i(g_\gamma) \right\rangle = \sum_{i=1}^{N} \langle U_i(g_\gamma) \rangle = 1, \quad \text{(D2)}$$

where the last equality uses the result of Eq. (D1) to demonstrate that $\mathcal{U}(g_\gamma)$ is also unbiased. As we relax assumptions of uniform or even low priors, the relevant updates, treated as random variables, can be shown to have unit expectation independent of grand spectrum correlations, but the formulas used to compute and combine the updates become less elegant.

The primary effect of grand spectrum correlations on the prior updates is not to bias the analysis, but only to alter the higher moments of the distribution that aggregate prior updates are drawn from. In particular, the number $N_I$ of *independent* grand spectrum bins is smaller than the number $N$ of grand spectrum excesses calculated. This increases the spread of the distribution that aggregate prior updates will be drawn from for a given haloscope scan window. However, since the dominant correlations are produced by the axion line shape (step 8, Appendix G), the number of independent bins has as a rough lower bound the number of axion linewidths $\Delta_a$ in the scan window $\Delta_W$,

[14] An example for a simple $N = 2$-bin grand spectrum using an $M = 1$-scan protocol will suffice to illustrate how it is that averaging in the space of confidence levels, Eq. (C1), may produce inherently conservative results relative to averaging in the space of couplings excluded at. If one bin has sensitivity to test at 90% confidence for a $g_\gamma = g_\gamma^{\mathrm{KSVZ}}$ axion (using distribution mean $\mu_a = 5.1$) and a second has sensitivity to perform the same test for a $g_\gamma = 3g_\gamma^{\mathrm{KSVZ}}$ axion, then the average coupling tested at 90% is $2g_\gamma^{\mathrm{KSVZ}}$. However, the bin with KSVZ sensitivity is tested at 100% confidence ($F_n = 0$) for the 2-KSVZ axion, whereas the bin with 3-KSVZ sensitivity is tested at only 6% confidence ($F_n = 94\%$) for the 2-KSVZ axion. Therefore the hypothesis that a 2-KSVZ axion lies in either bin is tested not at 90% but at a far weaker confidence of 53%.

[15] Systematic effects in the measurement will also correlate grand spectrum power excesses. The identification and removal of nonaxion power excesses and the Savitzky-Golay (SG) filtering exist largely to undo such experimentally induced correlations. We do not consider the effect of such correlations on the grounds that these processing steps appear largely successful [22].

and as an upper bound the number of bins used: $\Delta_W/\Delta_a \lesssim N_I \leq N$. The fact that $N_I \leq N$ broadens the probability distributions of the aggregate outcomes (Appendix B), but does not otherwise impact the analysis.

To see that correlations are indeed desirable, note that their absence would imply a grand spectrum bin spacing larger than the axion linewidth. As discussed in Ref. [22], such a coarse bin spacing would effectively leave unprobed a large fraction of the scan window. Since the correlations produce no deleterious effects upon the BPM framework, there is no reason not to oversample the space of possible axion masses, and the correlations should be considered a feature, not a bug.

## APPENDIX E: PRACTICAL DEPARTURES FROM GAUSSIANITY

The formula for updating priors using the BPM framework, Eq. (9), assumes that the axion and no-axion distributions are Gaussian, Eq. (1). In particular, measured distributions often rise in their extreme tails over what is ideally expected [48]. In the case of haloscopic axion detection, the presence of spurious rf tones increases the probability density of the high-power tail of the PDFs of Fig. 2. In this Appendix, we argue that the interpretation of the exclusion achieved with the BPM framework is robust to the resultant bias towards discovery. Conversely, there is little to be gained from Bayesian inference in the context of manually interrogating a persistent signal with the goal of reporting discovery. Efficiency is a far less important consideration in this context, and test statistics such as those of Refs. [31,49], in conjunction with manual interrogation (footnote 17) are well suited to the task of claiming discovery.

Sources of rf noise in the haloscope's physical environment can couple into the receiver chain. Real haloscopes go to significant effort to discriminate such artificial excesses from real axion signals, but excesses sufficiently similar in spectral profile to an axion necessarily contribute to the measured power distributions, leading to an effective increase in the high power probability density of both the axion and no-axion distribution PDFs.

The effect of increased probability density in the high-power tail in either distribution in Fig. 2 is to bias the BPM analysis towards discovery of an axion. In any one frequency bin, this effect is very unlikely because it only affects the distribution tails. Indeed, for existing haloscopes, the measured power excesses seem very well approximated out to several standard deviations by Gaussian distributions [16,22], indicating that the consequences of non-Gaussian features will be modest. However, with tens or hundreds of thousands of independent frequency bins measured, there may be a small augmentation of the BPM framework's aggregate prior update, Eq. (11).

In the context of reporting exclusion over axion parameter space—the only context in which data from haloscopic searches has appeared to date—a bias towards discovery is conservative, and hence acceptable. In the context of a FT framework, the effect of a discovery bias is not to change the reported exclusion, but rather to increase expected rescan time and the chance of a false positive. Since, as shown in Fig. 3(b), the BPM framework still achieves a 36% scan rate enhancement relative to thresholding for the HAYSTAC phase 1 dataset, the conservative bias is manifestly not large enough (at least for this dataset) to overwhelm the improved exclusion over the whole scan range.

In the context of reporting the discovery of axionic dark matter, conversely, a bias towards discovery would appear to be harmful. However, the nature of axion detection guards against adverse outcomes in practice. A power excess presenting as an axion, once identified, is fairly straightforwardly shown to persist or not by manual interrogation. The role of the statistical analysis frameworks discussed in this article is not to firmly identify the axion as such once a persistent signal is found, but rather to indicate with the highest fidelity where such a signal is likely to lie. While a BPM analysis may bias the experimentalist towards believing that a few nonaxionic excesses might be axionic, the real certification of an axion signal would look like that for other discoveries in fundamental physics: a series of careful measurements precluding beyond doubt that no other known physical phenomena could account for the signal. In other words, the asymmetry between the criteria for exclusion and discovery ensure that BPM's discovery-bias would never lead to accidental discovery.

## APPENDIX F: CHOICE OF PRIORS

For any choice of prior probability $P_{a,i}^{(0)}$ in axionic dark matter, Bayes' theorem can be applied and updates $U_i$ calculated and reported. However, the utility of the BPM framework rests on the fact that its prior updates $U_i$ can be treated as independent of the priors chosen for the axion and competing hypotheses, i.e., as reasonably approximated by Bayes factors. Additionally, as argued in Appendix A, it is highly convenient that in the appropriate limit the prior updates of the BT2 framework (Appendix A), itself directly comparable to the BPM framework, become the conditional probabilities quoted in frequentist exclusion. These conveniences hold only so long as one's priors are infinitesimal but nonzero.[16] In this Appendix, we justify the assumption of infinitesimal priors, providing a toy estimate of a prior for illustrative purposes. Additionally, we discuss how the prior update can be used to directly inform scan strategy in the presence of nonaxionic rf power excesses.

[16]The fact that the prior must be nonzero is justified by Cromwell's rule, which states that priors of exactly 0 or 1 should be avoided generally. More directly, there would be no sense in performing a haloscope search if the chance of success was truly zero.

The constraint that the prior be infinitesimal is only used in simplifying the denominator of Bayes' theorem, Eq. (8), according to the law of total probability:

$$P(x_i^{(j)}) = \int_0^\infty P(x_i^{(j)}|\mathcal{A}_i(g_\gamma))P(\mathcal{A}_i(g_\gamma))dg_\gamma + P(x_i^{(j)}|\mathcal{N}_i)P(\mathcal{N}_i) \approx P(x_i^{(j)}|\mathcal{N}_i), \quad \text{(F1)}$$

where the axion hypothesis $\mathcal{A}_i(g_\gamma)$ is here written explicitly as a family of hypotheses parametrized by $g_\gamma > 0$. The case of $g_\gamma = 0$ is indistinguishable from the no-axion hypothesis $\mathcal{N}_i$. The approximation in Eq. (F1) is justified so long as the total probability of there being an axion in any bin $i$,

$$P(\mathcal{A}_{i,\mathrm{tot}}) = \int_0^\infty P(\mathcal{A}_i(g_\gamma))dg_\gamma, \quad \text{(F2)}$$

is considerably smaller than any prior update that would be applied according to Eqs. (9) and (10).

To make a coarse estimate of $P(\mathcal{A}_{i,\mathrm{tot}})$, we can write down a Drake-like equation for axionic dark matter to which a typical haloscope searching for QCD axions such as HAYSTAC or ADMX might be sensitive:

$$P(\mathcal{A}_{i,\mathrm{tot}}) = P(\mathrm{PQ}) \times P(\mathrm{DM}) \times P(m_a \approx h\nu_i/c^2), \quad \text{(F3)}$$

where PQ denotes the event of the Peccei-Quinn hypothesis [50,51] being correct and the QCD axion field existing, DM denotes the event of that axion field actually accounting for an appreciable fraction of the galaxy's dark matter [52–55], $m_a \approx h\nu_i/c^2$ denotes the event that the axion is within roughly a linewidth of the bin of interest, $i$, and the probability of each event in parentheses is implicitly conditional upon all leftward events. In practice several other propositions must be true for an axion to be detectable. For instance, the axion must not be coincident with any other rf spikes and its line shape should roughly match that used in the data processing [18,25].

It is beyond the scope of this article to attempt serious estimates of these probabilities, but for the sake of justifying the infinitesimal prior approximation, we only need to consider the sample optimistic set of probabilities shown in Table II. For logarithmically uniform distributed priors [2]

$$P_{a,i} \propto \frac{1}{\nu_i} \quad \text{(F4)}$$

in a $Q_a \sim 10^6$ axion between $m_{\mathrm{SN}} \sim 1\ \mu$eV and $m_{\mathrm{OC}} \sim 1$ meV, the approximate mass range not disfavored by evidence from the SN1987a neutrino burst [56] or overclosure arguments [52–54], $P(m_a \approx h\nu_i/c^2) = \log(1+Q_a)/\log(m_{\mathrm{SN}}/m_{\mathrm{OC}}) = 1.4 \times 10^{-7}$. If overclosure arguments do not apply [57], then a QCD axion could be far lighter than $m_{\mathrm{OC}}$. Moreover, in practice, the prior probability for an axion to be in any given narrow coupling window must be less than $P(\mathcal{A}_{i,\mathrm{tot}})$. Hence, the rough estimate of $P(\mathcal{A}_{i,\mathrm{tot}}) \sim 6 \times 10^{-8}$ should be considered a generous upper bound on the prior probability as the term is used in this article.

TABLE II. Sample set of probabilities used to estimate prior probability of an axion according to Eq. (F3). The probabilities $P(\mathrm{PQ})$ and $P(\mathrm{DM})$ are set to high (optimistic) values, such that raising them would not qualitatively change the conclusion that the priors are in the appropriate infinitesimal limit. The probability of the axion mass coinciding with a given bin assumes logarithmic priors [2], and is itself optimistic in assuming that the axion exists within a 3 mass decade window in which the haloscope operates.

| Event | Optimistic probability |
|---|---|
| PQ | 0.8 |
| DM | 0.5 |
| $m_a \approx h\nu_i/c^2$ | $1.4 \times 10^{-7}$ |
| $\mathcal{A}_{i,\mathrm{tot}}$ | $6 \times 10^{-8}$ |

In practice, one should not wait for a prior update of order $1/P(\mathcal{A}_{i,\mathrm{tot}})$ to manually interrogate a bin[17]; instead, an approximate prior estimate of the prevalence of spurious rf tones (Appendix E) should inform decision making. The prevalence of these rf spikes is *a priori* unknown, but the fact that in practice [18,22,26] the number of rescans required agrees with the predictions from Gaussian statistics (i.e., the measured distribution of excess powers looks Gaussian even in its right tail) indicates that only a few rf spikes large enough to push a bin above threshold are expected in a scan. For an initial scan with $N_I$ independent bins, therefore, a prior update of $U \sim N_I/N_{\mathrm{rf}}$ should be sufficient cause to manually interrogate the bin. Here, $N_{\mathrm{rf}}$ is the experimentalist's best guess for the expected number of real rf spikes in the scan window. Insofar as this guess is incorrect, the haloscope search will take longer than would be optimal. Rescanning and eliminating spurious rf tones ensures that the bias in the updates remains small, as well as conservative.[18]

The prior probability $P(\mathcal{A}_{i,\mathrm{tot}})$ estimated in this Appendix makes our approximation of infinitesimal priors valid for any single-bin updates $U \ll 1/P(\mathcal{A}_{i,\mathrm{tot}})$. Since the prior update at which manual interrogation ought to begin, $N_I/N_{\mathrm{rf}}$, is generally orders of magnitude less than $1/P(\mathcal{A}_{i,\mathrm{tot}})$, the approximation of infinitesimal priors will always be valid.

[17]Manual interrogation means performing tests that would discriminate between most spurious rf excesses and an axion: for example tuning the spike well off cavity resonance and/or ramping down the magnet and seeing if the signal persists.

[18]The conservative bias discussed in Appendix E results from interpreting the no axion hypothesis, $\mathcal{N}_i$ in Eq. (F1), to incorporate spurious rf tones. This makes $P(x_i^{(j)}|\mathcal{N}_i)$ larger in reality than as calculated from Eq. (1). Using a slightly undersized denominator in Eq. (9) results in a slight antiexclusion bias.

## APPENDIX G: EFFECTS OF DATA PROCESSING ON INFORMATION CONTENT

The main text of this article takes as its motivation the unacceptably long scan times that will likely be required to verify or falsify the axionic dark matter hypothesis over any meaningful fraction of the plausible parameter space. However, our framework optimizes only the statistical analysis of the already-processed data of a specific measurement, raising two complementary questions: is the measurement optimal, and is the data processing optimal? The first question, of whether the measurement itself is optimal—i.e., is the haloscope as typically operated the most efficient tool for evaluating the axionic dark matter hypothesis?—is beyond the scope of this article, and we refer the reader to Ref. [2]. Subject to a haloscope or haloscope-like search platform, however, this Appendix addresses the effect of the data processing—the intermediate set of steps between measurement and statistical analysis—on the information content of the haloscope data. We show that the current data processing protocol of the leading haloscope experiments [18,26] are already highly optimized, with very little room for further improvements, though this has not historically been the case.

We define the data processing as any numerical manipulation of a measured quantity from the moment of its digitization until it is put into a statistical analysis framework [16,22]. The processing of real haloscope data involves many steps for dealing with practical nonidealities that are out of the scope of this Appendix. We do not address issues such as the identification and removal of persistent but nonaxionic spectral features, or the digital high-pass filtering of the data. For a thorough treatment of the HAYSTAC experiment's processing protocol[19] including all these steps, see Ref. [22]. This Appendix evaluates only the mathematical operations as performed on an ideal dataset where the goal is to determine the presence or absence of an axion at each frequency.

The data that may contain an axionic imprint arrives as a pair (one for each quadrature)[20] of discretely sampled, identically distributed time streams of measured voltages. Many such time streams are obtained, and they are broken up and discretely Fourier transformed[21] into so-called subspectra, with frequency bin spacing $\Delta_b$ roughly 2 orders of magnitude narrower than the axion linewidth $\Delta_a$. The subspectra are acquired in batches of $N_b$ at each of $N_t$ distinct haloscope tuning steps. The Fourier frequencies in each subspectrum can be mapped faithfully back to the $N_f$ rf frequencies[22] of the fluctuations within the haloscope cavity. Ultimately, each rf frequency is associated with a single bin in each of a large number of distinct subspectra. Therefore, the complex subspectrum voltages are denoted $v^{\alpha}_{Q,ijk}$, where $Q = X$ or $Y$ denotes the quadrature, $i = 1, \ldots, N_f$ indexes the rf frequency, $j = 1, \ldots, N_t$ the haloscope tuning step, and $k = 1, \ldots, N_b$ the subspectrum at a given tuning step.

At each of these tuning steps, sensitivity data are acquired as well. The sensitivity data, in a simplified picture, fully specify the expected thermal noise and possible axion signal power at each frequency bin $i$. At each tuning step and for each frequency, the sensitivity parameters $\eta^{\gamma}_{ij}$, defined as the ratio of signal power to noise power for a axion with $g_\gamma = 1$ that delivers 100% of its power at the $i$th rf frequency, can then be calculated.

This Appendix evaluates the consequences of each data processing step on the axion-pertinent (AP) information content of the data, where AP is understood to mean "potentially bearing on the probability of the axion hypothesis being true." Using the terminology of Ref. [22], the subspectrum data are processed ultimately into grand spectrum data through the following steps:

(1) The real and imaginary parts of each subspectrum quadrature voltage are summed in quadrature so as to obtain the quadrature powers $x^{\alpha}_{Q,ijk}$ within each bin:

$$x^{\alpha}_{Q,ijk} = (\Im[v^{\alpha}_{Q,ijk}])^2 + (\Re[v^{\alpha}_{Q,ijk}])^2. \quad \text{(G1)}$$

(2) The quadrature powers are summed to obtain the total power $x^{\alpha}_{ijk}$ in each subspectrum bin:

$$x^{\alpha}_{ijk} = x^{\alpha}_{X,ijk} + x^{\alpha}_{Y,ijk}. \quad \text{(G2)}$$

(3) All subspectrum powers at each tuning step and frequency are summed to obtain the raw spectrum powers:

$$x^{\beta}_{ij} = \sum_{k=1}^{N_b} x^{\alpha}_{ijk}. \quad \text{(G3)}$$

(4) The expected noise power, obtained from the application of a Savitzky-Golay (SG) filter [58,59] at each frequency, is subtracted from each raw spectrum to obtain the processed spectra:

[19]The HAYSTAC processing protocol is in its essentials the same as that presently used by ADMX [18].

[20]It is theoretically equally efficient to perform single-quadrature measurement [3], in which case there is only one quadrature time stream, but none of the conclusions of this Appendix are altered.

[21]While in principle it is possible to evaluate the axionic dark matter hypothesis in the time domain, we do not expect that this would provide any benefit, given the unitarity of the Fourier transform and the fact that the phase of the axion field at any moment is unknown.

[22]The number of rf frequencies in earlier processing steps $N_f$ is generally larger than $N$, the number of frequencies in the grand spectrum, as a later data processing step will average sets of spectrally adjacent bins.

$$x_{ij}^{\gamma} = x_{ij}^{\beta} - \langle x_{ij}^{\beta} \rangle. \tag{G4}$$

The processed spectra are power excesses: that is, departures of measured powers from the values expected absent an axion.

(5) The processed spectra are divided by the independently obtained sensitivities to obtain the rescaled spectra:

$$x_{ij}^{\delta} = x_{ij}^{\gamma} / \eta_{ij}^{\gamma}. \tag{G5}$$

The rescaled spectra have the property that an axion with $g_\gamma = 1$ unrealistically depositing all of its power entirely within any one bin shifts the mean power excess for that bin up from 0 to 1. Less axion-sensitive bins are therefore drawn from distributions with larger variances $(\sigma_{ij}^{\delta})^2$.

(6) In each rf frequency bin $i$, the combined spectrum power excess $x_i^{\epsilon}$ is taken as the ML-weighted sum of all rescaled spectra power excesses:

$$\frac{x_i^{\epsilon}}{(\sigma_i^{\epsilon})^2} = \sum_{j=1}^{N_t} \frac{x_{ij}^{\delta}}{(\sigma_{ij}^{\delta})^2}. \tag{G6}$$

The variance of the $i$th combined spectrum bin is obtained from the variance of all contributing rescaled spectrum bins as

$$(\sigma_i^{\epsilon})^2 = \left( \sum_{j=1}^{N_t} \frac{1}{(\sigma_{ij}^{\delta})^2} \right)^{-1}, \tag{G7}$$

where $(\sigma_{ij}^{\delta})^2 = \infty$ at tuning steps that do not contribute to a given frequency bin $i$.

(7) Sets of $n_c \ll \Delta_a/\Delta_b$ adjacent combined spectrum power excesses are rescaled, and the rebinned spectrum power excesses $x_i^{\zeta}$ are formed from their ML-weighted sums:

$$\frac{x_i^{\zeta}}{(\sigma_i^{\zeta})^2} = \frac{1}{n_c n_r} \sum_{l=1}^{n_c} \frac{x_{n_c(i-1)+l}^{\epsilon}}{(\sigma_{n_c(i-1)+l}^{\epsilon})^2}, \tag{G8}$$

where only every $n_c$th rebinned power excess is calculated and $n_r$ is the number of rebinned bins that contribute to each grand spectrum bin (step 8). The rebinned variances are the analogously scaled, ML-estimated variances given by their combined spectrum counterparts:

$$(\sigma_i^{\zeta})^2 = \frac{1}{(n_c n_r)^2} \left( \sum_{l=1}^{n_c} \frac{1}{(\sigma_{n_c(i-1)+l}^{\epsilon})^2} \right)^{-1}. \tag{G9}$$

The rebinned bin frequencies are taken as the average of the contributing combined bin frequencies. For the rebinned spectrum and the subsequent grand spectrum, $i$ now indexes frequency in the more sparsely populated space.

(8) The axion line shape $f_a(\nu|\nu_a)$ given an axion with rest mass $\nu_a$ is calculated as the PDF of axion particle energies given their Maxwellian velocity distribution in the galactic rest frame, and our planet's velocity through that rest frame [60]. This PDF is then discretized into a PMF $\bar{p}_a(\nu_j|\nu_{a,i})$ by integration over rebinned spectrum bins.[23] The grand spectrum is then constructed. The grand spectrum has the same frequency spacing as the rebinned spectrum, but each grand spectrum bin $i$ includes contributions from the $n_r = \Delta_a/n_c\Delta_b$ rebinned spectrum bins beginning at $i$. Sets of $2n_r + 1$ grand spectrum bins are thus by construction correlated (Appendix D). The grand spectrum power excesses $x_i^{\eta}$ are constructed from the sets of $n_r$ adjacent rebinned power excesses as

$$\frac{x_i^{\eta}}{(\sigma_i^{\eta})^2} = \sum_{l=1}^{n_r} \frac{\bar{p}_a(\nu_{i+l-1}|\nu_{a,i}) x_{i+l-1}^{\zeta}}{(\sigma_{i+l-1}^{\zeta})^2}, \tag{G10}$$

where $\nu_{i+l-1}$ denote the rebinned frequencies. Equation (G10) provides the ML estimate of the power delivered by an axion to the $i$th grand spectrum bin, which has variance given by

$$(\sigma_i^{\eta})^2 = \left( \sum_{l=1}^{n_r} \left[ \frac{\bar{p}_a(\nu_{i+l-1}|\nu_{a,i})}{\sigma_{i+l-1}^{\zeta}} \right]^2 \right)^{-1}. \tag{G11}$$

(9) If the combined spectrum bins did not contain correlations from the SG filtering (step 4), then the division by the independently obtained sensitivities (step 5) would ensure that $x_i^{\eta}/\sigma_i^{\eta}$ is drawn from a standard normal distribution (no axion) or a normal distribution with unit mean and standard deviation (axion with $g_\gamma = 1$). The negative correlations imprinted by the SG filter (Appendix D), however, reduce the grand spectrum standard deviations $\sigma_i^{\eta}$ by a spectrally uniform factor $\xi$,[24] which can be extracted directly from the data and validated through simulation.

Scaling the grand spectrum standard deviations down to

$$\tilde{\sigma}_i^{\eta} = \xi \sigma_i^{\eta} \tag{G12}$$

makes the corrected grand spectrum excesses

[23] This conversion is made difficult by the fact that the axion frequency will in general not align perfectly with any bin frequency, an effect that matters less the smaller the rebinned bin spacing $n_c\Delta_b$ is. Reference [22] accounts for this uncertainty.

[24] For phase 1 of HAYSTAC, $\xi = 0.93$ [22].

$$x_i^{(1)} = x_i^{\eta}/\tilde{\sigma}_i^{\eta} \tag{G13}$$

standard normal random variables absent an axion. The superscript (1) denotes that steps 1–9 have applied for the initial scan, and must be repeated for rescans. Equation (G13) represents the grand spectrum excesses referred to in the main text.

The SG filter has the additional effect of attenuating the visibility of a potentially present axion. The frequency-independent magnitude of this attenuation can be simulated as well, and is captured along with other, frequency-dependent effects, in the sensitivity parameters $\eta_i^{(1)}$ of Eq. (2).

(10) Steps 1–9 are repeated, possibly multiple times, at frequencies where rescans are deemed appropriate.

Information pertinent to the existence of the axion can only be generated during the measurement, and may be degraded by the subsequent data processing. Our aim is to determine where in steps 1–10, if at all, AP information could be lost. A guaranteed way to make use of all of the AP information content would therefore be to express the predictions of the axion and no-axion hypotheses, $\mathcal{A}_{Q,ijk}$ and $\mathcal{N}_{Q,ijk}$, respectively, in the space of the quadrature voltages $v_{Q,ijk}^{\alpha}$ and apply Bayes' theorem as

$$u_{Q,ijk}^{\alpha,\Re} = \frac{P(\mathcal{A}_{Q,ijk}|\Re[v_{Q,ijk}^{\alpha}])}{P(\mathcal{N}_{Q,ijk}|\Re[v_{Q,ijk}^{\alpha}])}, \tag{G14}$$

where the standard infinitesimal prior limit has been assumed and the same update applies for the imaginary components $\Im[v_{Q,ijk}^{\alpha}]$. While this approach is impractical, it nonetheless serves as an excellent baseline to compare the effects of the data processing protocol against.

Step 1 includes two operations: squaring the real and imaginary parts of each quadrature voltage, and then adding them. The real and imaginary parts of each quadrature voltage can be approximated as the same independent, identically distributed Gaussian random variables with mean zero [31,61]. Absent an axion, we label the variances of these random variables $\sigma_{v,ij}^2$. An axion, assumed to be of a given coupling and at a given frequency, increases the variance to $\lambda_{ij}\sigma_{v,ij}^2$, where $\lambda_{ij}$ is extremely close to but greater than unity. The prior update, Eq. (G14), is the ratio of Gaussian PDFs given by Eq. (1):

$$u_{Q,ijk}^{\alpha,\Re} = \frac{1}{\sqrt{\lambda_{ij}}}\exp\left[\frac{1-\lambda_{ij}^{-1}}{2\sigma_{v,ij}^2}(\Re[v_{Q,ijk}^{\alpha}])^2\right]. \tag{G15}$$

Since $\Re[v_{Q,ijk}]$ only occurs squared in this prior update (likewise $\Im[v_{Q,ijk}]$ in its identical equation), no AP information is lost by squaring the real and imaginary quadrature voltage components. Furthermore, since Bayesian updates are by their nature combined multiplicatively, the update accounting for both components of the quadrature voltage is

$$u_{Q,ijk}^{\alpha,v} = \frac{1}{\lambda_{ij}}\exp\left[\frac{1-\lambda_{ij}^{-1}}{2\sigma_{v,ij}^2}((\Re[v_{Q,ijk}^{\alpha}])^2 + (\Im[v_{Q,ijk}^{\alpha}])^2)\right]. \tag{G16}$$

Because the voltage quadrature components appear in the prior update only in their Eq. (G1) combination, step 1 preserves AP information content.

The same argument applies to steps 2 and 3. Multiplicative prior updates have their powers add in the exponent, hence steps 2 and 3 do not degrade AP information content. For step 2, this argument assumes that the axion and the thermal noise fluctuations distribute power evenly between the $X$ and $Y$ quadratures. For a two-quadrature measurement where the phase of the axion is unknown, that assumption is valid.

Step 4 is among the more vulnerable steps to meaningful information loss. If $\langle x_{ij}^{\beta}\rangle$ in Eq. (G4) is treated as a number, then Eq. (G4) amounts merely to the subtraction of a constant from the raw spectrum random variables $x_{ij}^{\beta}$, which will not degrade the information content. In practice, however, $\langle x_{ij}^{\beta}\rangle$ must itself be estimated, making it in effect a random variable. The estimation used by HAYSTAC and ADMX is performed via SG filters, which estimate $\langle x_{ij}^{\beta}\rangle$ at each frequency bin $i$ via the polynomial generalization of a moving average. If a wide spectral window for the generalized moving average is used, then $\langle x_{ij}^{\beta}\rangle$ is estimated with low variance, and acts like a constant. However, the SG filtering creates undesired correlations between bins up to two window-lengths apart, while also slightly attenuating an axion's visibility, an effect accounted for in step 9. The trade-off between the desired filtering effects and the undesired correlations and attenuation is beyond the scope of this article, but is discussed at length in Ref. [22]. So far, no proof exists that SG is the optimal high-pass digital filter for the data, and we therefore identify step 4 as potentially admitting of meaningful optimization.

Calculating the processed spectrum sensitivity parameters $\eta_{ij}^{\gamma}$ of step 5 relies on independent measurements of the thermal and added noise, magnetic field, and cavity quality factor and mode-structure properties of the haloscope, as well as calculations of the power that a dark matter axion would deliver. Subject to the accuracy of these measurements and calculations, step 5 simply divides the processed spectra power excess random variables $x_{ij}^{\gamma}$ by constants, preserving information content.

The ML weighting of step 6 is the information-preserving generalization of the straightforward addition of steps 1–3, for the case where the random variables being added have different variances. By the central limit theorem, the rescaled power excesses $x_{ij}^{\delta}$ are Gaussian

distributed with known variances $\sigma_{ij}^{\delta}$ well approximated as independent of the presence of any axion which delivers power far less than vacuum. The pertinent effect of an axion is to shift the mean of the rescaled power excess from 0 to $\mu^{\delta}$, which has no bin or tuning step dependence by construction. The appropriate prior update delivered by each rescaled power excess $x_{ij}^{\delta}$ is thus

$$u_{ij}^{\delta} = \exp\left[\frac{-(\mu^{\delta})^2/2 + \mu^{\delta} x_{ij}^{\delta}}{(\sigma_{ij}^{\delta})^2}\right]. \tag{G17}$$

When all $N_t$ tuning steps' updates are multiplied, the result is

$$\begin{aligned}\prod_{j=1}^{N_t} u_{ij}^{\delta} &= \exp\left[\sum_{j=1}^{N_t}\left(\frac{-(\mu^{\delta})^2/2 + \mu^{\delta} x_{ij}^{\delta}}{(\sigma_{ij}^{\delta})^2}\right)\right] \\ &= \exp\left[\frac{-(\mu^{\delta})^2/2 + \mu^{\delta} x_i^{\epsilon}}{(\sigma_i^{\epsilon})^2}\right].\end{aligned} \tag{G18}$$

Because the updates from all bin-$i$ rescaled spectra power excesses $x_{ij}^{\delta}$ and variances $(\sigma_{ij}^{\delta})^2$ can thus be obtained using the single bin-$i$ combined spectrum power excess $x_i^{\epsilon}$ and variance $(\sigma_i^{\epsilon})^2$, given by Eqs. (G6) and (G7), respectively, the ML estimation of step 6 preserves AP information content.

The same argument about ML estimation guarantees that steps 7 and 8 preserve AP information as well, in the limit where the axion line shape is truly approximately constant over spectral scales of $n_c\Delta_b$. Step 9 is in effect properly accounting for the information loss suffered during step 4, but is itself simply multiplication by a scalar, and causes no additional information loss.

In summary, the use of ML estimation, not consistently applied in previous haloscope analyses [16], makes the grand spectrum excesses $x_i^{(j)}$ together with the sensitivity parameters $\eta_i^{(j)}$ nearly sufficient statistics for evaluating the cold dark matter axion hypothesis. We identify only two areas where the sensitivity can be improved. First, it is possible that a high-pass filter with lower stop-band attenuation that maintains acceptable pass-band performance would improve the processing sensitivity. Second, the rationale behind the rebinning (step 7)—that it will reduce the expected number of axion candidates exceeding threshold, mitigating the look-elsewhere effect—is unnecessary in the Bayesian power measured framework. By not rebinning (setting $n_c \to n_c' = 1$ and $n_r \to n_r' = n_c n_r$), the curvature of the expected axion line shape over $n_c\Delta_b$-sized spectral scales is optimally accounted for. The expected $n_c$-fold growth of such candidates introduced by not rebinning is precisely negated in the aggregate prior update, Eq. (11), by the $n_c$-fold growth in the number $N$ of grand spectrum bins.[25]

## APPENDIX H: ADDITIONAL FEATURES OF THE HAYSTAC PHASE 1 DATASET

In addition to the LU candidate at $\nu_{LU}^{\star}$ discussed in Sec. IV, the dashed, blue curve of Fig. 3(b) also neglects two other features of the HAYSTAC phase 1 dataset that appear in Fig. 3(a). First, the BPM reanalysis includes data from rescans over the upper 100 MHz of the scan window that were performed during phase 1, run 1 of HAYSTAC [25]. These rescans were discarded, and new ones performed over some of the same frequencies, because an error in estimating the expected axion line shape lowered their sensitivities below the levels required by the predetermined confidence levels and initial scan sensitivities [22]. The extra rescan data itself is valid, however, and is thus included in Figs. 3(a) and 3(b), showcasing the flexibility of BPM to integrate all available information. We do not include this extra data in the comparison (dashed blue) curve, as the existence of extraneous rescan data is not ordinary in haloscope analyses.

Second, a $3.94\sigma$ grand spectrum excess at grand spectrum bin 550 ($\nu_{550} = 5.59886155$) GHz was recorded in the initial scan. This excess occurs 0.3% of the way into the scan window, where the number of contributing raw spectra (Appendix G) was 14, considerably less than the typical 40. The lower number of spectra implies at once an increased susceptibility to systematic error and a reduced axion sensitivity $\eta_{550}^{(1)} = 0.23$. The reduced sensitivity implies that an axion here has its prior update, Eq. (9), maximized at $g_{\gamma} \approx 4.1$. This excess ought to have been rescanned during phase 1 of HAYSTAC, but was not. As such, we include its positive prior update in the BPM reanalysis. Its presence is indicated by the dark column at the left of Fig. 3(a), and is solely responsible for the failure of the aggregate prior update (solid blue line, Fig. 3) to vanish at high couplings $\gtrsim 3.2|g_{\mathrm{KSVZ}}|$. Had the required rescans been performed, a source of excess would likely have been discovered or rejected in short order, as in the case of the LU candidate at $\nu_{LU}^{\star}$. As such, we remove it from the comparison (dashed blue) aggregate update curve.

[25]Future analyses may on the same principle set grand spectrum bin spacings closer than those of the combined spectrum. Doing so eliminates the need to account for misalignment (footnote 23) of the axion rest mass with grand spectrum bin frequency.

[1] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, Experimental searches for the axion and axion-like particles, Annu. Rev. Nucl. Part. Sci. **65**, 485 (2015).

[2] S. Chaudhuri, K. Irwin, P. W. Graham, and J. Mardon, Fundamental limits of electromagnetic axion and hidden-photon dark matter searches: Part I–The quantum limit, arXiv:1803.01627.

[3] M. Malnou, D. A. Palken, B. M. Brubaker, L. R. Vale, G. C. Hilton, and K. W. Lehnert, Squeezed Vacuum used to Accelerate the Search for a Weak Classical Signal, Phys. Rev. X **9**, 021023 (2019).

[4] M. Mück, M.-O. André, J. Clarke, J. Gail, and C. Heiden, Radio-frequency amplifier based on a niobium dc superconducting quantum interference device with microstrip input coupling, Appl. Phys. Lett. **72**, 2885 (1998).

[5] M. A. Castellanos-Beltran and K. W. Lehnert, Widely tunable parametric amplifier based on a superconducting quantum interference device array resonator, Appl. Phys. Lett. **91**, 083509 (2007).

[6] M. Malnou, D. A. Palken, L. R. Vale, G. C. Hilton, and K. W. Lehnert, Optimal operation of a Josephson parametric amplifier for vacuum squeezing, Phys. Rev. Applied **9**, 044023 (2018).

[7] W. F. Kindel, M. D. Schroer, and K. W. Lehnert, Generation and efficient measurement of single photons from fixed-frequency superconducting qubits, Phys. Rev. A **93**, 033817 (2016).

[8] B. R. Johnson, M. D. Reed, A. A. Houck, D. Schuster, L. S. Bishop, E. Ginossar, J. M. Gambetta, L. DiCarlo, L. Frunzio, S. Girvin, and R. J. Schoelkopf, Quantum non-demolition detection of single microwave photons in a circuit, Nat. Phys. **6**, 663 (2010).

[9] H. Zheng, M. Silveri, R. T. Brierley, S. M. Girvin, and K. W. Lehnert, Accelerating dark-matter axion searches with quantum measurement technology, arXiv:1607.02529.

[10] I. Stern, A. A. Chisholm, J. Hoskins, P. Sikivie, N. S. Sullivan, D. B. Tanner, G. Carosi, and K. van Bibber, Cavity design for high-frequency axion dark matter detectors, Rev. Sci. Instrum. **86**, 123305 (2015).

[11] P. Sikivie, Detection rates for "invisible"-axion searches, Phys. Rev. D **32**, 2988 (1985).

[12] C. Hagmann, D. Kinion, W. Stoeffl, K. van Bibber, E. Daw, H. Peng, Leslie J. Rosenberg, J. LaVeigne, P. Sikivie, N. S. Sullivan, D. B. Tanner, F. Nezrick, M. S. Turner, D. M. Moltz, J. Powell, and N. A. Golubev, Results from a High-Sensitivity Search for Cosmic Axions, Phys. Rev. Lett. **80**, 2043 (1998).

[13] S. J. Asztalos, E. Daw, H. Peng, L. J. Rosenberg, D. B. Yu, C. Hagmann, D. Kinion, W. Stoeffl, K. van Bibber, J. LaVeigne, P. Sikivie, N. S. Sullivan, D. B. Tanner, F. Nezrick, and D. M. Moltz, Experimental constraints on the axion dark matter halo density, Astrophys. J. Lett. **571**, L27 (2002).

[14] S. J. Asztalos, R. F. Bradley, L. Duffy, C. Hagmann, D. Kinion, D. M. Moltz, L. J. Rosenberg, P. Sikivie, W. Stoeffl, N. S. Sullivan, D. B. Tanner, K. van Bibber, and D. B. Yu, Improved rf cavity search for halo axions, Phys. Rev. D **69**, 011101(R) (2004).

[15] S. J. Asztalos, G. Carosi, C. Hagmann, D. Kinion, K. van Bibber, M. Hotz, L. J. Rosenberg, G. Rybka, J. Hoskins, J. Hwang, P. Sikivie, D. B. Tanner, R. Bradley, and J. Clarke, SQUID-Based Microwave Cavity Search for Dark-Matter Axions, Phys. Rev. Lett. **104**, 041301 (2010).

[16] S. Asztalos, E. Daw, H. Peng, L. J. Rosenberg, C. Hagmann, D. Kinion, W. Stoeffl, K. van Bibber, P. Sikivie, N. S. Sullivan, D. B. Tanner, F. Nezrick, M. S. Turner, D. M. Moltz, J. Powell, M.-O. André, J. Clarke, M. Mück, and R. F. Bradley, Large-scale microwave cavity search for dark-matter axions, Phys. Rev. D **64**, 092003 (2001).

[17] J. V. Sloan *et al.*, Limits on axion-photon coupling or on local axion density: Dependence on models of the Milky Way's dark halo, Phys. Dark Universe **14**, 95 (2016).

[18] N. Du *et al.* (ADMX Collaboration), Search for Invisible Axion Dark Matter with the Axion Dark Matter Experiment, Phys. Rev. Lett. **120**, 151301 (2018).

[19] T. Braine *et al.* (ADMX Collaboration), Extended Search for the Invisible Axion with the Axion Dark Matter Experiment, Phys. Rev. Lett. **124**, 101303 (2020).

[20] C. Boutan *et al.* (ADMX Collaboration), Piezoelectrically Tuned Multimode Cavity Search for Axion Dark Matter, Phys. Rev. Lett. **121**, 261302 (2018).

[21] N. Crisosto, G. Rybka, P. Sikivie, N. S. Sullivan, D. B. Tanner, and J. Yang, ADMX SLIC: Results from a superconducting LC circuit investigating cold axions, arXiv:1911.05772.

[22] B. M. Brubaker, L. Zhong, S. K. Lamoreaux, K. W. Lehnert, and K. A. van Bibber, HAYSTAC axion search analysis procedure, Phys. Rev. D **96**, 123008 (2017).

[23] B. M. Brubaker *et al.*, First Results from a Microwave Cavity Axion Search at 24 $\mu$eV, Phys. Rev. Lett. **118**, 061302 (2017).

[24] S. Al Kenany, M. A. Anil, 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, D. A. Palken, N. M. Rapidis, J. R. Root, M. Simanovskaia, T. M. Shokair, I. Urdinaran, K. A. van Bibber, and L. Zhong, Design and operational experience of a microwave cavity axion detector for the 20–100 $\mu$eV range, Nucl. Instrum. Methods Phys. Res., Sect. A **854**, 11 (2017).

[25] B. M. Brubaker, First results from the HAYSTAC axion search, Ph.D. Thesis, Yale University, 2017.

[26] L. Zhong *et al.*, Results from phase 1 of the HAYSTAC microwave cavity axion experiment, Phys. Rev. D **97**, 092001 (2018).

[27] M. Dine, W. Fischler, and M. Srednicki, A simple solution to the strong *CP* problem with a harmless axion, Phys. Lett. B **104**, 199 (1981).

[28] A. R. Zhitnitsky, On possible suppression of the axion hadron interactions, Sov. J. Nucl. Phys. **31**, 260 (1980).

[29] Y. Kahn, B. R. Safdi, and J. Thaler, Broadband and Resonant Approaches to Axion Dark Matter Detection, Phys. Rev. Lett. **117**, 141801 (2016).

[30] J. L. Ouellet, C. P. Salemi, J. W. Foster, R. Henning, Z. Bogorad, J. M. Conrad, J. A. Formaggio, Y. Kahn, J. Minervini, A. Radovinsky, N. L. Rodd, B. R. Safdi, J. Thaler, D. Winklehner, and L. Winslow, First Results from ABRACADABRA-10 cm: A Search for sub-$\mu$eV Axion Dark Matter, Phys. Rev. Lett. **122**, 121802 (2019).

[31] 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).
[32] M. Silva-Feaver, S. Chaudhuri, H. Cho, C. Dawson, P. Graham, K. Irwin, S. Kuenstner, D. Li, J. Mardon, H. Moseley, R. Mule, A. Phipps, S. Rajendran, Z. Steffen, and B. Young, Design overview of DM radio pathfinder experiment, IEEE Trans. Appl. Supercond. **27**, 1 (2017).
[33] 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).
[34] A. Caldwell, G. Dvali, B. Majorovits, A. Millar, G. Raffelt, J. Redondo, O. Reimann, F. Simon, and F. Steffen (The MAMDAX Working Group), Dielectric Haloscopes: A New Way to Detect Axion Dark Matter, Phys. Rev. Lett. **118**, 091801 (2017).
[35] G. Rybka, A. Wagner, K. Patel, R. Percival, K. Ramos, and A. Brill, Search for dark matter axions with the Orpheus experiment, Phys. Rev. D **91**, 011701(R) (2015).
[36] S. Lee, S. Ahn, J. Choi, B. R. Ko, and Y. K. Semertzidis, Axion Dark Matter Search Around 6.7 $\mu$eV, Phys. Rev. Lett. **124**, 101802 (2020).
[37] D. Alesini *et al.*, Galactic axions search with a superconducting resonant cavity, Phys. Rev. D **99**, 101101(R) (2019).
[38] S. DePanfilis, A. C. Melissinos, B. E. Moskowitz, J. T. Rogers, Y. K. Semertzidis, W. U. Wuensch, H. J. Halama, A. G. Prodell, W. B. Fowler, and F. A. Nezrick, Limits on the Abundance and Coupling of Cosmic Axions at $4.5 < m_a < 5.0$ $\mu$eV, Phys. Rev. Lett. **59**, 839 (1987).
[39] W. U. Wuensch, S. DePanfilis-Wuensch, Y. K. Semertzidis, J. T. Rogers, A. C. Melissinos, H. J. Halama, B. E. Moskowitz, A. G. Prodell, W. B. Fowler, and F. A. Nezrick, Results of a laboratory search for cosmic axions and other weakly coupled light particles, Phys. Rev. D **40**, 3153 (1989).
[40] C. Hagmann, P. Sikivie, N. S. Sullivan, and D. B. Tanner, Results from a search for cosmic axions, Phys. Rev. D **42**, 1297 (1990).
[41] C. Hagmann, A search for cosmic axions, Ph. D. thesis, University of Florida, 1990.
[42] L. A. Wainstein and V. D. Zubakov, in *Extraction of Signals from Noise* (Dover, New York, 1962), pp. 142–153.
[43] W. W. Harman, in *Principles of the Statistical Theory of Communication* (McGraw-Hill, New York, 1963), pp. 217–248.
[44] L. Lista, *Statistical Methods for Data Analysis in Particle Physics*; 2nd ed. (Springer, New York, 2017).
[45] J. E. Kim, Weak-Interaction Singlet and Strong *CP* Invariance, Phys. Rev. Lett. **43**, 103 (1979).
[46] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Can confinement ensure natural *CP* invariance of strong interactions?, Nucl. Phys. **B166**, 493 (1980).
[47] T. S. Roussy, D. A. Palken, W. B. Cairncross, B. M. Brubaker, D. N. Gresh, M. Grau, K. C. Cossel, K. B. Ng, Y. Shagam, Y. Zhou, V. V. Flambaum, J. Ye, and E. A. Cornell, Experimental constraint on axionlike particle coupling over 7 orders of magnitude in mass (to be published).
[48] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*; 3rd ed. (McGraw-Hill, New York, 2003).
[49] F. Beaujean, A. Caldwell, and O. Reimann, Is the bump significant? An axion-search example, Eur. Phys. J. C **78**, 793 (2018).
[50] R. D. Peccei and H. R. Quinn, *CP* Conservation in the Presence of Pseudoparticles, Phys. Rev. Lett. **38**, 1440 (1977).
[51] R. D. Peccei and H. R. Quinn, Constraints imposed by *CP* conservation in the presence of pseudoparticles, Phys. Rev. D **16**, 1791 (1977).
[52] J. Preskill, M. B. Wise, and F. Wilczek, Cosmology of the invisible axion, Phys. Lett. B **120**, 127 (1983).
[53] L. F. Abbott and P. Sikivie, A cosmological bound on the invisible axion, Phys. Lett. B **120**, 133 (1983).
[54] M. Dine and W. Fischler, The not-so-harmless axion, Phys. Lett. B **120**, 137 (1983).
[55] M. S. Turner, Cosmic and local mass density of "invisible" axions, Phys. Rev. D **33**, 889 (1986).
[56] G. G. Raffelt, in *Axions: Theory, Cosmology, and Experimental Searches*, Lecture Notes in Physics Vol. 741 (Springer, New York, 2008), pp. 51–71.
[57] P. W. Graham and S. Rajendran, New observables for direct detection of axion dark matter, Phys. Rev. D **88**, 035023 (2013).
[58] A. Savitzky and M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. **36**, 1627 (1964).
[59] A. T. Malagon, Search for 140 micro-eV pseudoscalar and vector dark matter using microwave cavities, Ph.D. Thesis, Yale University, 2014.
[60] M. S. Turner, Periodic signatures for the detection of cosmic axions, Phys. Rev. D **42**, 3572 (1990).
[61] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, Rev. Mod. Phys. **82**, 1155 (2010).