Untargeted Bayesian search of anisotropic gravitational-wave backgrounds through the analytical marginalization of the posterior

Adrian Ka-Wai Chung[*] and Nicolás Yunes

Illinois Center for Advanced Studies of the Universe and Department of Physics,
University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

We develop a method to perform an untargeted Bayesian search for anisotropic gravitational-wave backgrounds that can efficiently and accurately reconstruct the background intensity map. Our method employs an analytic marginalization of the posterior of the spherical-harmonic components of the intensity map, without assuming the background possesses any specific angular structure. The key idea is that the likelihood function of the spherical-harmonic components is a multivariate Gaussian when the intensity map is expressed as a linear combination of the spherical-harmonic components and the noise is stationary and Gaussian. If a uniform and wide prior of these spherical-harmonic components is prescribed, the marginalized posterior and the Bayes factor can be well approximated by a high-dimensional Gaussian integral. The analytical marginalization allows us to regard the spherical-harmonic components of the intensity map of the background as free parameters and to construct their individual marginalized posterior distribution in a reasonable time, even though many spherical-harmonic components are required. The marginalized posteriors can, in turn, be used to accurately construct the intensity map of the background. By applying our method to mock data, we show that we can recover precisely the angular structures of various simulated anisotropic backgrounds, without assuming prior knowledge of the relation between the spherical-harmonic components predicted by a given model. Our method allows us to bypass the time-consuming numerical sampling of a high-dimensional posterior, leading to a more model-independent and untargeted Bayesian measurement of the angular structures of the gravitational-wave background.

I. INTRODUCTION

The direct detection of gravitational waves (GWs) emitted by compact binary coalescence (CBC) is a milestone in GW astrophysics [1–10]. The detection of a GW background (GWB), formed by the random and incoherent superposition of numerous individually unresolvable GW signals emitted by different types of sources, may be the milestone that can be achieved next, in the foreseeable future [11–15]. The North American Nanohertz Observatory for Gravitational Waves collaboration has just published $\sim 4\sigma$ significance evidence of the detection of a GWB by analyzing its 15-year dataset [16–19]. This discovery immediately opens up new directions of astronomical research [20,21]. Astrophysical sources, including CBCs [22–25], rapidly rotating asymmetric neutron stars [26–30], and core-collapse supernova [31–34], can generate GWs that form a GWB. Alternatively, a GWB can also be generated by GWs emitted by cosmological sources, like cosmological inflation [35–42], the phase transitions that may have occurred in the early Universe [43–52], and cosmic strings [53–61], if they exist. A GWB

may even be generated by physics that has yet to be fully explored, such as ultralight bosons [62–66] and primordial black holes [67–70], which are candidates to explain dark matter. As a GWB can be formed by sources significantly different from those generating individually detectable GW signals, detecting a GWB constitutes a unique probe of the Universe [15].

While a GWB is expected to be dominantly isotropic, it should also contain angular structures. In general, different sources and generation mechanisms could form GWBs with different angular structures [25,59,71–78]. This source and mechanism dependence suggests that accurately mapping the angular structure of the GWB could be very informative, allowing us to pinpoint GWB sources and deduce their properties [79]. To this end, several methods have been developed to extract the angular distribution of the GWB power spectrum. Broadly speaking, these methods can be classified as either *frequentist* or *Bayesian*. The frequentist approach amounts to constructing some maximum-likelihood estimator with different basis to characterize GWB anisotropies. Examples of the frequentist approach include radiometer search [80] and spherical-harmonic decomposition [81], which have been widely used in analyzing the actual data measured by the LIGO

---
[*]akwchung@illinois.edu

043032-1

and Virgo detectors [82]. The Bayesian approach amounts to constructing the posterior of random variables related to GWB anisotropies, such as done very recently in [83,84].

These two approaches are useful in probing GWB anisotropies, but they also have limitations. For example, since the radiometer search works in the pixel basis, it is not suitable for searching for extended sources [85,86]. Working in the spherical-harmonic basis, one can use a spherical-harmonic decomposition to search for widespread sources and probe the anisotropies in a model-independent way, but it may lead to some nonphysical maximum likelihood estimates, such as complex estimates for some coefficients that, on physical grounds, should be real.

One way to remedy the drawback of the spherical-harmonic decomposition is to perform a model-independent Bayesian search for an anisotropic GWB. However, to describe an anisotropic GWB without assuming any source models, one needs a model-independent framework that is typically characterized by many parameters, such as (the formally infinite number of) the spherical-harmonic components required in the spherical-harmonic decomposition. This large number of variables makes the construction of the posterior of these variables computationally untenable, even if the posterior is estimated through numerical sampling [87]. Thus, the Bayesian search of an anisotropic GWB has been restricted to either a *targeted* Bayesian analysis [83], inferring the overall amplitude of a GWB whose angular structures are given by a specific model, or to a model-independent framework characterized by only a few parameters [84].

The goal of this paper is to develop a computationally efficient, fast and *untargeted* Bayesian search that can construct the Bayesian marginalized posterior of the spherical-harmonic components of the angular structure of the anisotropic GWB without prior knowledge of the relation among the spherical harmonic components. We start with a spherical-harmonic decomposition of the intensity map of the GWB. This decomposition allows us to express the energy flux of the GWB as a function of the sky direction through linear combinations of the spherical harmonic components. If the noise is stationary and Gaussian, then the likelihood function of the spherical-harmonic components is a multivariate Gaussian function. Thus, the marginalized posterior of a specific spherical-harmonic component and the Bayes factor (between an anisotropic GWB and a nondetection hypotheses) can be well approximated by a high-dimensional Gaussian integral. After evaluating this integral, the marginalized posterior of the real or imaginary part of a particular spherical-harmonic component is also a Gaussian function, whose mean and variance are given by the convolution between the cross-spectral density of the data [i.e. the product of the frequency-domain data measured by a detector and the complex conjugate of the frequency-domain data measured by another

detector, see Eq. (22)] and the spherical-harmonic component of the overlap reduction function.

To fully illustrate the power of our analysis, we apply our scheme to mock data containing (i) no GWB signal, (ii) a time-independent dipole GWB signal, and (iii) a GWB formed by Galactic plane binaries. We show that our analysis is capable of extracting the angular structures of all of these signals, despite each type corresponding to different levels of anisotropy. In particular, we show that, in the strong signal-to-noise ratio limit, our analysis can recover an accurate sky map that is almost identical to the simulated Galactic plane signal without bias. Through our Bayes factor calculations, we show that the data can be used to infer the suitable angular length scale of anisotropies that should be included in the analysis.

Our marginalization scheme has several advantages compared to other existing search methods of anisotropic GWBs. First, the analytical formulae derived in this work allow us to reconstruct the intensity map of a GWB extremely accurately and rapidly, and compute the Bayes factors efficiently, completely bypassing the numerical sampling of an extremely high-dimensional posterior, which creates severe computational challenges. Second, since our analysis does not require prior knowledge about the relationship between various spherical-harmonic components, our work represents a major step toward a model-independent search for anisotropic GWBs, which is crucial for understanding the properties of their sources.

The remainder of this paper presents the details of the calculations summarized above, and it is organized as follows. Section II lays the foundation of our analysis by first reviewing the basic properties of GWBs. Section III explains the method we develop and defines different probability distribution functions and hypothesis ranking for the Bayesian search of GWBs. Section IV presents the details of the analytic marginalization and of the evaluation of the Bayes factor. Section V applies the marginalization to mock data. Section VI concludes and points to future research. Throughout this paper, we adopt the following conventions: bold lowercase characters represent a vector; bold uppercase characters represent a square matrix, and their corresponding italic unbolded characters with subscript(s) represent the elements of this matrix. Complex conjugation of a number is denoted by an asterisk. For example, a_i is the ith element of the vector **a**, A_{ij} is the (i, j)th element of the square matrix **A**, and a^* is the complex conjugate of a. Following [11,12,82,88], we take the value of the Hubble constant to be $H_0 = 67.9 \text{ kms}^{-1} \text{Mpc}^{-1}$, which is the *Planck* 2015 value [89], although our conclusions will not depend on this choice. While the analysis presented in this paper can be performed in any coordinate system, in this work we define the sky position $\hat{\Omega}$ and analyze the anisotropy in celestial coordinates (in right accession and declination).

II. PROPERTIES OF ANISOTROPIC GRAVITATIONAL-WAVE BACKGROUND

In this section, we will briefly review the properties and Bayesian analysis of an anisotropic GWB. Only GW properties that are strictly relevant to our work will be reviewed. We refer the reader to, for example, [43,81,90] for a more detailed and exhaustive review of aniso-tropic GWB.

In general, metric perturbations at a given spacetime position can be written as a sum of contributions coming from all directions in the sky through a plane-wave expansion [90–92],

$$h_{ij}(t,\mathbf{x}) = \sum_{A=+,\times}\int_{-\infty}^{\infty} df \int d^2\hat{\Omega}\,\tilde{h}_A(f,\hat{\Omega}) e_{ij}^A(\hat{\Omega}) e^{-2\pi i f(t-\hat{\Omega}\cdot\mathbf{x})},$$

$$(1)$$

where $A = +$ and $\times$ stand for the GW polarization, $\hat{\Omega}$ is a unit vector pointing in a sky direction, $e_{ij}^A(\hat{\Omega})$ are the GW polarization tensors, and the overhead tilde stands for the Fourier transform. Without loss of generality, the expect-ation value[1] of h_{ij} produced by a GWB is usually assumed to be zero [93],

$$\langle h_{ij}(t,\mathbf{x}) \rangle = 0. \qquad (2)$$

However, the quadratic expectation value of $\tilde{h}_A$ is not zero [90],

$$\langle \tilde{h}_A(f,\hat{\Omega})\tilde{h}_{A'}^*(f',\hat{\Omega}') \rangle = \frac{1}{4}\delta(f-f')\delta^2(\hat{\Omega},\hat{\Omega}')\delta_{AA'}\mathcal{H}(f,\hat{\Omega}),$$

$$(3)$$

where $\delta(\cdot)$ and $\delta^2(\cdot)$ are one- and two-dimensional Dirac delta functions, respectively, $\delta_{AA'}$ is a Kronecker delta, and $\mathcal{H}(f,\hat{\Omega})$ is a function of frequency and the sky position $\hat{\Omega}$ that is related to the one-sided strain power spectral density (PSD) of the GWB via [81,90,91],

$$S_h(f) = \frac{1}{4\pi}\int_{S^2} d\hat{\Omega}\,\mathcal{H}(f,\hat{\Omega}). \qquad (4)$$

In other words, $\mathcal{H}(f,\hat{\Omega})$ characterizes the angular distri-bution of the GWB power in different sky directions. The one-sided PSD is related to the dimensionless energy density (also known as the "spectrum") of the GWB via

$$\Omega_{\rm GW}(f) \equiv \frac{f}{\rho_c}\frac{d\rho_{\rm GW}}{df} = \frac{8\pi^3}{3H_0^2}f^3 S_h(f), \qquad (5)$$

where $d\rho_{\rm GW}$ is the energy density of GWs of frequencies between f and $f + df$, ρ_c is the cosmological critical energy density $(\rho_c = 3H_0^2/8\pi G)$.[2] Closely related to $\Omega_{\rm GW}(f)$ is the power of the GWB per unit frequency per unit solid angle [86,90,94],

$$\mathcal{F}(f,\Theta) = \frac{c^3\pi}{4G}f^2\mathcal{H}(f,\hat{\Omega}), \qquad (6)$$

where c is the speed of light and G is Newton's gravita-tional constant. As one would expect then, $\mathcal{F}(f,\Theta)$ has units of $\mathrm{W\,Hz^{-1}\,sr^{-1}}$.

In general, the spectral content and the anisotropy of a GWB are correlated. However, the correlation may be difficult to measure with existing ground-based detectors, like advanced LIGO, advanced Virgo and KAGRA [82,85,94]. Hence, following past search on anisotropic GWBs [82,85,86,94], we assume [81,83,90,91]

$$\mathcal{H}(f,\hat{\Omega}) = H(f)\mathcal{P}(\hat{\Omega}), \qquad (7)$$

where $H(f)$ represents the spectral shape of the GWB and $\mathcal{P}(\hat{\Omega})$ encapsulates the strength and angular distribution of the intensity of the GWB, a function of the sky position. As in any other GWB search, we need to specify the spectral shape of the GWB, $H(f)$, that we are trying to detect. Within the sensitivity band of ground-based detectors, the energy spectrum of many GWBs can be well approximated by a power law in frequency [11–14], which means we can choose $H(f)$ to also have a power law structure, namely

$$H(f) = H_\alpha(f) = \left(\frac{f}{f_{\rm ref}}\right)^{\alpha-3}, \qquad (8)$$

where $f_{\rm ref}$ is a reference frequency and α is the tilt index. Following the existing search of GWB from the actual data, we will fix α and infer $\mathcal{P}(\hat{\Omega})$. Throughout this paper, we also follow the existing LIGO/Virgo search of a GWB and choose $f_{\rm ref} = 25$ Hz, $\alpha = 0, 2/3$ or 3 [11–14]. The choice of $f_{\rm ref}$ does not affect the rest of the analysis at all because it just provides the overall normalization for $\Omega_{\rm GW}$.

To extract the angular structure of the GWB from data, we perform a spherical-harmonic decomposition to express $\mathcal{P}(\hat{\Omega})$ as a linear combination of (scalar) spherical har-monics $Y_{\ell m}(\hat{\Omega})$,

---

[1]To be more specific, if one assumes ergodicity, then the expectation value is equivalent to the ensemble average, which is also the spatial average or the temporal average, see, e.g. [15,90] for a more detailed review.

[2]Note that the normalization convention of $S_h(f)$ is different from that in some of the literature, like [82,83,85,86,90,94]. Here, we include a factor of $(4\pi)^{-1}$ so that for the monopole part of GWB, we have $S_h(f) = \mathcal{H}_{00}(f)$, where $\mathcal{H}_{00}(f)$ is the monopole part of $\mathcal{H}(f;\hat{\Omega})$.

$$\mathcal{P}(\hat{\Omega}) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{+\ell} \mathcal{P}_{\ell m} Y_{\ell m}(\hat{\Omega}), \qquad (9)$$

where $\mathcal{P}_{\ell m}$ are referred to as *spherical-harmonic components* of the spectrum of the GWB, in units of strain^2 Hz^{-1} rad^{-1}. In principle, this sum must include an infinite number of ℓ terms, but in practice, one must truncate the sum at some $\ell = \ell_{\max}$. The value of $\ell_{\max}$ will be specified in subsequent calculations. From Eq. (9), we notice two things. First, upon sky averaging (i.e. integration over sky angle), all terms vanish except $\mathcal{P}_{00}$. This implies that

$$\mathcal{P}_{00} = \frac{S_h(f_{\mathrm{ref}})}{\sqrt{4\pi}} = \frac{3H_0^2}{2\pi^2 f_{\mathrm{ref}}^3} \frac{\Omega_{\mathrm{GW}}(f_{\mathrm{ref}})}{\sqrt{4\pi}}. \qquad (10)$$

Second, the real nature of $\mathcal{P}(\hat{\Omega})$ and $Y_{\ell 0}$ and the complex conjugation property of $Y_{\ell,m}$ imply that

$$Y_{\ell 0} \in \mathbb{R} \Rightarrow \mathcal{P}_{\ell 0} \in \mathbb{R},$$
$$Y_{\ell,-m} = (-1)^m Y_{\ell m}^* \Rightarrow \mathcal{P}_{\ell,-m} = (-1)^m \mathcal{P}_{\ell m}^*. \qquad (11)$$

These requirements imply that, in order to specify the angular distribution of a GWB, we only need $(\ell_{\max} + 1)^2$ real numbers,

$$\mathcal{P}_{00}, \mathcal{P}_{10}, \ldots, \mathcal{P}_{\ell 0},$$
$$\mathcal{P}_{11}^{\mathrm{Re}}, \mathcal{P}_{21}^{\mathrm{Re}}, \ldots, \mathcal{P}_{\ell m}^{\mathrm{Re}},$$
$$\mathcal{P}_{11}^{\mathrm{Im}}, \mathcal{P}_{21}^{\mathrm{Im}}, \ldots, \mathcal{P}_{\ell m}^{\mathrm{Im}}, \qquad (12)$$

where $\mathcal{P}_{\ell m}^{\mathrm{Re}}$ and $\mathcal{P}_{\ell m}^{\mathrm{Im}}$ are, respectively, the real and imaginary parts of $\mathcal{P}_{\ell m}$. For the sake of clarity, we introduce a $(\ell_{\max} + 1)^2$-vector to denote these numbers,

$$\mathbf{w} = (\mathcal{P}_{00}, \mathcal{P}_{10}, \ldots, \mathcal{P}_{\ell 0},$$
$$\mathcal{P}_{11}^{\mathrm{Re}}, \mathcal{P}_{21}^{\mathrm{Re}}, \ldots, \mathcal{P}_{\ell m}^{\mathrm{Re}},$$
$$\mathcal{P}_{11}^{\mathrm{Im}}, \mathcal{P}_{21}^{\mathrm{Im}}, \ldots, \mathcal{P}_{\ell m}^{\mathrm{Im}})^{\mathrm{T}}. \qquad (13)$$

III. BASICS OF BAYESIAN SEARCH FOR GRAVITATIONAL-WAVE BACKGROUND

Searching for an anisotropic GWB amounts to determining the spherical-harmonic components of the intensity map from the data. In the presence of overwhelming noise, the spherical-harmonic components, just as the estimation of the parameters of other GW signals, should be determined by Bayesian inference. This section is devoted to reviewing the basics of a Bayesian search for an anisotropic GWB. Before we explain the Bayesian strategy we propose, we will first state the assumptions and simplifications we make for the calculations, and then we will justify them. Then, we will define the likelihood, prior and posterior for

the Bayesian search. In terms of the spherical-harmonic components defined in the last section, we will explicitly write down the likelihood function as a Gaussian of the real and imaginary parts of the spherical-harmonic components of the GWB. Finally, we also define the Bayes factor, which competes the hypothesis that an anisotropic GWB is detected against that hypothesis that the data consists of only noises.

A. Approximations and simplifications

To construct various probability distribution functions that will be used in our Bayesian GWB analysis, we make the following assumptions and approximations:

(A.1) The GWB signal is weak, in the sense that the autocorrelated power and the cross-correlated power of the GWB signal are much smaller than that of the detector noise. This assumption is justified because the current observational constraints on the strength of GWB indicates that, if a GWB is present at all, it must be weak [82,88].

(A.2) The instrumental noise is stationary and Gaussian. In practice, this assumption is not always realistic as individual GW signals or non-Gaussian noise transients, known as glitches, will occasionally be present. Nonetheless, data segments containing non-Gaussian transients will be removed by applying data cuts [95]. Hence, in our analysis, we can assume that the detector response is Gaussian.

(A.3) The noise across different detectors is not correlated. This assumption is realistic because the existing ground-based detectors are spatially well separated.

(A.4) We assume that different segments of the data are independent (uncorrelated in time and frequency) to simplify the statistical calculations. This assumption implies that the likelihood function can be written as a product of the likelihood over different time and frequency segments. In practice, the segments are correlated for two reasons. First, the serial dependence in the entire segment of the time-domain data introduces autocorrelations over time and frequency. Second, when we transform time-domain data segments into the frequency domain, the data segments are windowed. To make full use of the windowed data, the time-domain data segments need to overlap, which introduces correlations between them. In practice, to address these correlations, Eq. (53) (one of our key results) should first be applied to each individual segment (with appropriate windowing factors multiplied), and then optimally combined [96–98]. Reference [99], however, has shown that, after taking all these correlations into account, the optimally combined results from all data segments agree well with predictions obtained from a likelihood that assumes the data segments are

independent of each other. Such consistency justifies the simplifications used in this paper.

B. Likelihood and posterior in Bayesian GWB analysis

When a GWB is present, it induces responses on GW detectors that force the latter to measure strain data consisting of two parts:

$$\tilde{s}_I(f,t) = \tilde{n}_I(f,t) + \tilde{h}_I(f,t), \tag{14}$$

where I labels the detector and $\tilde{s}_I(f,t)$ is the finite- or short-time Fourier transform of the time-domain data in detector I, $s_I(t)$, within a time interval $[t-\tau/2, t+\tau/2]$,

$$\tilde{s}_I(f,t) = \int_{t-\tau/2}^{t+\tau/2} dt' w(t') s_I(t') e^{-2i\pi ft'}, \tag{15}$$

with $w(t)$ a windowing function. Similarly, $\tilde{n}_I(f,t)$ and $\tilde{h}_I(f,t)$ are the short-time Fourier transform of the instrumental noise, $n_I(t)$, and of the GWB-induced response, $h_I(t)$, of detector I, respectively. The time-domain GWB-induced response on detector I is related to the metric perturbations of the GWB [Eq. (1)] via

$$h_I(t) = D_I^{ij} h_{ij}(t, \mathbf{x}_I), \tag{16}$$

where D_I^{ij} is a tensor (known as the detector response tensor) that encapsulates the geometry and orientations geometry of detector I, and $\mathbf{x}_I$ is the position vector of detector I.

The expectation value of the GWB-induced response satisfies

$$\langle \tilde{h}_I(f,t) \rangle = 0, \tag{17}$$

which descends directly from Eq. (2). As $\tilde{h}_I(f,t)$ is random and has zero mean, the GWB-induced response just looks like noise within individual detectors. However, the responses induced on two GW detectors, say I and J, should be proportional to each other (in the time domain), which means that they should be correlated across among detectors [81],

$$\langle \tilde{h}_I(f,t)\tilde{h}_J^*(f,t) \rangle = \frac{\tau}{2} H_\alpha(f) \sum_{\ell m} \gamma_{\ell m}^{(IJ)}(f,t)\mathcal{P}_{\ell m}, \tag{18}$$

where τ is the time length of the data segment analyzed and $\gamma_{\ell m}^{(IJ)}(f,t)$ is the spherical-harmonic components of the overlap reduction function (ORF) of detectors I and J, defined by [90]

$$\gamma^{(IJ)}(f,t,\hat{\Omega}) = \frac{1}{2} \sum_A R_A^{(I)}(f,t,\hat{\Omega})[R_A^{(J)}(f,t,\hat{\Omega})]^*,$$

$$\gamma_{\ell m}^{(IJ)}(f,t) = \int d^2\hat{\Omega}\gamma^{(IJ)}(f,t,\hat{\Omega})Y_{\ell m}(\hat{\Omega}), \tag{19}$$

where $A = +, \times$ stands for the GW polarization, and $R_A^{(I)}(f,t,\hat{\Omega})$ is the polarization-basis response function of detector I. The latter depends on time because of Earth's rotation. As the definition suggests, $\gamma_{\ell m}^{(IJ)}(f,t)$ encapsulates information about the detectors' geometry, location, orientation, and antenna pattern, and they should not be confused with the spherical-harmonic components of the spectrum of the GWB, $\mathcal{P}_{\ell m}$. Instrumental noise, on the other hand, has very different properties: if I and J are well separated, their instrumental noise should be uncorrelated,

$$\langle \tilde{n}_I(f,t)\tilde{n}_J^*(f,t) \rangle = 0. \tag{20}$$

By the same token

$$\langle \tilde{n}_I(f,t)\tilde{h}_J^*(f,t) \rangle = \langle \tilde{n}_J(f,t)\tilde{h}_I^*(f,t) \rangle$$
$$= \langle \tilde{n}_I(f,t)\tilde{h}_I^*(f,t) \rangle = \langle \tilde{n}_J(f,t)\tilde{h}_J^*(f,t) \rangle = 0. \tag{21}$$

These correlation properties suggest that a GWB can be searched for by cross-correlating the strain data measured by different detectors. To this end, we define the cross-spectral density, $C(f,t)$, between two detectors, I and J, via

$$C(f,t) \equiv \frac{2}{\tau} \tilde{s}_I(f,t)\tilde{s}_J^*(f,t). \tag{22}$$

If a GWB is present, then the expectation value of $C(f,t)$ is [81]

$$\langle C(f,t) \rangle = H_\alpha(f) \sum_{\ell m} \gamma_{\ell m}(f,t)\mathcal{P}_{\ell m}. \tag{23}$$

We can also derive the (approximate) variance of $C(f,t)$ for a weak GWB signal by considering the covariance matrix,

$$\mathrm{Cov}(f,t;f',t') = \langle C(f,t)C^*(f',t') \rangle - \langle C(f,t) \rangle \langle C^*(f',t') \rangle,$$
$$\approx \langle C(f,t)C^*(f',t') \rangle, \tag{24}$$

where we have dropped the term $\langle C(f,t) \rangle \langle C^*(f',t') \rangle$ because it corresponds to a second-order contribution in the weak-signal approximation (cf. A.1). Then, using Eqs. (14) and (20)–(22), we have

$$\mathrm{Cov}(f,t;f',t') = \delta_{ff'}\delta_{tt'}N_I(f,t)N_J(f,t), \tag{25}$$

where $N_{I,J}(f,t)$ is the one-sided PSD of the output of detectors I and J. We remind the reader that the derivation

is valid only if we assume A.1. Reference [100] has verified that Eq. (25) gives an accurate estimate of the variance of the cross-correlation for the search of isotropic GWBs. Since anisotropic GWBs is an extension of an isotropic GWB search, we expect that Eq. (25) remains accurate in this case as well.

The Bayesian search of an anisotropic GWB amounts to determining the posterior of $\mathbf{w}$, given the cross-spectral density, of all data segments $\{C\}$. According to Bayes's theorem, the posterior is related to the likelihood by

$$p(\mathbf{w}|\{C\}, H) = \frac{p(\mathbf{w}|H)p(\{C\}|\mathbf{w}, H)}{p(\{C\}|H)}. \qquad (26)$$

Here, $p(\mathbf{w}|\{C\}, H)$ is the posterior of $\mathbf{w}$, given the cross-spectral density and the hypothesis H [e.g. that the measured data contain a GWB signal, which will be more precisely defined in Eq. (30)]. The quantity $p(\{C\}|H)$ is the Bayesian evidence, which is a normalization constant of the posterior. The quantity $p(\mathbf{w}|H)$ is the prior of $\mathbf{w}$, prescribed according to our hypothesis. The quantity $p(\{C\}|\mathbf{w}, H)$ is the likelihood that we will measure $\{C\}$, given that there is a GWB with spherical-harmonic components $\mathbf{w}$. Using the weak-signal approximation, and the expectation value and the variance derived above, the likelihood $p(\{C\}|\mathbf{w}, H)$ can be modeled by [81,83,90]

$$p(\{C\}|\mathbf{w}, H) = \mathcal{N} \exp\left\{-\frac{1}{2}\sum_{f,t}\frac{|C(f,t) - H_\alpha(f)\gamma_\mu\mathcal{P}_\mu|^2}{N_I(f,t)N_J(f,t)}\right\}, \qquad (27)$$

where $\sum_{f,t}$ stands for summation over frequency bins and the center times of the short-timed Fourier transform, $\mathcal{N}$ is a proportionality constant that does not depend on $\mathcal{P}_{\ell m}$, and $\gamma_\mu\mathcal{P}_\mu$ is shorthand notation for

$$\gamma_\mu\mathcal{P}_\mu = \sum_{\ell=0}^{\ell_{\max}^{(\inf)}}\sum_{m=-\ell}^{+\ell}\gamma_{\ell m}(f,t)\mathcal{P}_{\ell m}, \qquad (28)$$

where $\mu = (\ell, m)$ labels the mode, and $\ell_{\max}^{(\inf)}$ is the maximum ℓ that we include in the inference analysis.

When searching for an anisotropic GWB, Eq. (26) represents a high-dimensional posterior probability distribution function, which is difficult to visualize. Thus, it is very convenient to present the marginalized posterior of a particular spherical-harmonic component. To this end, one can marginalize the posterior [Eq. (26)] over all components of $\mathbf{w}$ that one is not interested in (at the moment) to obtain the marginalized posterior of, say, w_i,

$$p(w_i|\{C\}, H) = \prod_{j\neq i}\int dw_j\, p(\mathbf{w}|\{C\}, H). \qquad (29)$$

The lower and upper limit of the integral involved in the marginalization depends on $p(\mathbf{w}|H)$, which will be prescribed in Sec. IV.

C. The Bayes factor

Other than constructing the marginalized posterior, Bayesian theory also provides a framework to compute the so-called Bayes factor. The latter is a measure that allows one to compare two hypotheses in light of the data within Bayesian inference. In the context of GWBs, the Bayes factor can be used to quantify whether an anisotropic GWB has been detected or not by comparing the following two hypotheses:

$H_{\ell_{\max}}$: the data $\{C\}$ contain a GWB signal whose

$\mathcal{P}_{\ell_{\max}m} \neq 0$ for at least one $m \in$

$[-\ell_{\max}, -\ell_{\max}+1, \ldots, \ell_{\max}-1, \ell_{\max}]$, and

H_{null}: the data $\{C\}$ contain only noise. $\qquad (30)$

In Bayesian inference, we can compare these two hypotheses by computing their odds ratio, namely, the ratio of their respective evidences given the data:

$$\mathcal{O}(\ell_{\max}) = \frac{p(H_{\ell_{\max}}|\{C\})}{p(H_{\text{null}}|\{C\})} = \frac{p(H_{\ell_{\max}})}{p(H_{\text{null}})}\frac{p(\{C\}|H_{\ell_{\max}})}{p(\{C\}|H_{\text{null}})}. \quad (31)$$

The term $p(H_{\ell_{\max}})/p(H_{\text{null}})$ is known as the prior odds, and it represents our prior belief of one hypothesis over the other. The second term in the above equation is known as the Bayes factor,

$$\mathcal{B}(\ell_{\max}) = \frac{p(\{C\}|H_{\ell_{\max}})}{p(\{C\}|H_{\text{null}})}, \qquad (32)$$

which implies that the odds ratio is nothing but the product of the prior odds with the Bayes factor.

One can think of the Bayes factor as the odds ratio between two hypotheses under the assumption of equal prior belief between them. As we have no information about whether we have detected a GWB before we analyze the data, we naturally assume the two hypotheses are equally likely. Thus,

$$p(H_{\ell_{\max}}) = p(H_{\text{null}}) \Rightarrow \mathcal{O}(\ell_{\max}) = \mathcal{B}(\ell_{\max}). \qquad (33)$$

If $\mathcal{B}(\ell_{\max}) > 1$, then hypothesis $H_{\ell_{\max}}$ is favored over hypothesis H_{null}, which implies it is more likely that we have detected a GWB than not; the opposite is true, of course, if $\mathcal{B}(\ell_{\max}) \leq 1$. For convincing evidence that we have indeed detected a GWB, one typically requires that $\mathcal{B}(\ell_{\max}) \gg 1$, where precisely how much larger than unity this requirement must depend on the statistician's definition of "convincing" [90].

IV. ANALYTIC MARGINALIZATION OF THE POSTERIOR AND BAYES FACTOR

As pointed out in the last section, the posterior of the spherical-harmonic components is a probability distribution function of high dimension. In principle, one can numerically sample the posterior using nested sampling or Markov chain Monte Carlo techniques. But given the high dimensionality of the distribution, both sampling approaches will take an extremely long time to complete in the GWB case. In this section, we will show that, if a wide-enough uniform prior is prescribed, the marginalized posterior and Bayes factor for the search for an anisotropic GWB can be *analytically* evaluated as a high-dimensional Gaussian integral, with the former also a Gaussian function.

A. Marginalized posterior

Let us begin by explicitly writing down the exponent of the likelihood as a quadratic form of $\mathbf{w}$. To start, we rewrite the likelihood as

$$p(\{C\}|\mathbf{w}, H_{\ell_{\max}^{(\mathrm{inf})}}) \propto \exp\left\{-\frac{1}{2}\sum_{f,t}\frac{R(f,t)R^*(f,t)}{N_I(f,t)N_J(f,t)}\right\}, \quad (34)$$

where $R(f,t)$ stands for the residual

$$R(f,t) \equiv C(f,t) - H(f)\sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\sum_{m=-\ell}^{\ell}\gamma_{\ell m}(f,t)\mathcal{P}_{\ell m}. \quad (35)$$

Explicitly writing out the summation over ℓ and m, we have

$$\sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\sum_{m=-\ell}^{+\ell}\gamma_{\ell m}(f,t)\mathcal{P}_{\ell m} = \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\gamma_{\ell 0}(f)\mathcal{P}_{\ell 0} + \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\sum_{m=1}^{+\ell}[\gamma_{\ell m}(f,t)\mathcal{P}_{\ell m} + \gamma_{\ell,-m}(f,t)\mathcal{P}_{\ell,-m}],$$

$$= \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\gamma_{\ell 0}(f)\mathcal{P}_{\ell 0} + \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\sum_{m=1}^{+\ell}[\gamma_{\ell m}(f,t)\mathcal{P}_{\ell m} + (-1)^{\ell}\gamma_{\ell m}^*(f,t)\mathcal{P}_{\ell m}^*], \quad (36)$$

where, in the last line, we have used Eqs. (B1) and (B4) of [81], namely

$$\gamma_{\ell m}^*(f,t) = (-1)^{\ell+m}\gamma_{\ell,-m}(f,t),$$
$$\mathcal{P}_{\ell m}^* = (-1)^m \mathcal{P}_{\ell,-m}. \quad (37)$$

We further decompose $\gamma_{\ell m}\mathcal{P}_{\ell m}$ into its real and imaginary parts,

$$\Re[\gamma_{\ell m}\mathcal{P}_{\ell m}] = \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\gamma_{\ell 0}^{\mathrm{Re}}(f)\mathcal{P}_{\ell 0} + \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\sum_{m=1}^{+\ell}[1+(-1)^{\ell}][\gamma_{\ell m}^{\mathrm{Re}}(f,t)\mathcal{P}_{\ell m}^{\mathrm{Re}} - \gamma_{\ell m}^{\mathrm{Im}}(f,t)\mathcal{P}_{\ell m}^{\mathrm{Im}}],$$

$$\Im[\gamma_{\ell m}\mathcal{P}_{\ell m}] = \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\gamma_{\ell 0}^{\mathrm{Im}}(f)\mathcal{P}_{\ell 0} + \sum_{\ell=0}^{\ell_{\max}^{(\mathrm{inf})}}\sum_{m=1}^{+\ell}[1-(-1)^{\ell}][\gamma_{\ell m}^{\mathrm{Im}}(f,t)\mathcal{P}_{\ell m}^{\mathrm{Re}} + \gamma_{\ell m}^{\mathrm{Re}}(f,t)\mathcal{P}_{\ell m}^{\mathrm{Im}}]. \quad (38)$$

These expressions can be more compactly expressed if we define two $(\ell_{\max}^{(\mathrm{inf})}+1)^2$ vectors, $\mathbf{u}(f,t)$ and $\mathbf{v}(f,t)$, such that

$$\Re[R(f,t)] = C^{\mathrm{Re}}(f,t) - \mathbf{u}^{\mathrm{T}}(f,t)\cdot\mathbf{w},$$
$$\Im[R(f,t)] = C^{\mathrm{Im}}(f,t) - \mathbf{v}^{\mathrm{T}}(f,t)\cdot\mathbf{w}, \quad (39)$$

where

$$\mathbf{u}(f,t) \equiv H_\alpha(f)(\gamma_{00}^{\mathrm{Re}}, \gamma_{10}^{\mathrm{Re}}, ..., \gamma_{\ell 0}^{\mathrm{Re}}, [1+(-1)^1]\gamma_{11}^{\mathrm{Re}},$$
$$[1+(-1)^2]\gamma_{21}^{\mathrm{Re}}, ..., [1+(-1)^{\ell}]\gamma_{\ell m}^{\mathrm{Re}}, -[1+(-1)^1]\gamma_{11}^{\mathrm{Im}}, -[1+(-1)^2]\gamma_{21}^{\mathrm{Im}}, ..., -[1+(-1)^{\ell}]\gamma_{\ell m}^{\mathrm{Im}})^{\mathrm{T}},$$
$$\mathbf{v}(f,t) \equiv H_\alpha(f)(\gamma_{00}^{\mathrm{Im}}, \gamma_{10}^{\mathrm{Im}}, ..., \gamma_{\ell 0}^{\mathrm{Im}}, [1-(-1)^1]\gamma_{11}^{\mathrm{Im}},$$
$$[1-(-1)^2]\gamma_{21}^{\mathrm{Im}}, ..., [1-(-1)^{\ell}]\gamma_{\ell m}^{\mathrm{Im}}, [1-(-1)^1]\gamma_{11}^{\mathrm{Re}}, [1-(-1)^2]\gamma_{21}^{\mathrm{Re}}, ..., [1-(-1)^{\ell}]\gamma_{\ell m}^{\mathrm{Re}})^{\mathrm{T}}. \quad (40)$$

Note that $\mathbf{u}(f,t)$ and $\mathbf{v}(f,t)$ depend only on the ORF of the detectors, but each element of $\mathbf{u}(f,t)$ and $\mathbf{v}(f,t)$ is a function of f and t because they inherit the frequency and time dependence from $H(f)$ and $\gamma_{\ell m}(f,t)$.

With $\mathbf{u}$ and $\mathbf{v}$ defined, the square of the modulus of $R(f,t)$ can be computed as a quadratic function of $\mathbf{w}$

$$R(f,t)R^*(f,t)$$
$$= C(f,t)C^*(f,t) - 2\mathbf{g}^{\mathrm{T}}(f,t)\cdot\mathbf{w} + \mathbf{w}^{\mathrm{T}}\cdot\mathbf{K}(f,t)\cdot\mathbf{w},$$
(41)

where $\mathbf{g}(f,t)$ is a $(\ell_{\max}^{(\mathrm{inf})}+1)^2$ vector,

$$\mathbf{g}(f,t) \equiv C^{\mathrm{Re}}(f,t)\mathbf{u}(f,t) + C^{\mathrm{Im}}(f,t)\mathbf{v}(f,t),$$
(42)

and $\mathbf{K}(f,t)$ is a symmetric-square matrix of order of $(\ell_{\max}^{(\mathrm{inf})}+1)^2$, whose elements are given by

$$K_{ij}(f,t) \equiv u_i u_j + v_i v_j.$$
(43)

Note that while $\mathbf{g}(f,t)$ depends on the data via the real and imaginary parts of $C(f,t)$, K_{ij} depends solely on the detectors' geometry via the dependence on the ORF. Similarly, we can also write the exponent of the likelihood as a quadratic function of $\mathbf{w}$,

$$\sum_{f,t}\frac{R(f,t)R^*(f,t)}{N_I(f,t)N_J(f,t)} = \sum_{f,t}\frac{|C(f,t)|^2}{N_I(f,t)N_J(f,t)}$$
$$- 2\mathbf{j}^{\mathrm{T}}\cdot\mathbf{w} + \mathbf{w}^{\mathrm{T}}\cdot\mathbf{Q}\cdot\mathbf{w},$$
(44)

where $\mathbf{j}$ is a $[\ell_{\max}^{(\mathrm{inf})}+1]^2$-vector and $\mathbf{Q}$ is another symmetric-square matrix of order of $[\ell_{\max}^{(\mathrm{inf})}+1]^2$. Explicitly, their elements are

$$j_i \equiv \sum_{f,t}\frac{g_i(f,t)}{N_I(f,t)N_J(f,t)},$$

$$Q_{ij} \equiv \sum_{f,t}\frac{K_{ij}(f,t)}{N_I(f,t)N_J(f,t)}.$$
(45)

Recall that $\mathbf{g}$ depends on the data, and so does $\mathbf{j}$. Even though $\mathbf{K}$ does not depend on data, $\mathbf{Q}$ does because it contains the PSD of the data. Unlike $\mathbf{g}$ and $\mathbf{K}$, $\mathbf{j}$, and $\mathbf{Q}$ are constant. In terms of $\mathbf{j}$, $\mathbf{w}$, and $\mathbf{Q}$, the likelihood and posterior are, respectively, given by

$$p(\{C\}|\mathbf{w}, H_{\ell_{\max}^{(\mathrm{inf})}}) = \bar{\mathcal{N}}\exp\left(\mathbf{j}^{\mathrm{T}}\cdot\mathbf{w} - \frac{1}{2}\mathbf{w}^{\mathrm{T}}\cdot\mathbf{Q}\cdot\mathbf{w}\right),$$

$$p(\mathbf{w}|\{C\}, H_{\ell_{\max}^{(\mathrm{inf})}}) = \bar{\mathcal{N}}\frac{p(\mathbf{w}|H_{\ell_{\max}^{(\mathrm{inf})}})}{p(\{C\}|H_{\ell_{\max}^{(\mathrm{inf})}})}$$
$$\times\exp\left(\mathbf{j}^{\mathrm{T}}\cdot\mathbf{w} - \frac{1}{2}\mathbf{w}^{\mathrm{T}}\cdot\mathbf{Q}\cdot\mathbf{w}\right),$$
(46)

where

$$\bar{\mathcal{N}} = \mathcal{N}\exp\left(-\frac{1}{2}\sum_{f,t}\frac{|C(f,t)|^2}{N_I(f,t)N_J(f,t)}\right).$$
(47)

At this point, let us summarize and remind the reader that $\mathbf{g}$, $\mathbf{j}$, and $\mathbf{Q}$ depend on the data, whereas $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{K}$ depend only on the geometry of the detectors via the ORF. Thus, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{K}$ can be precomputed and stored for given detectors to speed up the analysis.

We are now ready to marginalize the posterior. If we are particularly interested in knowing the posterior of w_i, then the argument of the exponential of the posterior can be written as

$$\mathbf{j}^{\mathrm{T}}\cdot\mathbf{w} - \frac{1}{2}\mathbf{w}^{\mathrm{T}}\cdot\mathbf{Q}\cdot\mathbf{w} = j_i w_i - \frac{1}{2}Q_{ii}w_i^2$$
$$+ \sum_{k\neq i}\left[j_k - \frac{1}{2}w_i(Q_{ki}+Q_{ik})\right]w_k$$
$$- \frac{1}{2}\sum_{k\neq i}\sum_{l\neq i}w_k Q_{kl}w_l,$$
(48)

where the index i in the first two terms, $j_i w_i$ and $Q_{ii}w_i^2$, does not imply summation. To facilitate subsequent calculations, we define the following $([\ell_{\max}^{(\mathrm{inf})}+1]^2-1)$-vectors and square matrices of order $[[\ell_{\max}^{(\mathrm{inf})}+1]^2-1]$:

$\tilde{\mathbf{w}}^{(i)}$ = the vector $\mathbf{w}$ with the ith element removed,

$\mathbf{b}^{(i)}$ = the vector $\mathbf{j}$ with the ith element removed,

$\mathbf{a}^{(i)}$ = a vector whose kth element is

$a_k = Q_{ik}$ (for $k \neq i$ given i)

$\mathbf{n}^{(i)} = \mathbf{b}^{(i)} - w_i\mathbf{a}^{(i)}$,

$\tilde{\mathbf{Q}}^{(i)}$ = the matrix $\mathbf{Q}$ with the ith row and the ith column removed,

$\mathbf{M}^{(i)}$ = the inverse of $\tilde{\mathbf{Q}}^{(i)}$.

Note that, except $\mathbf{w}$, all these vectors and matrices depend on the data. The vectors $\mathbf{b}$ and $\mathbf{n}$ depend on the data because $\mathbf{j}$ depends on the data [cf. Eq. (45)]. The vector $\mathbf{a}$ and the matrices $\tilde{\mathbf{Q}}^{(i)}$ and $\mathbf{M}^{(i)}$ all depend on the data because $\mathbf{Q}$ depends on the PSD of the data. Note also that,

since $\tilde{\mathbf{Q}}^{(i)}$ is symmetric, so is $\mathbf{M}^{(i)}$. The argument of the exponential of the posterior can then be more compactly written as

$$
\mathbf{j}^{\mathrm{T}} \cdot \mathbf{w} - \frac{1}{2} \mathbf{w}^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{w} = j_i w_i - \frac{1}{2} Q_{ii} w_i^2 + \mathbf{n}^{(i)\mathrm{T}} \cdot \tilde{\mathbf{w}}^{(i)} \\
- \frac{1}{2} \tilde{\mathbf{w}}^{(i)\mathrm{T}} \cdot \tilde{\mathbf{Q}}^{(i)} \cdot \tilde{\mathbf{w}}^{(i)}, \qquad (49)
$$

where we recall that $\mathbf{n}^{(i)}$ depends on both the index i and w_i, and the repeated subscript i does not imply summation.

The posterior can be analytically marginalized if we choose a prior for $\mathbf{w}$ with the following properties:

(1) The prior is factorized as a product of the prior of individual w_i,

$$
p(\mathbf{w}|H_{\ell_{\max}^{(\mathrm{inf})}}) = \prod_{i=1}^{[\ell_{\max}^{(\mathrm{inf})}+1]^2} p_i(w_i|H_{\ell_{\max}^{(\mathrm{inf})}}), \qquad (50)
$$

where $p_i(w_i|H_{\ell_{\max}^{(\mathrm{inf})}})$ is the prior of w_i. By choosing a factorized prior for $\mathbf{w}$, we are assuming that different w_i are independent of each other.

(2) Each $p_i(w_i|H_{\ell_{\max}^{(\mathrm{inf})}})$ is uniform for $w_i \in [-\Delta^{(i)}, \Delta^{(i)}]$, where $\Delta^{(i)} > 0$ is the width of the prior of w_i.

(3) When $w_i = \pm\Delta^{(i)}$, $\mathbf{j}^{\mathrm{T}} \cdot \mathbf{w} - \frac{1}{2}\mathbf{w}^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{w}$ is very negative, regardless of the value of the other $w_{j\neq i}$. This condition can always be met if we choose a large enough $\Delta^{(i)}$ such that

$$
p(\{C\}|w_i = \pm\Delta^{(i)}, H_{\ell_{\max}^{(\mathrm{inf})}}) \approx 0. \qquad (51)
$$

This prior corresponds to a square centered at the origin in the complex $\mathcal{P}_{\ell m}$ plane for $(\ell, m) \neq 0$. One may think that a more natural prior would be one that is uniform for, say, $|\mathcal{P}_{\ell m}| \leq \Delta$ with some $\Delta > 0$, which corresponds to a circle centered at the origin in the complex plane. However, if Δ is large enough, both the square and circle priors will lead to similar parameter estimation results. This is because, in the region between the square and the circle priors, the argument of the exponential in the posterior is very negative, and thus, the contribution to the posterior can be well approximated by zero. This condition is not contradictory to the weak signal approximation, because it can be met by a smaller $\Delta^{(i)}$, corresponding to a weaker signal if we have more data.

With these properties in place, the marginalized posterior of w_i can be evaluated as

$$
\begin{aligned}
p(w_i|\{C\}, H_{\ell_{\max}^{(\mathrm{inf})}}) &= \int d\tilde{\mathbf{w}}^{(i)} p(\mathbf{w}|\{C\}, H_{\ell_{\max}^{(\mathrm{inf})}}), \\
&= \frac{1}{p(\{C\}|H_{\ell_{\max}^{(\mathrm{inf})}})} \prod_{j\neq i} \int_{-\Delta^{(i)}}^{\Delta^{(i)}} dw_j \, p(\mathbf{w}|H) p(\{C\}|\mathbf{w}, H_{\ell_{\max}^{(\mathrm{inf})}}), \\
&= \frac{1}{p(\{C\}|H_{\ell_{\max}^{(\mathrm{inf})}})} \prod_{j\neq i} \int_{-\Delta^{(i)}}^{\Delta^{(i)}} \frac{dw_j}{2\Delta^{(i)}} p(\{C\}|\mathbf{w}, H_{\ell_{\max}^{(\mathrm{inf})}}) \\
&\propto \prod_{j\neq i} \int_{-\Delta^{(i)}}^{\Delta^{(i)}} dw_j \exp\left(j_i w_i - \frac{1}{2} Q_{ii} w_i^2 + \mathbf{n}^{(i)\mathrm{T}} \cdot \tilde{\mathbf{w}}^{(i)} - \frac{1}{2} \tilde{\mathbf{w}}^{(i)\mathrm{T}} \cdot \tilde{\mathbf{Q}}^{(i)} \cdot \tilde{\mathbf{w}}^{(i)} \right), \\
&\approx \prod_{j\neq i} \int_{-\infty}^{+\infty} dw_j \exp\left(j_i w_i - \frac{1}{2} Q_{ii} w_i^2 + \mathbf{n}^{(i)\mathrm{T}} \cdot \tilde{\mathbf{w}}^{(i)} - \frac{1}{2} \tilde{\mathbf{w}}^{(i)\mathrm{T}} \cdot \tilde{\mathbf{Q}}^{(i)} \cdot \tilde{\mathbf{w}}^{(i)} \right) \\
&\propto \exp\left(j_i w_i - \frac{1}{2} Q_{ii} w_i^2 + \frac{1}{2} \mathbf{n}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{n} \right), \\
&= \exp\left(j_i w_i - \frac{1}{2} Q_{ii} w_i^2 + \frac{1}{2} (\mathbf{b}^{(i)} - w_i \mathbf{a}^{(i)})^{\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot (\mathbf{b}^{(i)} - w_i \mathbf{a}^{(i)}) \right) \\
&\propto \exp\left[(j_i - \mathbf{a}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{b}^{(i)}) w_i - \frac{1}{2} (Q_{ii} - \mathbf{a}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{a}^{(i)}) w_i^2 \right], \qquad (52)
\end{aligned}
$$

where in going from the fourth to the fifth line we have made use of the third property of the uniform prior of $\mathbf{w}$, and from the sixth to the seventh line we have used that $\mathbf{M}^{(i)}$ is symmetric, so that $\mathbf{b}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{a}^{(i)} = \mathbf{a}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{b}^{(i)}$. We again remind the reader that the repeated index

i does not imply summation. We see that the marginalized posterior of w_i is a Gaussian function of w_i. The mean, μ_i, and the standard deviation, σ_i, of w_i can be read from the marginalized posterior of w_i readily, namely

$$\mu_i = \frac{j_i - \mathbf{a}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{b}^{(i)}}{Q_{ii} - \mathbf{a}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{a}^{(i)}},$$

$$\sigma_i = (Q_{ii} - \mathbf{a}^{(i)\mathrm{T}} \cdot \mathbf{M}^{(i)} \cdot \mathbf{a}^{(i)})^{-\frac{1}{2}}. \tag{53}$$

Note that, throughout the marginalization procedure, we do not ignore the correlations among the components of w_i; these correlations are encoded in the off-diagonal elements of $\mathbf{Q}$. If one neglects the correlations among the w_is, one sets the off-diagonal elements of $\mathbf{Q}$ to zero, resulting in a diagonal $\tilde{\mathbf{Q}}^{(i)}$ and $\tilde{\mathbf{M}}^{(i)}$.

Equation (53) can be more compactly written as

$$\mu_i = [\mathbf{Q}^{-1} \cdot \mathbf{j}]_i,$$

$$\sigma_i = [\mathbf{Q}^{-1}]_{ii}^{\frac{1}{2}}, \tag{54}$$

using the inverse formulas of a block matrix. To see this, we first consider the case of $i = 1$ and write $\mathbf{Q}$ as a block matrix and $\mathbf{j}$ as

$$\mathbf{Q} = \begin{pmatrix} Q_{11} & \mathbf{a}^{(1)\,\mathrm{T}} \\ \mathbf{a}^{(1)} & \tilde{\mathbf{Q}}^{(1)} \end{pmatrix}, \qquad \mathbf{j} = \begin{pmatrix} j_1 \\ \mathbf{b}^{(1)} \end{pmatrix}. \tag{55}$$

Then, we compute the inverse of $\mathbf{Q}$ using the block-inverse formula. For a block matrix $\mathbf{P}$ [101],

$$\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix}. \tag{56}$$

Taking $\mathbf{A} = Q_{11}$, $\mathbf{B} = \mathbf{a}^{(1)\,\mathrm{T}}$, $\mathbf{C} = \mathbf{a}^{(1)}$, and $\mathbf{D} = \tilde{\mathbf{Q}}^{(1)}$, we find

$$\mathbf{Q}^{-1} = \begin{pmatrix} (Q_{11} - \mathbf{a}^{(1)\,\mathrm{T}} \cdot \tilde{\mathbf{M}}^{(1)} \cdot \mathbf{a}^{(1)})^{-1} & -(Q_{11} - \mathbf{a}^{(1)\,\mathrm{T}} \cdot \tilde{\mathbf{M}}^{(1)} \cdot \mathbf{a}^{(1)})^{-1}\mathbf{a}^{(1)\,\mathrm{T}}\tilde{\mathbf{M}}^{(1)} \\ -\tilde{\mathbf{M}}^{(1)} \cdot \mathbf{a}^{(1)}(Q_{11} - \mathbf{a}^{(1)\,\mathrm{T}} \cdot \tilde{\mathbf{M}}^{(1)} \cdot \mathbf{a}^{(1)})^{-1} & \tilde{\mathbf{M}}^{(1)} + \tilde{\mathbf{M}}^{(1)} \cdot \mathbf{a}^{(1)}(Q_{11} - \mathbf{a}^{(1)\,\mathrm{T}} \cdot \tilde{\mathbf{M}}^{(1)} \cdot \mathbf{a}^{(1)})^{-1}\mathbf{a}^{(1)\,\mathrm{T}} \cdot \tilde{\mathbf{M}}^{(1)} \end{pmatrix}. \tag{57}$$

Reading the first row of $\mathbf{Q}^{-1} \cdot \mathbf{j}$ and the (1,1) element of $\mathbf{Q}^{-1}$, we find that

$$[\mathbf{Q}^{-1} \cdot \mathbf{j}]_1 = \frac{j_1 - \mathbf{a}^{(1)\mathrm{T}} \cdot \mathbf{M}^{(1)} \cdot \mathbf{b}^{(1)}}{Q_{11} - \mathbf{a}^{(1)\mathrm{T}} \cdot \mathbf{M}^{(1)} \cdot \mathbf{a}^{(1)}} = \mu_1,$$

$$[\mathbf{Q}^{-1}]_{11} = (Q_{11} - \mathbf{a}^{(1)\mathrm{T}} \cdot \mathbf{M}^{(1)} \cdot \mathbf{a}^{(1)})^{-1} = \sigma_1^2. \tag{58}$$

The above arguments can be generalized to other $i \neq 1$. A convenient way to generalize the argument is rearranging $\mathbf{Q}$ and $\mathbf{j}$ into

$$\mathbf{Q} \to \begin{pmatrix} Q_{ii} & \mathbf{a}^{(i)\,\mathrm{T}} \\ \mathbf{a}^{(i)} & \tilde{\mathbf{Q}}^{(i)} \end{pmatrix}, \qquad \mathbf{j} \to \begin{pmatrix} j_i \\ \mathbf{b}^{(i)} \end{pmatrix}. \tag{59}$$

Computing $[\mathbf{Q}^{-1} \cdot \mathbf{j}]_1$ and $\mathbf{Q}^{-1}$ through the above procedure can then prove the case for $i \neq 1$. The marginalization procedure described above, and in particular Eq. (54) [or Eq. (53)], are some of the key results of this paper.

We shall conclude this subsection by discussing the relation between our analysis and the Fisher information matrix analysis. First, the matrix $\mathbf{Q}$ is actually a Fisher information matrix, which can be seen by realizing that

$$Q_{ij} = -\frac{\partial^2}{\partial w_i \partial w_j} \log \, p(\mathbf{w}|\{C\}, H_{\ell_{\mathrm{max}}^{(\mathrm{inf})}}). \tag{60}$$

Second, the maximum-likelihood estimation obtained by the Fisher information matrix analysis, which amounts to solving the equations

$$\frac{\partial}{\partial w_i} \log p(\mathbf{w}|\{C\}, H_{\ell_{\mathrm{max}}^{(\mathrm{inf})}})|_{\mathbf{w}=\mathbf{w}_{\mathrm{ML}}} = 0 \Rightarrow \mathbf{Q} \cdot \mathbf{w}_{\mathrm{ML}} = \mathbf{j}, \tag{61}$$

is actually identical to the μ_i given by Eq. (54). Moreover, following from the usual Fisher information matrix analysis, the measurement uncertainty of w_i is given by the square root of the (i, i) element of the inverse of the Fisher information matrix,

$$\Delta w_i = [\mathbf{Q}^{-1}]_{ii}^{\frac{1}{2}}, \tag{62}$$

which is just σ_i, as given in Eq. (54). In other words, Eq. (54) recovers exactly the maximum-likelihood estimate and the measurement uncertainty of $\mathbf{w}$ obtained using a Fisher information matrix analysis. This is reasonable, because $p(\mathbf{w}|H_{\ell_{\mathrm{max}}^{(\mathrm{inf})}})$ is a constant, which implies that the posterior is proportional to likelihood. Hence, the maximum-posterior $\mathbf{w}$ and maximum-likelihood $\mathbf{w}$ are the same, and so is the measurement uncertainty. Another consequence of this connection is that μ_i, being the w_i that maximizes the marginalized posterior, is also a component of the maximum-posterior $\mathbf{w}$, the latter of which is defined as

$$\mathbf{w}_{\mathrm{MP}} = \arg \max_{\mathbf{w}} p(\mathbf{w}|\{C\}, H_{\ell_{\mathrm{max}}^{(\mathrm{inf})}}). \tag{63}$$

Recovering the results obtained through the Fisher information matrix analysis proves the correctness of our marginalization.

B. Bayes factor

Given a large enough $\Delta^{(i)}$, the Bayes factor between the two hypotheses defined in Eq. (30) can also be analytically evaluated in a similar manner. To calculate the Bayes factor, we need to evaluate $p(\{C\}|H_{\ell_{\max}^{(\text{inf})}})$ and $p(\{C\}|H_{\text{null}})$. We first evaluate $p(\{C\}|H_{\ell_{\max}^{(\text{inf})}})$ using Bayes's theorem,

$$
\begin{aligned}
p(\{C\}|H_{\ell_{\max}^{(\text{inf})}}) &= \int d\mathbf{w} \, p(\{C\}|\mathbf{w}, H_{\ell_{\max}^{(\text{inf})}}) p(\mathbf{w}|H_{\ell_{\max}^{(\text{inf})}}), \\
&= \bar{\mathcal{N}} \prod_{i=1}^{(\ell_{\max}^{(\text{inf})}+1)^2} \int_{-\Delta^{(i)}}^{\Delta^{(i)}} \frac{dw_i}{2\Delta^{(i)}} \exp\left(\mathbf{j}^{\mathrm{T}} \cdot \mathbf{w} - \frac{1}{2} \mathbf{w}^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{w} \right), \\
&\approx \bar{\mathcal{N}} \prod_{i=1}^{(\ell_{\max}^{(\text{inf})}+1)^2} \int_{-\infty}^{+\infty} \frac{dw_i}{2\Delta^{(i)}} \exp\left(\mathbf{j}^{\mathrm{T}} \cdot \mathbf{w} - \frac{1}{2} \mathbf{w}^{\mathrm{T}} \cdot \mathbf{Q} \cdot \mathbf{w} \right), \\
&= \frac{\bar{\mathcal{N}}}{|\mathbf{Q}|^{1/2} \Delta(\ell_{\max}^{(\text{inf})})} \left(\frac{\pi}{2} \right)^{\frac{[\ell_{\max}^{(\text{inf})}+1]^2}{2}} \exp\left(\frac{1}{2} \mathbf{j}^{\mathrm{T}} \cdot \mathbf{Q}^{-1} \cdot \mathbf{j} \right),
\end{aligned}
\tag{64}
$$

where from the third to the fourth line we have again made use of the third property of the uniform prior of $\mathbf{w}$, $|\mathbf{Q}|$ is the determinant of $\mathbf{Q}$, and

$$
\Delta(\ell_{\max}^{(\text{inf})}) = \prod_{i=1}^{[\ell_{\max}^{(\text{inf})}+1]^2} \Delta^{(i)}. \tag{65}
$$

When the hypothesis is H_{null}, the evidence simplifies significantly, as we show below:

$$
\begin{aligned}
p(\{C\}|H_{\text{null}}) &= \int d\mathbf{w} \, p(\{C\}|\mathbf{w}, H_{\text{null}}) p(\mathbf{w}|H_{\text{null}}), \\
&= \int d\mathbf{w} \, p(\{C\}|\mathbf{w}=\mathbf{0}, H_{\ell_{\max}^{(\text{inf})}}) p(\mathbf{w}|H_{\ell_{\max}^{(\text{inf})}}), \\
&= p(\{C\}|\mathbf{w}=\mathbf{0}, H_{\ell_{\max}^{(\text{inf})}}), \\
&= \bar{\mathcal{N}}.
\end{aligned}
\tag{66}
$$

Thus, the Bayes factor can be analytically evaluated as

$$
\begin{aligned}
\mathcal{B}(\ell_{\max}^{(\text{inf})}|\{\Delta^{(i)}\}) &= \frac{1}{|\mathbf{Q}|^{1/2} \Delta(\ell_{\max}^{(\text{inf})})} \left(\frac{\pi}{2} \right)^{\frac{[\ell_{\max}^{(\text{inf})}+1]^2}{2}} \\
&\times \exp\left(\frac{1}{2} \mathbf{j}^{\mathrm{T}} \cdot \mathbf{Q}^{-1} \cdot \mathbf{j} \right).
\end{aligned}
\tag{67}
$$

At this junction, a word of caution is necessary. Equation (67) is valid only if a large enough $\Delta^{(i)}$ is chosen, because otherwise one cannot extend the limits of integration in the fifth line of Eq. (52) and in the third equality of Eq. (64). Apart from this criterion, the width of the prior of $\mathbf{w}$ is arbitrary, which means that the Bayes factor is also, in this sense, arbitrary. This is because the Bayes factor depends on the prior volume of the parameters that characterize the hypothesis that is being compared. Thus, when computing the Bayes factor using Eq. (67), one should also be careful of and report the chosen $\Delta^{(i)}$. This is also the reason why the Bayes factor in Eq. (67) is written as $\mathcal{B}(\ell_{\max}^{(\text{inf})}|\{\Delta^{(i)}\})$ to emphasize its dependence on both $\ell_{\max}^{(\text{inf})}$ and $\Delta^{(i)}$, both specified according to our hypothesis. As we will show in the next section, however, for any reasonably large-enough choice of $\Delta^{(i)}$, the effects of the value of $\Delta^{(i)}$ on the Bayes factor is not significant and will not affect the ranking between the two hypotheses. Therefore, whether we choose $\Delta^{(i)} = 1$ or $\Delta^{(i)} = 10$, both of which are much larger than the astrophysically motivated value of $\mathcal{P}_{\ell m}$, corresponding to $|\mathcal{P}_{\ell m}| \sim \mathcal{O}(10^{-48})$, our conclusions will be unaffected.

Let us conclude this section by pointing out that the above calculations can be easily extended to a detector network that contains more detectors. To apply the method to a detector network, one just sums over the detector pairs when calculating the following quantities [102]

$$
\begin{aligned}
\mathbf{j} &= \sum_I \sum_{J>I} \mathbf{j}^{(IJ)}, \\
\mathbf{Q} &= \sum_I \sum_{J>I} \mathbf{Q}^{(IJ)},
\end{aligned}
\tag{68}
$$

where $\mathbf{j}^{(IJ)}$ and $\mathbf{Q}^{(IJ)}$ are respectively the j vector and $\mathbf{Q}$ of the detectors I and J [cf. Eq. (45)].

V. MOCK DATA ANALYSIS

In this section, we illustrate the accuracy of our analysis in extracting the angular structures of a GWB by applying it to mock data. We will first explain the general setup of

different mock data analyses. Then, we will apply our analysis to different sets of mock data, each corresponding to a different level of anisotropy. We will show that our analysis can extract the angular structure of different types of anisotropic sources with excellent accuracy.

A. General setup

As the likelihood [Eq. (27)] does not explicitly depend on the strain data measured by individual detectors but on their correlation, we follow [83] and directly simulate the cross-spectral density of data segments in the frequency domain,

$$C_{\mathrm{inj}}(f,t) = C_n(f,t) + H_\alpha(f) \sum_{\ell=0}^{\ell_{\mathrm{max}}^{(\mathrm{inj})}} \sum_{m=-\ell}^{\ell} \gamma_{\ell m}(f,t) \mathcal{P}_{\ell m}^{(\mathrm{inj})}.$$
(69)

In this expression, $\ell_{\mathrm{max}}^{(\mathrm{inj})}$ and $\mathcal{P}_{\ell m}^{(\mathrm{inj})}$ are the maximum ℓ and the spherical-harmonic components of the simulated GWB contained in the mock data, respectively. Note that $\ell_{\mathrm{max}}^{(\mathrm{inj})}$ is in general different from $\ell_{\mathrm{max}}^{(\mathrm{inf})}$ because the maximum ℓ that a GWB corresponds to can, in general, be different from the maximum ℓ that we choose to infer. Through the mock data analyses, we choose $\ell_{\mathrm{max}}^{(\mathrm{inf})} = 1, 2, \ldots, 10$, but in general $\ell_{\mathrm{max}}^{(\mathrm{inf})}$ can be freely adjusted for analyzing actual data. Note also that by directly simulating the cross-spectral density in the frequency domain, we are assuming A.4 and ignoring the cross- and autocorrelations present in the time-domain data. In practice, the analysis should thus be applied to individual (windowed) segments and then optimally combine the result from individual segments to address the cross- and autocorrelations. Nonetheless, as pointed out when assuming that condition A.4 holds; if the windowing and optimal combinations are properly executed, the final results should agree well with calculations that use the likelihood and ignores these correlations, as shown in [99].

We study the effects of noise fluctuations by including $C_n(f,t)$ in the injected cross-spectral density of data segments in Eq. (69). In particular, $C_n(f,t)$ represents the cross-spectral density of the stationary Gaussian noise contained in the data. We simulate $C_n(f,t)$ by generating a random complex frequency sequence of zero mean and variance that satisfies [83]

$$\langle |C_n(f,t)|^2 \rangle - |\langle C_n(f,t) \rangle|^2 \approx \frac{N_I^{(\mathrm{n})}(f,t) N_J^{(\mathrm{n})}(f,t)}{\tau \Delta f},$$
(70)

where recall that τ is the length of the data segments, Δf is the frequency resolution, and $N_{I,J}^{(\mathrm{n})}(f,t)$ are the noise PSDs of the detectors I and J, respectively. Since the measured strain data contain both the instrumental noise and the signal when a GWB presents, the PSD of the strain data

measured by individual detectors will contain both the instrumental-noise PSD, $N_{I,J}^{(\mathrm{n})}(f,t)$, and the autocorrelated power of the responses due to a GWB,

$$N_{I,J}(f,t) = N_{I,J}^{(\mathrm{n})}(f,t) + S_h(f,t).$$
(71)

Hence, in practice, this $N_{I,J}^{(n)}$ is not the same as the PSD in Eq. (27). These two PSDs are extremely difficult to separate in an actual detection. Since we expect the signal to be weak, we just regard the measured strain PSD as the noise PSDs for the evaluation of the likelihood, at the cost of slightly reducing our search sensitivity [103]. To account for this effect, in our mock-data analyses we include both PSDs in our search when evaluating the likelihood and marginalized posteriors, but we only include $N_{I,J}^{(\mathrm{n})}(f,t)$ when simulating $C_n(f,t)$.

Other properties of the injection are chosen to remain in line with current GWB searches with advanced LIGO and Virgo detectors [82,94]. More specifically, for each mock analysis, we simulate data that consist of segments of equal time length $\tau = 192$ s. Since these mock data analyses are meant to represent proof-of-principle demonstrations, we only simulate data measured by the advanced LIGO Hanford and Livingston detectors at their design sensitivity. The PSD of the detectors is estimated with the exact frequency resolution of the cross-spectral density segments to avoid the need for coarse-graining data [11–14]. As the mock data contain only stationary Gaussian noise and the responses induced by the simulated stationary GWB, we drop the time dependence of the PSDs, so that $N_{I,J}(f,t) = N_{I,J}(f)$ and we do not notch the data at particular frequency bins. We assume the data start at the starting time of the third observing run of the advanced LIGO and Virgo detectors. We also focus on simulating and searching for GWB with $\alpha = 0, 2/3$, and 3 because GWBs characterized by these α are under extensive search and correspond to astrophysical interesting sources. More explicitly, $\alpha = 0$ describes the GWB produced by cosmic strings formed during the end of cosmological inflation [35–51,53–61]. The spectral tilt $\alpha = 2/3$ characterizes the GWB produced by CBCs [22–25], and $\alpha = 3$ approximately describe the GWB produced by supernova [31–34]. The explicit value of the spherical-harmonic components of the simulated GWBs will be given individually in the corresponding sections below.

To gauge the accuracy of measuring w_i from the simulated data, for different i, we define two measures. The first measure is the error of a specific w_i relative to σ_i,

$$\delta_i = \frac{\mu_i - w_i^{(\mathrm{inj})}}{\sigma_i},$$
(72)

where recall that μ_i is the mean of the marginalized posterior of w_i, while σ_i is its variance. By examining
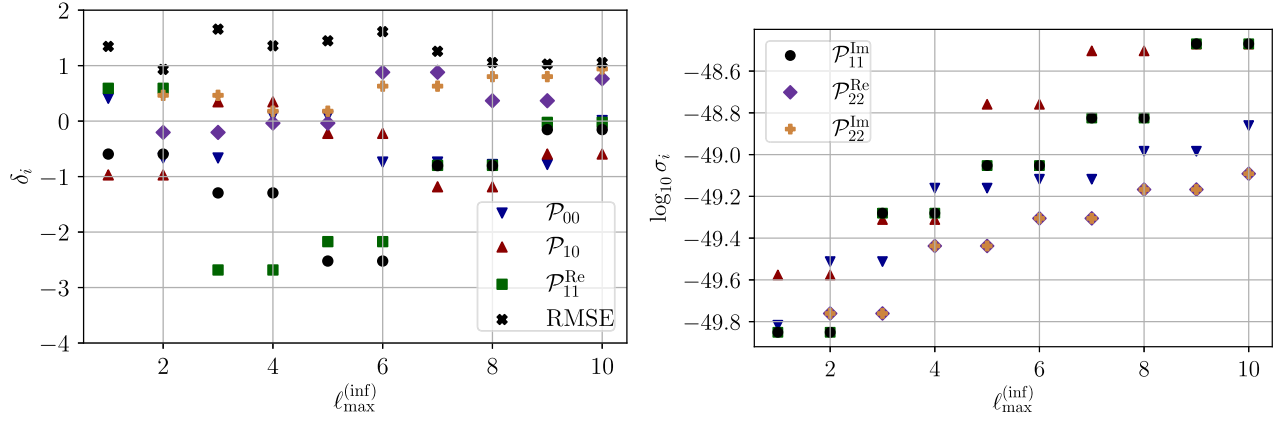
FIG. 1.   The measurement bias δ_i [left, see Eq. (72) in the main text] and measurement uncertainty σ_i (right) of some $\mathcal{P}_{\ell m}$, obtained by applying our analysis to one year of mock data, which contain solely stationary Gaussian noise, as a function of $\ell_{\max}^{(\inf)}$, assuming $\alpha = 2/3$. Note that δ_i has been scaled by σ_i in its definition. For the purpose of illustration, we only show δ_i and σ_i for $\mathcal{P}_{00}, \mathcal{P}_{10}, \mathcal{P}_{11}$, and $\mathcal{P}_{22}$. Observe that $|\delta_i| < 3$ for all $\mathcal{P}_{\ell m}$, indicating that the results are consistent with the fact that the mock data contain no signal to $3\sigma_i$ confidence. Observe also that, as $\ell_{\max}^{(\inf)}$ approaches 10, δ_{RMSE} [see Eq. (73) for definition] approaches to ~1 steadily, indicating that overall, the recovered w_i is consistent with zero within ~1σ.

δ_i, we can study the effects of $\ell_{\max}^{(\inf)}$ on measurement accuracy of w_i. If $\delta_i = N$, then the best-fit w_i is $N\sigma$ away from the injected value. Therefore, when δ_i is close to zero, then the recovered w_i is perfectly consistent with the injected $w_i^{(\inj)}$. However, due to the presence of noise fluctuations, we expect that $|\delta_i|$ can occasionally be as large as ~3 (see e.g. [82], where the SNR of a GWB is 3.6, but one still cannot claim a detection). In what follows, we calculate the marginalized posterior of many parameters, but we will only show results (e.g. δ_i and σ_i) for a subset of them. In the Supplemental Material [104], we present all results obtained by our mock data analysis.

The second measure is the root mean square error (RMSE),

$$\delta_{\mathrm{RMSE}} = \sqrt{\frac{1}{(\ell^{(\inf)} + 1)^2} \sum_i \sigma_i^2}, \qquad (73)$$

where $\sum_i$ stands for summation over the index i that labels the vector w_i, corresponding to $\mathcal{P}_{\ell m}$ for $0 \leq \ell \leq \ell_{\max}^{(\inf)}$. Heuristically, δ_{RMSE} gives the averaged deviation of the measured w_i from the simulated w_i relative to σ_i. Therefore, unlike δ_i, δ_{RMSE} measures the overall accuracy of all w_i.

B. Pure-noise injection

We first apply our formalism to 365 days of mock data that contain only pure noise. The left panel of Fig. 1 shows δ_i and the right panel the base-10 logarithm of σ_i, both as a function of $\ell_{\max}^{(\inf)}$. To illustrate, we show only the marginalized posterior of $\mathcal{P}_{00}, \mathcal{P}_{10}, \mathcal{P}_{11}^{\mathrm{Re}}, \mathcal{P}_{11}^{\mathrm{Im}}, \mathcal{P}_{22}^{\mathrm{Re}}$, and $\mathcal{P}_{22}^{\mathrm{Im}}$. Since the results of different α are quantitatively the same, for illustration, we only show $\alpha = 2/3$, corresponding to the

GWB formed by CBCs. First, we observe that for all $\ell_{\max}^{(\inf)}$, $|\delta_i| < 3$; this means that μ_i is consistent with $w_i^{(\inj)} = 0$ to $3\sigma_i$, indicating that we can accurately pinpoint the fact that the mock data contain no GWB. Second, we observe that σ_i increases with $\ell_{\max}^{(\inj)}$ and it is expected. Increasing $\ell_{\max}^{(\inj)}$ introduces more (unnecessary) free parameters whose measure uncertainty correlates with those associated with the spherical-harmonic components of smaller ℓ, deteriorating the overall measurement accuracy.

We also check that $\tilde{\mathbf{Q}}^{(i)}$ is numerically well conditioned because the evaluation of μ_i and σ_i involves the inverse of $\tilde{\mathbf{Q}}^{(i)}$. To this end, we compute the individual condition number κ_i of the matrix $\tilde{\mathbf{Q}}^{(i)}$, which is defined by[3]

$$\kappa_i = \frac{\lambda_i^{\min}}{\lambda_i^{\max}}, \qquad (74)$$

where $\lambda_i^{\min}$ and $\lambda_i^{\max}$ are the eigenvalues of $\tilde{\mathbf{Q}}^{(i)}$ that have the smallest and largest modulus, respectively. A larger κ_i implies that $\tilde{\mathbf{Q}}^{(i)}$ is easier to invert numerically and $\kappa_i = 0$ means that $\tilde{\mathbf{Q}}^{(i)}$ is singular. Then, we define the overall condition number κ as

$$\kappa = \min_i \kappa_i. \qquad (75)$$

Since κ is essentially the lower bound of κ_i, a larger κ implies that $\tilde{\mathbf{Q}}^{(i)}$ is easier to numerically invert for all i.

---

[3]This is not the usual definition of the condition number of a matrix, which is defined as the ratio between the eigenvalue of the largest modulus and that of the least modulus. The definition in this paper follows the convention in the literature of the search of anisotropic GWB, e.g., [81,105].
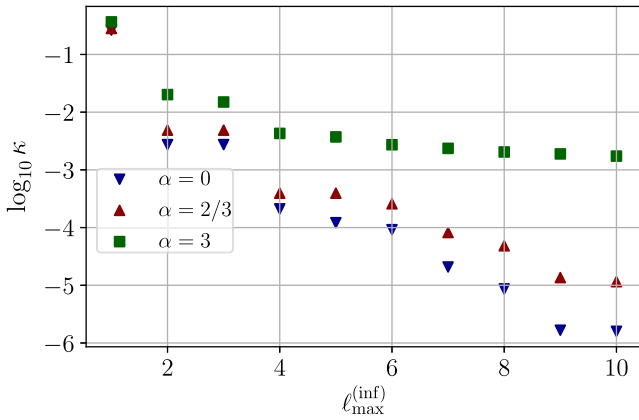
FIG. 2. The overall condition number [κ, defined by Eq. (75)] of $\tilde{\mathbf{Q}}^{(i)}$ [defined below Eq. (48)], whose inverse is required for the evaluation of the marginalized posterior, as a function of $\ell_{\max}^{(\mathrm{inf})}$ for $\alpha = 0, 2/3$, and 3. A larger κ implies that $\tilde{\mathbf{Q}}^{(i)}$ is more numerically invertible for all i. Observe that $\kappa > 10^{-6}$ for $\ell_{\max}^{(\mathrm{inf})} = 10$ and all α, implying that $\tilde{\mathbf{Q}}^{(i)}$ for all i and α can be inverted within double precision.

Figure 2 shows κ as a function of $\ell_{\max}^{(\mathrm{inf})}$ for $\alpha = 0, 2/3$, and 3. Observe that, for $\alpha = 0, 2/3$ and 3, $\kappa > 10^{-6}$ for $\ell_{\max}^{(\mathrm{inf})} = 10$, the upper limit of $\ell_{\max}$ considered throughout the paper. This means that $\tilde{\mathbf{Q}}^{(i)}$ for all i and α can be inverted within double precision without numerical issues.[4] To further check that $\tilde{\mathbf{Q}}^{(i)}$ is properly inverted, we compute the max norm, the maximum of the modulus of the elements of a matrix, of the following error matrix,

$$E^{(i)} = I - \tilde{\mathbf{Q}}^{(i)}\mathbf{M}^{(i)}, \qquad (76)$$

which should be a zero matrix if $\mathbf{M}^{(i)}$ is exactly equal to the inverse of $\tilde{\mathbf{Q}}^{(i)}$. We find that the max norm of E_i is at most 10^{-10} for different i and α, confirming that $\tilde{\mathbf{Q}}^{(i)}$ can be inverted within double precision without numerical issues.

C. Time-independent dipole

We now validate our method by recovering a simulated time-independent dipole with $\alpha = 0, 2/3$, and 3 from 365 days of mock data. The simulated dipole signals are motivated by the dipole produced by the peculiar motion of the Solar System barycenter relative to the cosmic rest frame.[5]

---

[4]In principle, a regularization scheme, such as that presented in [81,90,106], can also be applied when inverting $\tilde{\mathbf{Q}}^{(i)}$, but such regularization may bias results [80,81,107–110].

[5]The orbit of the Earth around the Solar System barycenter induces a smaller time-dependent kinematic dipole signal, even though it is routine to analyze the stochastic GWB from the perspective of the solar system barycenter. This requires special approaches to extract [99,111,112].

For all α, the nonzero spherical-harmonic components from the mock data injections are

$$\mathcal{P}_{00}^{(\mathrm{inj})} = 4.69 \times 10^{-46},$$
$$\mathcal{P}_{10}^{(\mathrm{inj})} = -1.16 \times 10^{-47},$$
$$\mathcal{P}_{11}^{(\mathrm{inj})} = (6.60 + 1.41i) \times 10^{-47}, \qquad (77)$$

and $\ell_{\max}^{(\mathrm{inj})} = 1$. These spherical-harmonic components are chosen so that their value is significantly larger than the corresponding measurement uncertainty, facilitating the validation of our analysis. The monopole signal is included so that the intensity map is positive in all sky directions.

Figure 3 shows δ_i and σ_i for $\mathcal{P}_{00}, \mathcal{P}_{10}, \mathcal{P}_{11}^{\mathrm{Re}}, \mathcal{P}_{11}^{\mathrm{Im}}, \mathcal{P}_{22}^{\mathrm{Re}}$, and $\mathcal{P}_{22}^{\mathrm{Im}}$ with $\alpha = 2/3$, obtained by analyzing the mock data with the simulated dipole signal. Observe that $|\delta_i| < 3$ for different $\ell_{\max}^{(\mathrm{inf})}$, which shows the robustness of our analysis in two ways. First, our analysis can correctly infer different $\mathcal{P}_{\ell m}$ to $3\sigma_i$ confidence. In other words, our analysis does not mistake the angular structure of $\ell \leq \ell_{\max}^{(\mathrm{inj})}$ with the angular structure of $\ell_{\max}^{(\mathrm{inj})} < \ell \leq \ell_{\max}^{(\mathrm{inf})}$. Second, choosing different $\ell_{\max}^{(\mathrm{inf})}$ does not significantly affect our measurement of $\mathcal{P}_{\ell m}^{(\mathrm{inj})}$. Thus, one can adjust $\ell_{\max}^{(\mathrm{inj})}$ for the search of different GWB without having to worry that the results will be significantly affected by this choice. Note that the measurement uncertainty for different i is slightly larger than those shown in Fig. 1 due to the contribution of the detectors' PSD from the monopole of the simulated GWB.

D. Galactic plane distribution

Our last mock data analysis concerns the GWB emitted by sources populating the galactic plane. For the mock-data challenge of the galactic-plane signal, we focus on $\alpha = 2/3$ because we expect that the results obtained with other choices of α will be quantitatively similar. We choose to focus on $\alpha = 2/3$ because this spectral index corresponds to the background due to CBCs, the only type of GW sources that the Advanced LIGO and Virgo detectors have detected so far. To investigate the performance of our analysis when extracting anisotropic GWB signals of different signal-to-noise ratio (SNR, ρ), we simulate galactic-plane signals of different SNRs but we reduce the total time length of the mock data of each SNR to 30 days. The measurement results from analyzing data of longer time length can be estimated by scaling the SNR, which is proportional to the square root of the integration time. The $\mathcal{P}_{\ell m}$ of the galactic-plane signal that we simulate are

$$\mathcal{P}_{\ell m}^{(\mathrm{inj})} = \epsilon \mathcal{P}_{\ell m}^{(\mathrm{GP})}, \qquad (78)$$

where ϵ controls the overall amplitude (and SNR) of the galactic-plane signal and $\mathcal{P}_{\ell m}^{(\mathrm{GP})}$ are explicitly given in the
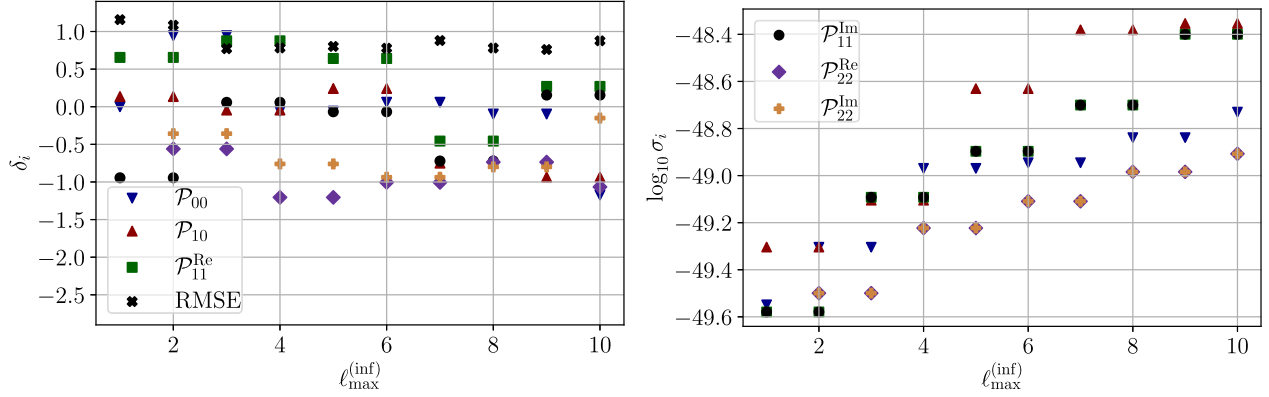
FIG. 3.   Same as Fig. 1 but with mock data that contains a simulated time-independent dipole (i.e. $\mathcal{P}_{\ell \geq 2m} \equiv 0$). Observe that $|\delta_i| < 3$ for different $\mathcal{P}_{\ell m}$, indicating that our analyses can accurately measure different $\mathcal{P}_{\ell m}$ to 3σ confidence. Moreover, comparing the δ_i and σ_i of $\mathcal{P}_{10}$ and $\mathcal{P}_{11}$ with those of $\mathcal{P}_{22}$ suggests that our analysis is unlikely to mistake a spherical-harmonic component of $\ell \leq \ell_{\max}^{(\text{inj})}$ with another spherical-harmonic component of $\ell > \ell_{\max}^{(\text{inj})}$. Observe also that, as $\ell_{\max}^{(\text{inf})}$ approaches 10, δ_{RMSE} approaches 1 steadily, indicating that overall, the recovered w_i is consistent with $w_i^{(\text{inj})}$ within $\sim 1\sigma$.

Appendix. When choosing these $\mathcal{P}_{\ell m}^{(\text{inj})}$, we set $\ell_{\max}^{(\text{inj})} = 7$, following [83], because this is sufficient to capture the fine angular structures of such a galactic-plane signal.

The intensity map of the simulated galactic signal is visualized in the top left panel of Fig. 4, produced using HEALPix [113,114]. The brightness of the color map in all panels represents the intensity, and the intensity of all maps is scaled by a number such that the maximum intensity of each panel is normalized to one. The top-right, middle-left, middle-right, bottom-left, and bottom-right panels show intensity maps when $\epsilon = 1, 10^{0.5}, 10^1, 10^{1.5}, 10^2$, respectively, constructed using our analysis, with the spherical harmonic components of the recovered background taken to be the μ_i of Eq. (53). As we increase ϵ, the SNR of the signal increases (see the top horizon axis of Fig. 5 for the monopole SNR of each ϵ), and the reconstructed intensity map is increasingly consistent with the simulated intensity map. At $\epsilon = 10^2$, the reconstructed intensity map shows almost no visual differences from the original intensity map. The close consistency between the simulated and recovered intensity maps demonstrates the ability of our formalism to resolve detailed and sophisticated angular structures of GWBs.

Despite the close visual consistency, we also quantitatively assess the consistency between the simulated and reconstructed intensity maps by defining the match,

$$\mathcal{M} = \frac{\sum_i w_i^{(\text{inj})} \mu_i}{\sqrt{\sum_i (w_i^{(\text{inj})})^2} \sqrt{\sum_i \mu_i^2}}, \qquad (79)$$

where $w_i^{(\text{inj})}$ is the value of the real or imaginary parts of $\mathcal{P}_{\ell m}^{(\text{inj})}$ corresponding to the index i, and μ_i is the recovered value, given by Eq. (53). A match closer to unity implies a

more faithful recovery. If the reconstructed intensity map is identical to the simulated intensity map, then $\mathcal{M} = 1$. Figure 5 shows $\mathcal{M}$ of the simulated galactic-plane signal as a function of ϵ, with the top horizontal axis denoting the SNR of the monopole part of the simulated background of the corresponding ϵ. Observe that $\mathcal{M}$ increases to ~ 1 as ϵ increases. This is reasonable because, as the background SNR increases, the angular structures of the simulated anisotropic background can also be more clearly detected. Moreover, when the SNR reaches ~ 400, which is an SNR that can be achieved within about a year if we detect a GWB of $\Omega_{\text{GW}}(f_{\text{Ref}}) \sim 10^{-10}$ with the next-generation detectors [99,115], our analysis can recover the intensity map with a match very close to one, indicating its applicability to the realistic detection of a GWB.

Besides the reconstruction of the intensity map, we also compute the Bayes factor between the hypotheses that there is an anisotropic GWB in the signal and that there is only noise [see Eq. (30)], given an injection of an anisotropic GWB from the mock galactic-plane signal. The left panels of Fig. 6 show the natural logarithm of the Bayes factor as a function of $\ell_{\max}^{(\text{inf})}$, obtained by analyzing the galactic-plane signals of different ϵ, choosing $\Delta^{(i)} = 1$ (inverted blue triangles) and $\Delta^{(i)} = 10$ (red triangles). Both of these choices of $\Delta^{(i)}$ correspond to a prior of width much larger than the astrophysically motivated value of w_i, which should be of $\mathcal{O}(10^{-48})$ (see Figs. 1 and 3). The dashed vertical line denotes the $\ell_{\max}^{(\text{inj})}$ of the simulated galactic-plane signal. Observe that, in general, for all ϵ, $\log \mathcal{B}(\ell_{\max}^{(\text{inf})}|\Delta^{(i)} = 1)$ is slightly larger than $\mathcal{B}(\ell_{\max}^{(\text{inf})}|\Delta^{(i)} = 10)$ because it has a narrower prior. Nonetheless, despite these slight differences, both choices of $\Delta^{(i)}$ lead to a similar Bayes factor. This suggests that, for a reasonably large $\Delta^{(i)}$, the explicit choice of the prior width does not
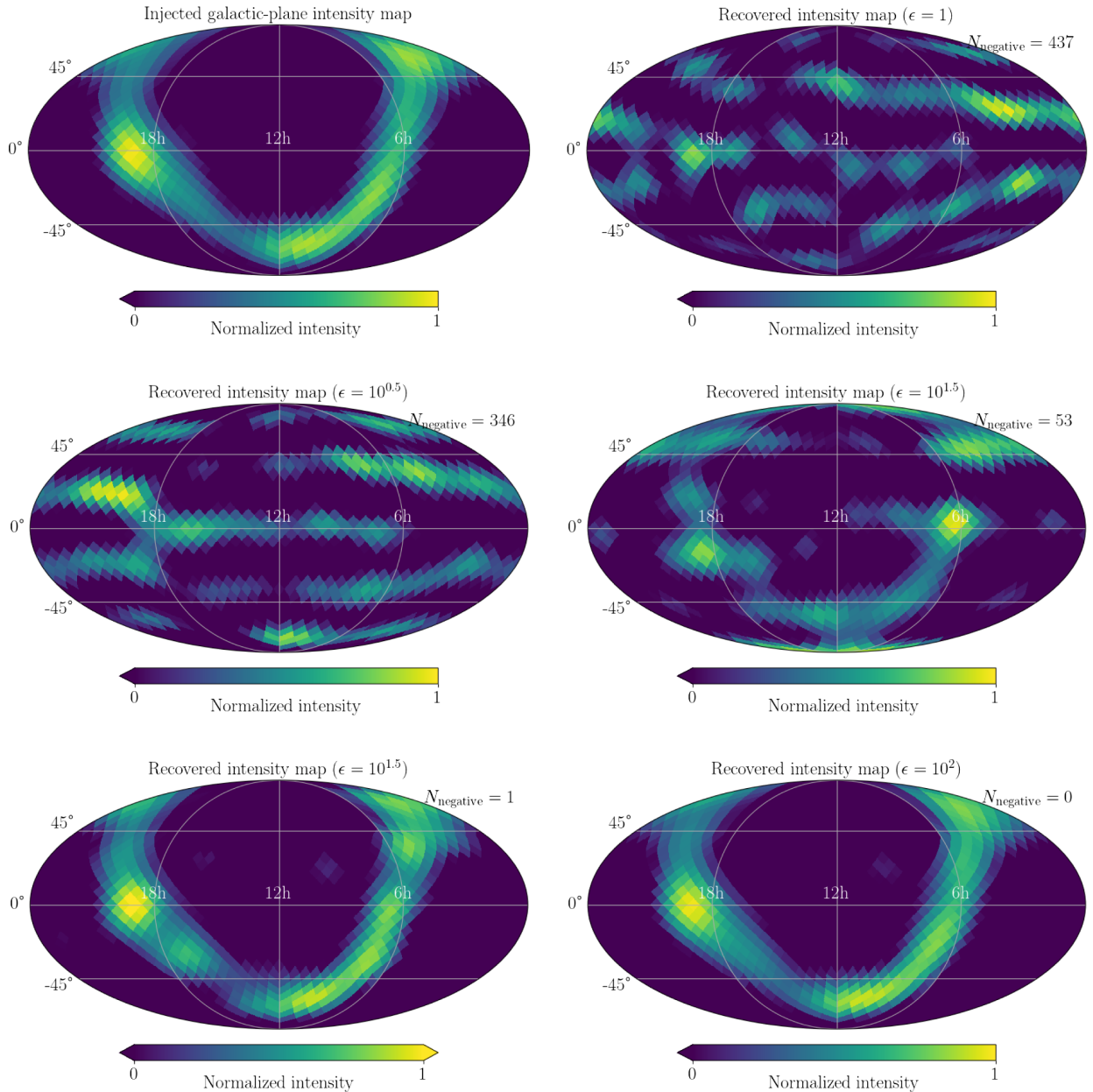
FIG. 4.   The top, left panel shows the angular distribution of gravitational-wave backgrounds produced by sources populating the galactic plane, which we simulate assuming $\alpha = 2/3$ and that the signal lasts for 30 sidereal days. The rest of the figures are the recovered intensity maps from mock data containing signals of different strengths, as characterized by ϵ [see Eq. (78) for definition]. The intensity maps are presented in the equatorial coordinate system. The brightest spot (on the left) is the galactic center. All figures are visualized by Mollweide projections and contain 1200 pixels ($N_{\mathrm{side}} = 10$). The number N_{negative} at the top right corner is the number of pixel of the recovered map which has negative intensity. The signal-to-noise ratio of the monopole of the background when $\epsilon = 1$ is 15.9 and that when $\epsilon = 10^2$ is 1050. To show the intensity contrast across different sky directions, the brightness of the color in all panels represents the intensity, and the intensity of all maps is scaled by a number such that the maximum intensity of the simulated map is normalized to one. Observe that, as the signal-to-noise ratio of the gravitational-wave background increases, our analysis can recover an intensity map that is increasingly accurate and consistent with the simulated angular distribution. Moreover, when the signal-to-noise ratio of the monopole part of the background has reached $\sim 10^3$, the reconstructed intensity maps show almost no visual difference relative to the simulated map. This close consistency shows that our formalism is capable of resolving detailed and sophisticated angular structures of a gravitational-wave background.
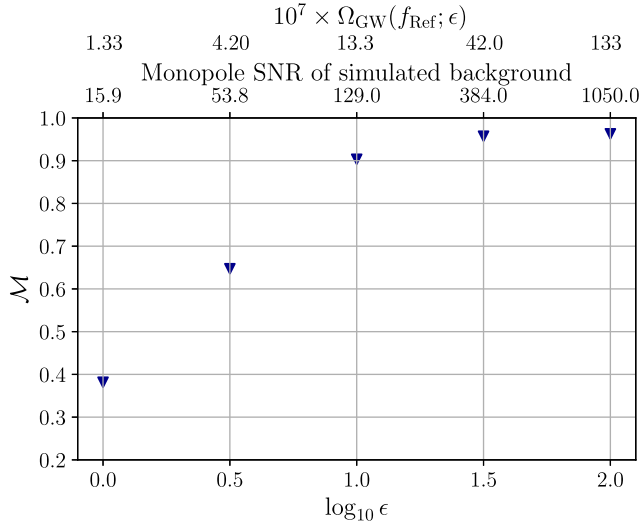
FIG. 5. The match between the simulated intensity map of the gravitational-wave background and that recovered using the untargeted Bayesian search [see Eq. (79) for the definition]. The lower horizontal axis represents the base-10 logarithm of ϵ, a proportionality constant that regulates the amplitude of the simulated gravitational-wave background. The upper horizontal axis represents the signal-to-noise ratio of the monopole of the simulated background of the corresponding ϵ. On top of the signal-to-noise ratio, we also include the value of $\Omega_{\mathrm{GW}}(f;\epsilon)$ at the corresponding ϵ for reference. A match closer to one indicates a more faithful recovery of the intensity map of the gravitational-wave background using our analysis. Observe that as ϵ increases (or equivalently, as the signal-to-noise ratio increases), $\mathcal{M}$ also increases, showing that the recovered intensity map is increasingly accurate for louder signals. Moreover, the match is close to one when the monopole signal-to-noise ratio is about 400, which can be achieved within approximately a year using next-generation detectors if a gravitational-wave background of $\Omega_{\mathrm{GW}}(f_{\mathrm{Ref}}) \sim 10^{-10}$ is present. This suggests that our untargeted Bayesian search can indeed be applied to actual gravitational-wave detection in the future.

significantly affect the Bayes factor and the hypothesis ranking for the search of anisotropic GWBs. In this sense, our analysis is robust against different choices of $\Delta^{(i)}$, provided that $\Delta^{(i)}$ is reasonably large. Individually, we observe that for a given ϵ, the Bayes factor first increases until it reaches a maximum at a given $\ell_{\max}^{(\mathrm{inf})}$, and then it decreases. Let us denote the $\ell_{\max}^{(\mathrm{inf})}$ that maximizes the Bayes factor $\ell_{\max}^{\mathcal{B}}$ and show it on Fig. 6 with a dotted vertical line. Observe further that $\ell_{\max}^{\mathcal{B}}$ depends on ϵ. For a larger ϵ, corresponding to a louder signal, $\ell_{\max}^{\mathcal{B}}$ is more consistent with $\ell_{\max}^{(\mathrm{inf})}$, until eventually $\ell_{\max}^{\mathcal{B}}$ coincides with $\ell_{\max}^{(\mathrm{inf})}$ in the high signal-to-noise ratio scenario. This behavior is reasonable if one interprets the $\ell_{\max}^{\mathcal{B}}$ as the maximal resolvable angular scale of the background. As we increase $\ell_{\max}^{(\mathrm{inf})}$ until $\ell_{\max}^{\mathcal{B}}$, we are introducing more parameters in the model that are necessary for a more faithful description of the

detectable anisotropic GWB signal. More precisely, even if we increase the number of inference parameters, the increase in the marginalized likelihood (the numerator of the Bayes factor) still compensates for the increase in the prior volume. Thus, the hypothesis that the detected GWB has nonzero $\mathcal{P}_{\ell_{\max}^{\mathcal{B}} m}$ for at least one m between $-\ell_{\max}^{\mathcal{B}}$ and $\ell_{\max}^{\mathcal{B}}$ is increasingly favored by the data. But as we further increase $\ell_{\max}^{(\mathrm{inf})}$, the new model parameters are redundant because the detected background shows no resolvable angular structures of the corresponding angular scale. The hypothesis that a GWB signal of $\ell_{\max}^{(\mathrm{inf})} > \ell_{\max}^{\mathcal{B}}$ is detected in the data is now *no longer* better supported by the data than the hypothesis that the signal contains only up to $\ell_{\max}^{\mathcal{B}}$, which explains the decrease. Finally, if the signal is louder, then we can naturally detect the finer angular structures (corresponding to a larger ℓ) of the simulated background more confidently. This explains the increasing consistency between $\ell_{\max}^{(\mathrm{inf})}$ and $\ell_{\max}^{\mathcal{B}}$ as ϵ increases until $\ell_{\max}^{\mathcal{B}}$ essentially coincides with $\ell_{\max}^{(\mathrm{inj})}$ in the high SNR limit, when ϵ is large. This behavior could be used to decide which $\ell_{\max}^{(\mathrm{inf})}$ is suitable for a particular search, which is also consistent with the discussion in [83].

Apart from competing $H_{\ell_{\max}^{(\mathrm{inf})}}$ against H_{null}, we can also compete $H_{\ell_{\max}^{(\mathrm{inf})}}$ against $H_{\overline{\ell}_{\max}^{(\mathrm{inf})}}$, where $\overline{\ell}_{\max}^{(\mathrm{inf})}$ is another maximum angular scale included in the inference. This can be done by computing the Bayes factor between $H_{\ell_{\max}}$ and $H_{\overline{\ell}_{\max}}$, which is simply

$$\mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}} = \frac{p(\{C\}|H_{\overline{\ell}_{\max}^{(\mathrm{inf})}})}{p(\{C\}|H_{\ell_{\max}^{(\mathrm{inf})}})} = \frac{\mathcal{B}(\overline{\ell}_{\max}^{(\mathrm{inf})})}{\mathcal{B}(\ell_{\max}^{(\mathrm{inf})})}. \tag{80}$$

If $\mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}} > 1$, then $H_{\overline{\ell}_{\max}^{(\mathrm{inf})}}$ is favored by the data.

The right panels of Fig. 6 show $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}}$ for $\overline{\ell}_{\max}^{(\mathrm{inf})} = 1$, 4, and 7 as a function of $\ell_{\max}^{(\mathrm{inf})}$, obtained by analyzing the galactic-plane signal of $\epsilon = 1$ (top right), 10 (middle right), and 10^2 (bottom right). We only show the results when $\Delta^{(i)} = 1$ because the $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}}$ of $\Delta^{(i)} = 10$ are qualitatively the same. From these panels, we observe the following four patterns in the behavior of $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}}$ as a function of $\ell_{\max}^{(\mathrm{inf})}$:

(1) $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}}$ *increases with* $\ell_{\max}^{(\mathrm{inf})}$, e.g. when $\epsilon = 1$, indicating that $H_{\ell_{\max}^{(\mathrm{inf})} \leq \overline{\ell}_{\max}^{(\mathrm{inf})}}$ is better preferred by the data.

(2) $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}}$ *decreases with* $\ell_{\max}^{(\mathrm{inf})}$, e.g. for $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}=1}$ when $\epsilon = 10$ and for $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}=1}$ and $\log \mathcal{B}_{\ell_{\max}^{(\mathrm{inf})}}^{\overline{\ell}_{\max}^{(\mathrm{inf})}=4}$

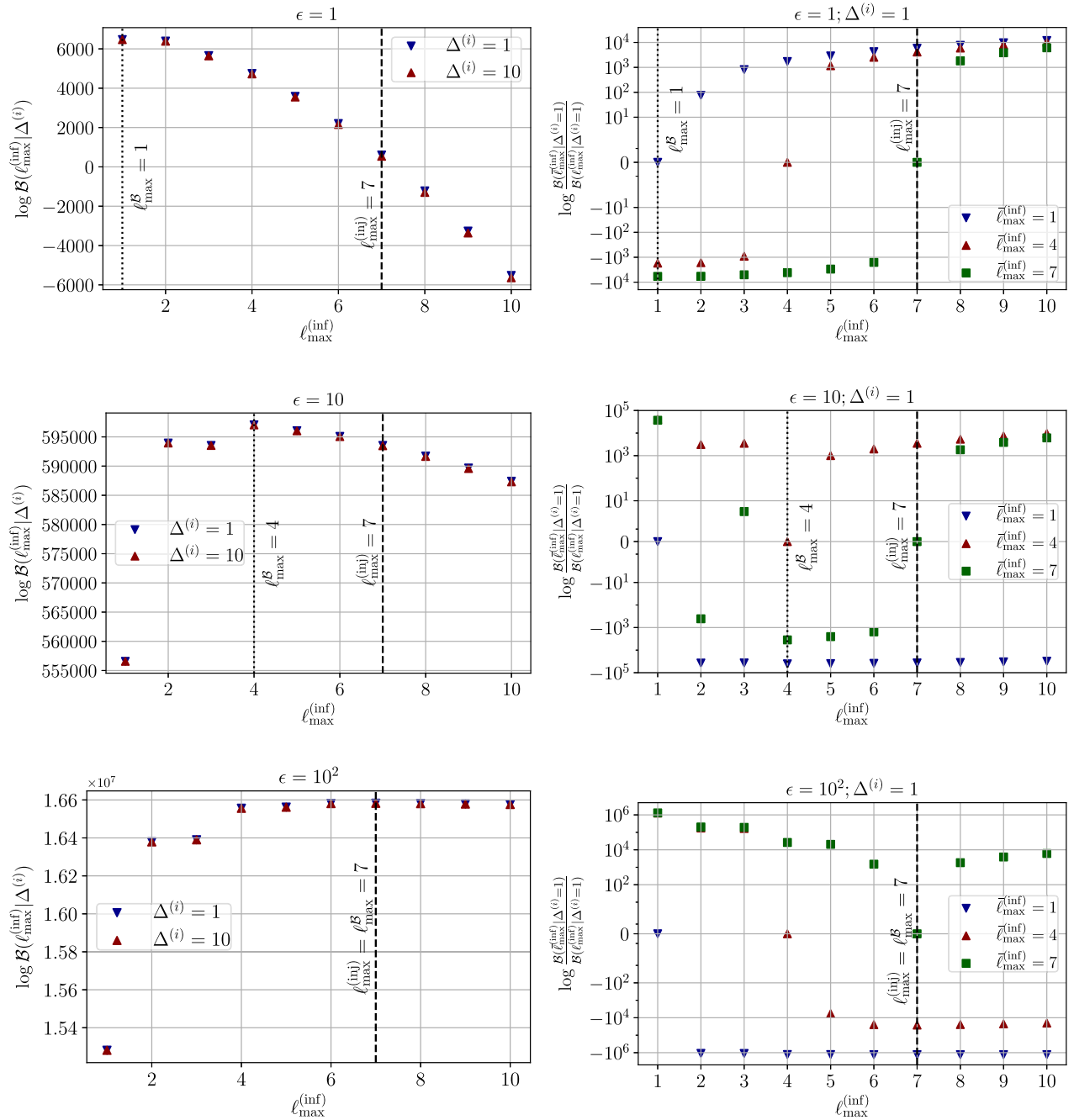FIG. 6.   To rank the hypotheses that we have detected a gravitational-wave background having angular structures up to the angular scale $\ell_{max}^{(inf)}$ ($H_{\ell_{max}^{(inf)}}$) from the data and that the data contain pure noise (H_{null}), we compute the Bayes factor between $H_{\ell_{max}^{(inf)}}$ and H_{null} (left panels) and that between $H_{\ell_{max}^{(inf)}}$ and $H_{\overline{\ell}_{max}^{(inf)}}$ (right panels), assuming different widths of the prior ($\Delta^{(i)}$). To facilitate the reading of the figures, we represent the maximal angular scale of the simulated background, $\ell_{max}^{(inj)} = 7$, with a dashed vertical line, and the angular scale at which the Bayes factor is maximized, $\ell_{max}^{\mathcal{B}}$, with a dotted vertical line. Observe that assuming different $\Delta^{(i)}$ does not significantly affect the resulting logarithm of the Bayes factor, indicating that our analysis is robust against the choice of prior. Observe also that as the amplitude of the background increases, as characterized by ϵ, $\ell_{max}^{\mathcal{B}}$ is increasingly consistent with $\ell_{max}^{(inj)}$, until they eventually coincide in the high signal-to-noise ratio scenario. This feature is reasonable if we interpret $\ell_{max}^{\mathcal{B}}$ as the maximum resolvable angular scale of the background. This pattern suggests that we can determine the angular scale that should be included in the inference analysis by locating the angular scale at which the Bayes factor is maximized, which is consistent with the finding of [83].

when $\epsilon = 10^2$, indicating that $H_{\ell_{\max}^{(\inf)} \geq \overline{\ell}_{\max}^{(\inf)}}$ is better preferred by the data.

(3) $\log \mathcal{B}_{\ell_{\max}^{(\inf)}}^{\overline{\ell}_{\max}^{(\inf)}}$ *first decreases, then increases with* $\ell_{\max}^{(\inf)}$ *and changes sign at some intermediate* $\ell_{\max}^{(\inf)}$, *e.g. for* $\log \mathcal{B}_{\ell_{\max}^{(\inf)}}^{\overline{\ell}_{\max}^{(\inf)}=7}$ *when* $\epsilon = 10$, *indicating that* $H_{\ell_{\max}^{(\inf)}}$ *is preferred over* $H_{\overline{\ell}_{\max}^{(\inf)}}$. *In other words,* $H_{\overline{\ell}_{\max}^{(\inf)}}$ *is not the hypothesis most preferred by the data.*

(4) $\log \mathcal{B}_{\ell_{\max}^{(\inf)}}^{\overline{\ell}_{\max}^{(\inf)}}$ *first decreases, then increases with* $\ell_{\max}^{(\inf)}$ *but remains non-negative, e.g. for* $\log \mathcal{B}_{\ell_{\max}^{(\inf)}}^{\overline{\ell}_{\max}^{(\inf)}=4}$ *when* $\epsilon = 10$ *and for* $\log \mathcal{B}_{\ell_{\max}^{(\inf)}}^{\overline{\ell}_{\max}^{(\inf)}=7}$ *when* $\epsilon = 10^2$, *indicating that* $H_{\overline{\ell}_{\max}^{(\inf)}}$ *is the hypothesis that is best supported by the data.*

By analyzing these patterns in the behavior of the log Bayes factor ratio shown in the right panels of Fig. 6, we again conclude that $\mathcal{B}(\ell_{\max}^{(\inf)})$ peaks at an angular scale that is increasingly consistent with the maximum angular scale contained in the injected background, which is consistent with what we observed from Fig. 5 and the left panel of Fig. 6.

VI. CONCLUDING REMARKS

In this paper, we presented a novel formalism to analytically marginalize the posterior of the spherical-harmonic components of the intensity map of a GWB in an untargeted Bayesian search. By prescribing a wide uniform prior for the real and imaginary parts of the spherical-harmonic components, we approximated the marginalized posterior (or likelihood) and Bayes factor as a Gaussian integral. The resulting marginalized posterior is also a Gaussian function. By reading off the mean and variance of the marginalized posterior, we can immediately determine the individual maximum posterior value of *many* spherical-harmonic components of the angular distribution of a GWB and gauge the associated measurement uncertainties. We validated our formalism by applying it to recover various anisotropic GWBs injections. For each simulated anisotropic GWB, our analysis accurately extracted the angular structures of the GWB within a 3σ interval. Furthermore, we are able to immediately evaluate the Bayes factor, which is largely unaffected by the width of the uniform prior. We showed that the Bayes factor is a reliable indicator of the angular scale that should be included in inference studies in a self-consistent way, which is also consistent with the findings of [83]. As the data products required for our analysis are similar and closely related to those used for existing spherical-harmonic decompositions of the actual data [81,82,85,86,90,116], we expect that, with minor

modifications, our analysis can be applied to actual data to efficiently extract GWB anisotropies along with other existing pipelines. Our analysis can also be applied to cross-check the results produced by other existing pipelines that search for anisotropic GWBs.

Our formalism presents several advantages in the detection of GWBs. First, our scheme makes possible Bayesian inference of a larger number of spherical-harmonic components of the angular distribution of a GWB in a reasonable timescale, leading to a much more model-independent Bayesian search of anisotropic GWBs. Prior to this work, in principle, we could treat all the spherical-harmonic components of interest as free parameters and attempt to infer them through Bayesian methods, but the computational cost and time needed to numerically sample the posterior would be huge [87]. To keep the computational time reasonable, previous Bayesian searches of anisotropic backgrounds either limited the number of spherical-harmonic components inferred (such as in [84]) or precomputed the spherical-harmonic components according to a given model and only inferred the overall amplitude of the anisotropic background (such as in [83]). By analytically marginalizing the posterior, we transform the problem into that of evaluating Gaussian integrals, greatly reducing the time needed to construct the marginalized posterior of spherical-harmonic components and compute the Bayes factor through Bayesian inference. The marginalized posterior of individual spherical-harmonic components can be used to construct an accurate intensity map of the GWB. The recovered intensity map can be compared with different GWB models, making the studies of GWBs more efficient. Second, our formalism is sufficiently flexible that it can be modified for the search of GWBs in various situations. Although this paper lays out the formalism of our method and presents a proof-of-principle analysis of synthetic data, considering only the joint detection of the LIGO Hanford and Livingston detectors, our approach can be straightforwardly extended to a network of detectors. Moreover, although this paper focused on searching for the GWB of a power-law spectrum, our approach can easily be adapted to the search for anisotropic GWBs of more sophisticated energy densities, such as those described by a broken-power law (such as in [117]).

Several aspects of our mock-data analyses differ from those carried out in real searches, but these differences do not undermine the performance of our method when applied to a future search. First, in our mock-data analyses, we only considered observations with the LIGO Hanford and Livingston detectors. In an actual search, the Virgo detector is operational, and while KAGRA is currently under development, this detector will join the network soon. Moreover, next-generation detectors, such as Cosmic Explorer [118] and Einstein Telescope [119], are also being planned. Our formalism can be easily extended to include

these detectors in a future analysis. With Virgo and future detectors included, the actual search sensitivity for the detection of a GWB will be greatly improved (assuming the LIGO-Virgo detectors are operating at their design sensitivity), which will also improve the accuracy and performance of our analysis. Hence, the results reported in this paper can be regarded as *conservative* estimates of what the future may hold. Second, when performing the short-time Fourier transform of the actual data in the time domain, this data will be Hann windowed to avoid spectral leakage [81,99]. To account for the windowing, we need to multiply the mean and variance by windowing factors [120]. The full use of the windowed data will then require that any windowed segment has an overlap of 50% with the Hann window and then be optimally combined. In this paper, since we are simulating the data in the frequency domain, we did not need to apply these procedures. By simulating the data directly in the frequency domain, we are effectively ignoring cross- and autocorrelations due to the serial dependence of the time-domain data. However, if the windowing and optimal combinations are correctly implemented, the results of the time-domain analysis should agree well with results that use the likelihood [Eq. (27)], which ignores these correlations, as shown in [99]. Third, the noise we considered in our mock-data challenges was stationary. In realistic data, nonstationary and/or non-Gaussian noise transients, also commonly known as "glitches," may occasionally occur and individual GW signals from CBCs may be present. When analyzing the actual data, data segments containing glitches and individual GW signals will be removed upon applying data-quality cuts [95,121–125]. Once these data segments are removed, our formalism can be applied as explained in this paper. Fourth, to fully demonstrate the accuracy of resolving the angular structures of GWBs with our method, we assumed strong GWB signals. In an actual detection scenario, we expect that GWBs to be much weaker. Nonetheless, the signal-to-noise ratio of a GWB detection is approximately proportional to the square root of the detection time [90,91]. Thus, in an actual detection, as the integration time is long enough, in principle, we can accumulate a sufficiently large signal-to-noise ratio so that the angular structures of the GWB can be accurately resolved by our analysis.

Several adaptations or explorations of our method can be carried out in the future to facilitate its implementation and improve its efficiency in the search for anisotropic GWBs in actual data. First, when no confident detection of a stochastic background is made, it is insightful to derive the 95% upper limit on the angular power spectrum, i.e. the 95% confidence region of

$$C_\ell = \left(\frac{2\pi^2 f_{\mathrm{ref}}^3}{3H_0^2}\right)^2 \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} [|\mathcal{P}_{\ell m}^{\mathrm{Re}}|^2 + |\mathcal{P}_{\ell m}^{\mathrm{Im}}|^2]. \qquad (81)$$

Since the individual $\mathcal{P}_{\ell m}^{\mathrm{Re}}$ and $\mathcal{P}_{\ell m}^{\mathrm{Im}}$ follow a Gaussian marginalized posterior whose mean is nonzero in general, as shown by our calculations, C_ℓ follows a *generalized* chi-squared distribution, which does not admit a simple closed-form analytic expression for its cumulative probability distribution function. Instead, numerical means are still required for constructing the cumulative probability distribution function of a generalized chi-squared distribution. Further effort must be devoted to either derive analytic results or to develop efficient numerical schemes that rapidly reconstruct the upper limit on C_ℓ when there is no GWB detection.

Second, to analytically marginalize the posterior, we prescribe a wide prior for the spherical harmonic components. Within the prior space, some spherical harmonic components actually correspond to an intensity map of negative intensity along some sky directions, which is not physical. Prescribing wide priors also makes our analysis suboptimal, in the sense that it may need much higher SNR to detect or resolve the angular structure of a GWB. One possible way to improve the method is to prescribe a conjugate normal prior for the spherical-harmonic components, which does not require the prior to be wide, thereby reducing the prior space that corresponds to negative intensity. However, the marginalized posterior and Bayes factor will then depend on the properties of the conjugate normal prior. Another possible way to improve the method is to make use of Clebsch-Gordan coefficients to parameterize the intensity map of a GWB [84]. However, the exponent of the likelihood in terms of Clebsch-Gordan coefficients becomes quartic in the relevant parameters. The analytical marginalization of such a posterior may be possible through an appropriate change of variables, but this requires further exploration.

Third, our analysis uses a spherical harmonic basis, which is well-suited for the search of wide-spread GWB sources. However, pointlike sources, such as nearby galaxy superclusters, may also contribute to anisotropic GWBs. These sources can be more adequately described using the pixel basis [80]. To include these pointlike sources in our search, we should explore extending our work to incorporate such a basis. Working with the pixel basis may require many more parameters to characterize GWB anisotropies than the spherical-harmonic basis. Therefore, in future work, one could explore how to perform the analysis with the pixel basis within a reasonable time frame.

Fourth, the marginalization of the likelihood in joint inferences of a GWB and individually resolvable GW signals requires further investigation. As mentioned here and also pointed out by [84], a motivation to measure the angular structure of GWBs in a Bayesian way is its integration with the existing search of other GW signals, such the those emitted by CBCs. One formalism that is capable of simultaneously searching for GWBs and individual GW signals is the "master-likelihood" method (also known as the hyper-likelihood approach) [126,127].

The marginalization of the master likelihood over the spherical harmonic components is certainly worth exploring to unite the search approaches of different types of GW signals for search efficiency reasons.

Finally, our formalism essentially assumes that we are searching for stationary GWBs. However, the kinematic dipole of a GWB induced by the proper motion of the Earth around the Solar System barycenter, a guaranteed anisotropic signal of GWBs [91,111,112], is time-dependent and requires a specially targeted method to implement in a search [99]. As this type of GWB signal varies over a timescale that is much longer than a sidereal day, we expect that our formalism can be straightforwardly adapted, say, by including this mild time dependence of the signal into the likelihood [Eq. (27)] before marginalization, to search for these GWB signals. Nonetheless, more exploration is still needed to determine the optimal way to modify our formalism to search for GWB signals with time dependence.

APPENDIX: $\mathcal{P}_{\ell m}$ OF THE GALACTIC-PLANE SIGNAL INJECTION

Below we provide the $\mathcal{P}_{\ell m}$ for the mock galactic plane signals that we simulated. The $\mathcal{P}_{\ell m}$ numbers are stored by m, in accordance with the convention of HEALPix. The relative intensity map is not alternated if one scales all $\mathcal{P}_{\ell m}$ by the same constant:

$$\mathcal{P}_{00}^{(\mathrm{GP})} = 6.24 \times 10^{-48},$$
$$\mathcal{P}_{10}^{(\mathrm{GP})} = -1.92 \times 10^{-50},$$
$$\mathcal{P}_{20}^{(\mathrm{GP})} = 1.28 \times 10^{-49},$$
$$\mathcal{P}_{30}^{(\mathrm{GP})} = -1.78 \times 10^{-49},$$

$$\mathcal{P}_{40}^{(\mathrm{GP})} = -1.03 \times 10^{-49},$$
$$\mathcal{P}_{50}^{(\mathrm{GP})} = -8.89 \times 10^{-50},$$
$$\mathcal{P}_{60}^{(\mathrm{GP})} = -3.63 \times 10^{-49},$$
$$\mathcal{P}_{70}^{(\mathrm{GP})} = -4.82 \times 10^{-50},$$
$$\mathcal{P}_{11}^{(\mathrm{GP})} = -2.90 \times 10^{-52} - 5.54 \times 10^{-50}i,$$
$$\mathcal{P}_{21}^{(\mathrm{GP})} = -9.05 \times 10^{-49} - 1.20 \times 10^{-49}i,$$
$$\mathcal{P}_{31}^{(\mathrm{GP})} = -1.39 \times 10^{-50} + 2.90 \times 10^{-49}i,$$
$$\mathcal{P}_{41}^{(\mathrm{GP})} = -5.72 \times 10^{-50} + 9.50 \times 10^{-50}i,$$
$$\mathcal{P}_{51}^{(\mathrm{GP})} = 6.89 \times 10^{-51} - 6.96 \times 10^{-50}i,$$
$$\mathcal{P}_{61}^{(\mathrm{GP})} = 3.86 \times 10^{-51} - 4.58 \times 10^{-50}i,$$
$$\mathcal{P}_{71}^{(\mathrm{GP})} = -2.10 \times 10^{-50} + 3.32 \times 10^{-50}i,$$
$$\mathcal{P}_{22}^{(\mathrm{GP})} = -9.95 \times 10^{-49} - 2.92 \times 10^{-49}i,$$
$$\mathcal{P}_{32}^{(\mathrm{GP})} = -8.70 \times 10^{-50} - 1.36 \times 10^{-49}i,$$
$$\mathcal{P}_{42}^{(\mathrm{GP})} = 3.92 \times 10^{-49} + 8.90 \times 10^{-50}i,$$
$$\mathcal{P}_{52}^{(\mathrm{GP})} = 7.52 \times 10^{-50} - 2.21 \times 10^{-49}i,$$
$$\mathcal{P}_{62}^{(\mathrm{GP})} = 3.31 \times 10^{-49} - 2.03 \times 10^{-50}i,$$
$$\mathcal{P}_{72}^{(\mathrm{GP})} = 3.46 \times 10^{-50} - 5.64 \times 10^{-50}i,$$
$$\mathcal{P}_{33}^{(\mathrm{GP})} = -1.85 \times 10^{-50} + 1.53 \times 10^{-50}i,$$
$$\mathcal{P}_{43}^{(\mathrm{GP})} = 5.59 \times 10^{-49} + 9.57 \times 10^{-50}i,$$
$$\mathcal{P}_{53}^{(\mathrm{GP})} = 1.14 \times 10^{-50} - 1.19 \times 10^{-50}i,$$
$$\mathcal{P}_{63}^{(\mathrm{GP})} = 1.69 \times 10^{-49} - 1.29 \times 10^{-50}i,$$
$$\mathcal{P}_{73}^{(\mathrm{GP})} = -4.30 \times 10^{-50} + 1.01 \times 10^{-49}i,$$
$$\mathcal{P}_{44}^{(\mathrm{GP})} = 3.46 \times 10^{-49} + 2.89 \times 10^{-49}i,$$
$$\mathcal{P}_{54}^{(\mathrm{GP})} = 1.00 \times 10^{-49} + 1.47 \times 10^{-49}i,$$
$$\mathcal{P}_{64}^{(\mathrm{GP})} = -2.59 \times 10^{-49} - 1.33 \times 10^{-49}i,$$
$$\mathcal{P}_{74}^{(\mathrm{GP})} = -1.68 \times 10^{-50} + 5.37 \times 10^{-50}i,$$
$$\mathcal{P}_{55}^{(\mathrm{GP})} = 1.29 \times 10^{-50} + 1.42 \times 10^{-49}i,$$
$$\mathcal{P}_{65}^{(\mathrm{GP})} = -3.35 \times 10^{-49} - 1.58 \times 10^{-49}i,$$
$$\mathcal{P}_{75}^{(\mathrm{GP})} = -7.04 \times 10^{-50} - 4.37 \times 10^{-51}i,$$
$$\mathcal{P}_{66}^{(\mathrm{GP})} = -1.66 \times 10^{-48} - 1.51 \times 10^{-49}i,$$
$$\mathcal{P}_{67}^{(\mathrm{GP})} = 3.05 \times 10^{-50} - 1.95 \times 10^{-49}i,$$
$$\mathcal{P}_{77}^{(\mathrm{GP})} = -5.54 \times 10^{-51} - 1.72 \times 10^{-49}i$$

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116,** 061102 (2016).

[2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence, Phys. Rev. Lett. **116,** 241103 (2016).

[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2, Phys. Rev. Lett. **118,** 221101 (2017); **121,** 129901(E) (2018).

[4] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170608: Observation of a 19-solar-mass binary black hole coalescence, Astrophys. J. Lett. **851,** L35 (2017).

[5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence, Phys. Rev. Lett. **119,** 141101 (2017).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, Phys. Rev. X **9,** 031040 (2019).

[7] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, Phys. Rev. X **11,** 021053 (2021).

[8] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run, arXiv:2111.03606 [Phys. Rev. X (to be published)].

[9] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190412: Observation of a binary-black-hole coalescence with asymmetric masses, Phys. Rev. D **102,** 043015 (2020).

[10] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190521: A Binary Black Hole Merger with a Total Mass of $150M_\odot$, Phys. Rev. Lett. **125,** 101102 (2020).

[11] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Upper Limits on the Stochastic Gravitational-Wave Background from Advanced LIGO's First Observing Run, Phys. Rev. Lett. **118,** 121101 (2017); **119,** 029901(E) (2017).

[12] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Search for the isotropic stochastic background using data from Advanced LIGO's second observing run, Phys. Rev. D **100,** 061101 (2019).

[13] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW150914: Implications for the Stochastic Gravitational Wave Background from Binary Black Holes, Phys. Rev. Lett. **116,** 131102 (2016).

[14] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170817: Implications for the Stochastic Gravitational-Wave Background from Compact Binary Coalescences, Phys. Rev. Lett. **120,** 091101 (2018).

[15] A. I. Renzini, B. Goncharov, A. C. Jenkins, and P. M. Meyers, Stochastic gravitational-wave backgrounds: Current detection efforts and future prospects, Galaxies **10,** 34 (2022).

[16] G. Agazie *et al.* (NANOGrav Collaboration), The NANOGrav 15-year data set: Evidence for a gravitational-wave background, arXiv:2306.16213.

[17] G. Agazie *et al.* (NANOGrav Collaboration), The NANOGrav 15-year data set: Detector characterization and noise budget, arXiv:2306.16218.

[18] G. Agazie *et al.* (NANOGrav Collaboration), The NANOGrav 15-year data set: Observations and timing of 68 millisecond pulsars, Astrophys. J. Lett. **951,** L9 (2023).

[19] A. D. Johnson *et al.* (NANOGrav Collaboration), The NANOGrav 15-year gravitational-wave background analysis pipeline, arXiv:2306.16223.

[20] G. Agazie *et al.* (NANOGrav Collaboration), The NANOGrav 15-year data set: Constraints on supermassive black hole binaries from the gravitational wave background, arXiv:2306.16220.

[21] A. Afzal *et al.* (NANOGrav Collaboration), The NANOGrav 15-year data set: Search for signals from new physics, Astrophys. J. Lett. **951,** L11 (2023).

[22] X.-J. Zhu, E. J. Howell, D. G. Blair, and Z.-H. Zhu, On the gravitational wave background from compact binary coalescences in the band of ground-based interferometers, Mon. Not. R. Astron. Soc. **431,** 882 (2013).

[23] Y. Zhao and Y. Lu, Stochastic gravitational wave background and eccentric stellar compact binaries, Mon. Not. R. Astron. Soc. **500,** 1421 (2020).

[24] A. J. Farmer and E. S. Phinney, The gravitational wave background from cosmological compact binaries, Mon. Not. R. Astron. Soc. **346,** 1197 (2003).

[25] G. Capurri, A. Lapi, C. Baccigalupi, L. Boco, G. Scelfo, and T. Ronconi, Intensity and anisotropies of the stochastic gravitational wave background from merging compact binaries in galaxies, J. Cosmol. Astropart. Phys. 11 (2021) 032.

[26] F. De Lillo, J. Suresh, and A. L. Miller, Stochastic gravitational-wave background searches and constraints on neutron-star ellipticity, Mon. Not. R. Astron. Soc. **513,** 1105 (2022).

[27] V. Ferrari, S. Matarrese, and R. Schneider, Stochastic background of gravitational waves generated by a cosmological population of young, rapidly rotating neutron stars, Mon. Not. R. Astron. Soc. **303,** 258 (1999).

[28] M. Surace, K. D. Kokkotas, and P. Pnigouras, The stochastic background of gravitational waves due to the f-mode instability in neutron stars, Astron. Astrophys. **586,** A86 (2016).

[29] P. D. Lasky, M. F. Bennett, and A. Melatos, Stochastic gravitational wave background from hydrodynamic turbulence in differentially rotating neutron stars, Phys. Rev. D **87,** 063004 (2013).

[30] P. A. Rosado, Gravitational wave background from rotating neutron stars, Phys. Rev. D **86,** 104007 (2012).

[31] B. Finkel, H. Andresen, and V. Mandic, Stochastic gravitational-wave background from stellar core-collapse events, Phys. Rev. D **105,** 063022 (2022).

[32] K. Crocker, T. Prestegard, V. Mandic, T. Regimbau, K. Olive, and E. Vangioni, Systematic study of the stochastic

gravitational-wave background due to stellar core collapse, Phys. Rev. D **95,** 063015 (2017).

[33] B. Finkel, H. Andresen, and V. Mandic, Stochastic gravitational-wave background from stellar core-collapse events, Phys. Rev. D **105,** 063022 (2022).

[34] A. Buonanno, G. Sigl, G. G. Raffelt, H.-T. Janka, and E. Muller, Stochastic gravitational wave background from cosmological supernovae, Phys. Rev. D **72,** 084001 (2005).

[35] C. Caprini and D. G. Figueroa, Cosmological backgrounds of gravitational waves, Classical Quantum Gravity **35,** 163001 (2018).

[36] M. C. Guzzetti, N. Bartolo, M. Liguori, and S. Matarrese, Gravitational waves from inflation, Riv. Nuovo Cimento **39,** 399 (2016).

[37] C. R. Contaldi and J. Magueijo, Unsqueezing of standing waves due to inflationary domain structure, Phys. Rev. D **98,** 043523 (2018).

[38] R. Easther, J. T. Giblin, Jr., and E. A. Lim, Gravitational Wave Production at the End of Inflation, Phys. Rev. Lett. **99,** 221301 (2007).

[39] A. A. Starobinskiĭ, Spectrum of relict gravitational radiation and the early state of the universe, Sov. J. Exp. Theor. Phys. Lett. **30,** 682 (1979).

[40] J. L. Cook and L. Sorbo, Particle production during inflation and gravitational waves detectable by ground-based interferometers, Phys. Rev. D **85,** 023534 (2012); **86,** 069901(E) (2012).

[41] M. S. Turner, Detectability of inflation produced gravitational waves, Phys. Rev. D **55,** R435 (1997).

[42] R. Easther and E. A. Lim, Stochastic gravitational wave production after inflation, J. Cosmol. Astropart. Phys. 04 (2006) 010.

[43] A. Romero, K. Martinovic, T. A. Callister, H.-K. Guo, M. Martínez, M. Sakellariadou, F.-W. Yang, and Y. Zhao, Implications for First-Order Cosmological Phase Transitions from the Third LIGO-Virgo Observing Run, Phys. Rev. Lett. **126,** 151301 (2021).

[44] M. Hindmarsh, M. Lüben, J. Lumma, and M. Pauly, Phase transitions in the early universe, SciPost Phys. Lect. Notes **24** (2021).

[45] D. J. Weir, Gravitational waves from a first order electroweak phase transition: A brief review, Phil. Trans. R. Soc. A **376,** 20170126 (2018).

[46] C. Caprini *et al.,* Science with the space-based interferometer eLISA. II: Gravitational waves from cosmological phase transitions, J. Cosmol. Astropart. Phys. 04 (2016) 001.

[47] T. Kahniashvili, L. Campanelli, G. Gogoberidze, Y. Maravin, and B. Ratra, Gravitational radiation from primordial helical inverse cascade MHD turbulence, Phys. Rev. D **78,** 123006 (2008); **79,** 109901(E) (2009).

[48] T. Kahniashvili, L. Kisslinger, and T. Stevens, Gravitational radiation generated by magnetic fields in cosmological phase transitions, Phys. Rev. D **81,** 023004 (2010).

[49] C. Caprini, R. Durrer, and G. Servant, The stochastic gravitational wave background from turbulence and magnetic fields generated by a first-order phase transition, J. Cosmol. Astropart. Phys. 12 (2009) 024.

[50] L. Kisslinger and T. Kahniashvili, Polarized gravitational waves from cosmological phase transitions, Phys. Rev. D **92,** 043006 (2015).

[51] A. Roper Pol, S. Mandal, A. Brandenburg, T. Kahniashvili, and A. Kosowsky, Numerical simulations of gravitational waves from early-universe turbulence, Phys. Rev. D **102,** 083512 (2020).

[52] G. Boileau, N. Christensen, C. Gowling, M. Hindmarsh, and R. Meyer, Prospects for LISA to detect a gravitational-wave background from first order phase transitions, J. Cosmol. Astropart. Phys. 02 (2023) 056.

[53] M. Sakellariadou, Cosmic strings and cosmic superstrings, Nucl. Phys. B, Proc. Suppl. **192–193,** 68 (2009).

[54] M. Sakellariadou, Cosmic strings, Lect. Notes Phys. **718,** 247 (2007).

[55] P. Auclair *et al.,* Probing the gravitational wave background from cosmic strings with LISA, J. Cosmol. Astropart. Phys. 04 (2020) 034.

[56] X. Siemens, V. Mandic, and J. Creighton, Gravitational Wave Stochastic Background from Cosmic (Super)Strings, Phys. Rev. Lett. **98,** 111101 (2007).

[57] W. Buchmuller, V. Domcke, and K. Schmitz, Stochastic gravitational-wave background from metastable cosmic strings, J. Cosmol. Astropart. Phys. 12 (2021) 006.

[58] B. P. Abbott *et al.* (LIGO and Virgo Collaborations), Constraints on cosmic strings using data from the first Advanced LIGO observing run, Phys. Rev. D **97,** 102002 (2018).

[59] A. C. Jenkins and M. Sakellariadou, Anisotropies in the stochastic gravitational-wave background: Formalism and the cosmic string case, Phys. Rev. D **98,** 063509 (2018).

[60] R. Abbott *et al.* (LIGO, Virgo, and KAGRA Collaborations), Constraints on Cosmic Strings using Data from the Third Advanced LIGO–Virgo Observing Run, Phys. Rev. Lett. **126,** 241102 (2021).

[61] G. S. F. Guedes, P. P. Avelino, and L. Sousa, Signature of inflation in the stochastic gravitational wave background generated by cosmic string networks, Phys. Rev. D **98,** 123505 (2018).

[62] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Stochastic and Resolvable Gravitational Waves from Ultralight Bosons, Phys. Rev. Lett. **119,** 131101 (2017).

[63] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Gravitational wave searches for ultralight bosons with LIGO and LISA, Phys. Rev. D **96,** 064050 (2017).

[64] L. Tsukada, T. Callister, A. Matas, and P. Meyers, First search for a stochastic gravitational-wave background from ultralight bosons, Phys. Rev. D **99,** 103015 (2019).

[65] L. Tsukada, R. Brito, W. E. East, and N. Siemonsen, Modeling and searching for a stochastic gravitational-wave background from ultralight vector bosons, Phys. Rev. D **103,** 083005 (2021).

[66] C. Yuan, Y. Jiang, and Q.-G. Huang, Constraints on an ultralight scalar boson from Advanced LIGO and Advanced Virgo's first three observing runs using the stochastic gravitational-wave background, Phys. Rev. D **106,** 023020 (2022).

[67] V. Mandic, S. Bird, and I. Cholis, Stochastic Gravitational-Wave Background Due to Primordial Binary Black Hole Mergers, Phys. Rev. Lett. **117,** 201102 (2016).

[68] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914, Phys. Rev. Lett. **117,** 061101 (2016); **121,** 059901(E) (2018).

[69] S. Wang, Y.-F. Wang, Q.-G. Huang, and T. G. F. Li, Constraints on the Primordial Black Hole Abundance from the First Advanced LIGO Observation Run using the Stochastic Gravitational-Wave Background, Phys. Rev. Lett. **120,** 191102 (2018).

[70] A. L. Miller, S. Clesse, F. De Lillo, G. Bruno, A. Depasse, and A. Tanasijczuk, Probing planetary-mass primordial black holes with continuous gravitational waves, Phys. Dark Universe **32,** 100836 (2021).

[71] G. Cusin, C. Pitrou, and J.-P. Uzan, Anisotropy of the astrophysical gravitational wave background: Analytic expression of the angular power spectrum and correlation with cosmological observations, Phys. Rev. D **96,** 103019 (2017).

[72] M. Geller, A. Hook, R. Sundrum, and Y. Tsai, Primordial Anisotropies in the Gravitational Wave Background from Cosmological Phase Transitions, Phys. Rev. Lett. **121,** 201303 (2018).

[73] A. C. Jenkins, M. Sakellariadou, T. Regimbau, and E. Slezak, Anisotropies in the astrophysical gravitational-wave background: Predictions for the detection of compact binaries by LIGO and Virgo, Phys. Rev. D **98,** 063501 (2018).

[74] A. C. Jenkins, R. O'Shaughnessy, M. Sakellariadou, and D. Wysocki, Anisotropies in the Astrophysical Gravitational-Wave Background: The Impact of Black Hole Distributions, Phys. Rev. Lett. **122,** 111101 (2019).

[75] N. Bartolo, D. Bertacca, S. Matarrese, M. Peloso, A. Ricciardone, A. Riotto, and G. Tasinato, Anisotropies and non-Gaussianity of the cosmological gravitational wave background, Phys. Rev. D **100,** 121501 (2019).

[76] G. Cusin, I. Dvorkin, C. Pitrou, and J.-P. Uzan, First Predictions of the Angular Power Spectrum of the Astrophysical Gravitational Wave Background, Phys. Rev. Lett. **120,** 231101 (2018).

[77] E. Dimastrogiovanni, M. Fasiello, and G. Tasinato, Searching for Fossil Fields in the Gravity Sector, Phys. Rev. Lett. **124,** 061302 (2020).

[78] E. Dimastrogiovanni, M. Fasiello, A. Malhotra, and G. Tasinato, Enhancing gravitational wave anisotropies with peaked scalar sources, J. Cosmol. Astropart. Phys. 01 (2023) 018.

[79] Y. Zheng, N. Kouvatsos, J. Golomb, M. Cavaglià, A. I. Renzini, and M. Sakellariadou, The angular power spectrum of gravitational-wave transient sources as a probe of the large-scale structure, arXiv:2305.02652.

[80] S. W. Ballmer, A radiometer for stochastic gravitational waves, Classical Quantum Gravity **23,** S179 (2006).

[81] E. Thrane, S. Ballmer, J. D. Romano, S. Mitra, D. Talukder, S. Bose, and V. Mandic, Probing the anisotropies of a stochastic gravitational-wave background using a network of ground-based laser interferometers, Phys. Rev. D **80,** 122002 (2009).

[82] R. Abbott et al. (LIGO, Virgo, and KAGRA Collaborations), Search for anisotropic gravitational-wave backgrounds using data from Advanced LIGO and Advanced Virgo's first three observing runs, Phys. Rev. D **104,** 022005 (2021).

[83] L. Tsukada, S. Jaraba, D. Agarwal, and E. Floden, Bayesian parameter estimation for targeted anisotropic gravitational-wave background, Phys. Rev. D **107,** 023024 (2023).

[84] S. Banagiri, A. Criswell, T. Kuan, V. Mandic, J. D. Romano, and S. R. Taylor, Mapping the gravitational-wave sky with LISA: A Bayesian spherical harmonic approach, Mon. Not. R. Astron. Soc. **507,** 5451 (2021).

[85] J. Abadie et al. (LIGO Collaboration), Directional Limits on Persistent Gravitational Waves using LIGO S5 Science Data, Phys. Rev. Lett. **107,** 271102 (2011).

[86] B. P. Abbott et al. (LIGO and Virgo Collaborations), Directional Limits on Persistent Gravitational Waves from Advanced LIGO's First Observing Run, Phys. Rev. Lett. **118,** 121102 (2017).

[87] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, Pub. Astron. Soc. Aust. **36,** e010 (2019).

[88] R. Abbott et al. (LIGO, Virgo, and KAGRA Collaborations), Upper limits on the isotropic gravitational-wave background from Advanced LIGO and Advanced Virgo's third observing run, Phys. Rev. D **104,** 022004 (2021).

[89] P. A. R. Ade et al. (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. **594,** A13 (2016).

[90] J. D. Romano and N. J. Cornish, Detection methods for stochastic gravitational-wave backgrounds: A unified treatment, Living Rev. Relativity **20,** 2 (2017).

[91] B. Allen and A. C. Ottewill, Detection of anisotropies in the gravitational wave stochastic background, Phys. Rev. D **56,** 545 (1997).

[92] B. Allen and J. D. Romano, Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities, Phys. Rev. D **59,** 102001 (1999).

[93] M. Maggiore, Gravitational Waves: Volume 1: Theory and Experiments, Gravitational Waves (OUP, Oxford, 2008).

[94] B. P. Abbott et al. (LIGO and Virgo Collaborations), Directional limits on persistent gravitational waves using data from Advanced LIGO's first two observing runs, Phys. Rev. D **100,** 062001 (2019).

[95] C. Biwer et al. (LIGO Collaboration), Validating gravitational-wave detections: The Advanced LIGO hardware injection system, https://dcc.ligo.org/LIGO-P1600285/public (2016).

[96] A. Lazzarini, S. Bose, P. Fritschel, M. McHugh, T. Regimbau, K. Reilly, J. D. Romano, J. T. Whelan, S. E. Whitcomb, and B. F. Whiting, Optimal combination of signals from co-located gravitational wave interferometers for use in searches for a stochastic background, Phys. Rev. D **70,** 062001 (2004).

[97] A. Lazzarini and J. Romano (LIGO Collaboration), Use of overlapping windows in the stochastic background search, https://dcc.ligo.org/T040089/public (2004).

[98] A. Lazzarini (LIGO Collaboration), Bias from power spectrum measurement in parameter estimation for the stochastic gravitational wave background, https://dcc.ligo .org/LIGO-T040128/public (2004).

[99] A. K.-W. Chung, A. C. Jenkins, J. D. Romano, and M. Sakellariadou, Targeted search for the kinematic dipole of the gravitational-wave background, Phys. Rev. D **106**, 082005 (2022).

[100] A. Matas and J. D. Romano, Frequentist versus Bayesian analyses: Cross-correlation as an approximate sufficient statistic for LIGO-Virgo stochastic background searches, Phys. Rev. D **103**, 062003 (2021).

[101] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory* (Princeton University Press, Princeton, NJ, 2005).

[102] S. Mitra, S. Dhurandhar, T. Souradeep, A. Lazzarini, V. Mandic, S. Bose, and S. Ballmer, Gravitational wave radiometry: Mapping a stochastic gravitational wave background, Phys. Rev. D **77**, 042002 (2008).

[103] S. Biscoveanu, C. Talbot, E. Thrane, and R. Smith, Measuring the Primordial Gravitational-Wave Background in the Presence of Astrophysical Foregrounds, Phys. Rev. Lett. **125**, 241101 (2020).

[104] See the Supplemental Material at http://link.aps.org/ supplemental/10.1103/PhysRevD.108.043032 for the mean and sigma of the real and imaginary parts of different $\mathcal{P}_{\ell m}$ for $\ell \leq 10$ inferred from different simulated data.

[105] E. Floden, V. Mandic, A. Matas, and L. Tsukada, Angular resolution of the search for anisotropic stochastic gravitational-wave background with terrestrial gravitational-wave detectors, Phys. Rev. D **106**, 023010 (2022).

[106] S. Panda, S. Bhagwat, J. Suresh, and S. Mitra, Stochastic gravitational wave background mapmaking using regularized deconvolution, Phys. Rev. D **100**, 043541 (2019).

[107] J. Suresh, A. Ain, and S. Mitra, Unified mapmaking for an anisotropic stochastic gravitational wave background, Phys. Rev. D **103**, 083024 (2021).

[108] D. Agarwal, J. Suresh, S. Mitra, and A. Ain, Upper limits on persistent gravitational waves using folded data and the full covariance matrix from Advanced LIGO's first two observing runs, Phys. Rev. D **104**, 123018 (2021).

[109] E. Floden, V. Mandic, A. Matas, and L. Tsukada, Angular resolution of the search for anisotropic stochastic gravitational-wave background with terrestrial gravitational-wave detectors, Phys. Rev. D **106**, 023010 (2022).

[110] D. Agarwal, J. Suresh, S. Mitra, and A. Ain, Angular power spectra of anisotropic stochastic gravitational wave background: Developing statistical methods and analyzing data from ground-based detectors, Phys. Rev. D **108**, 023011 (2023).

[111] N. Bartolo *et al.* (LISA Cosmology Working Group), Probing anisotropies of the stochastic gravitational wave background with LISA, J. Cosmol. Astropart. Phys. 11 (2022) 009.

[112] D. Chowdhury, G. Tasinato, and I. Zavala, On the response of the Einstein Telescope to Doppler anisotropies, Phys. Rev. D **107**, 083516 (2023).

[113] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, HEALPY: Equal area pixelization and spherical harmonics transforms for data on the sphere in PYTHON, J. Open Source Software **4**, 1298 (2019).

[114] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPIX: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, Astrophys. J. **622**, 759 (2005).

[115] E. D. Hall and M. Evans, Metrics for next-generation gravitational-wave detectors, Classical Quantum Gravity **36**, 225002 (2019).

[116] A. Ain, J. Suresh, and S. Mitra, Very fast stochastic gravitational wave background map making using folded data, Phys. Rev. D **98**, 024001 (2018).

[117] A. Romero, K. Martinovic, T. A. Callister, H.-K. Guo, M. Martínez, M. Sakellariadou, F.-W. Yang, and Y. Zhao, Implications for First-Order Cosmological Phase Transitions from the Third LIGO-Virgo Observing Run, Phys. Rev. Lett. **126**, 151301 (2021).

[118] D. Reitze *et al.*, Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, Bull. Am. Astron. Soc. **51**, 035 (2019), https://baas.aas.org/pub/ 2020n7i035/release/1.

[119] M. Punturo *et al.*, The Einstein Telescope: A third-generation gravitational wave observatory, Classical Quantum Gravity **27**, 194002 (2010).

[120] B. Abbott *et al.* (LIGO Collaboration), Analysis of first LIGO science data for stochastic gravitational waves, Phys. Rev. D **69**, 122004 (2004).

[121] R. Abbott *et al.* (LIGO Collaboration), Data for 'Search for the isotropic stochastic background using data from Advanced LIGO's second observing run', https://dcc .ligo.org/T1900058/public (2019).

[122] R. Abbott *et al.* (LIGO, Virgo, and KAGRA Collaborations), Data products and supplemental information for O3 stochastic directional paper, https://dcc.ligo.org/LIGO-G2002165/public (2021).

[123] B. P. Abbott *et al.* (LIGO and Virgo Collaborations), Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914, Classical Quantum Gravity **33**, 134001 (2016).

[124] E. Thrane, N. Christensen, and R. Schofield, Correlated magnetic noise in global networks of gravitational-wave interferometers: Observations and implications, Phys. Rev. D **87**, 123009 (2013).

[125] E. Thrane, N. Christensen, R. M. S. Schofield, and A. Effler, Correlated noise in networks of gravitational-wave detectors: Subtraction and mitigation, Phys. Rev. D **90**, 023013 (2014).

[126] R. Smith and E. Thrane, Optimal Search for an Astrophysical Gravitational-Wave Background, Phys. Rev. X **8**, 021019 (2018).

[127] S. Biscoveanu, C. Talbot, E. Thrane, and R. Smith, Measuring the Primordial Gravitational-Wave Background in the Presence of Astrophysical Foregrounds, Phys. Rev. Lett. **125**, 241101 (2020).