



Forecasting Pulsar Timing Array Sensitivity to Anisotropy in the Stochastic Gravitational Wave Background

Nihan Pol¹ , Stephen R. Taylor¹ , and Joseph D. Romano² ¹ Department of Physics and Astronomy, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235, USA² Department of Physics and Astronomy, Texas Tech University, Lubbock, TX 79409-1051, USA

Received 2022 June 15; revised 2022 September 16; accepted 2022 October 5; published 2022 December 2

Abstract

Statistical anisotropy in the nanohertz-frequency gravitational wave background (GWB) is expected to be detected by pulsar timing arrays (PTAs) in the near future. By developing a frequentist statistical framework that intrinsically restricts the GWB power to be positive, we establish scaling relations for multipole-dependent anisotropy decision thresholds that are a function of the noise properties, timing baselines, and cadences of the pulsars in a PTA. We verify that (i) a larger number of pulsars, and (ii) factors that lead to lower uncertainty on spatial cross-correlation measurements between pulsars, lead to a higher overall GWB signal-to-noise ratio, and lower anisotropy decision thresholds with which to reject the null hypothesis of isotropy. Using conservative simulations of realistic NANOGrav data sets, we predict that an anisotropic GWB with angular power $C_{l=1} > 0.3C_{l=0}$ may be sufficient to produce tension with isotropy at the $p = 3 \times 10^{-3}$ ($\sim 3\sigma$) level in near-future NANOGrav data with a 20 yr baseline. We present ready-to-use scaling relationships that can map these thresholds to any number of pulsars, configuration of pulsar noise properties, or sky coverage. We discuss how PTAs can improve the detection prospects for anisotropy, as well as how our methods can be adapted for more versatile searches.

Unified Astronomy Thesaurus concepts: Millisecond pulsars (1062); Gravitational wave astronomy (675); Gravitational waves (678); Radio pulsars (1353)

1. Introduction

All of the long-baseline pulsar timing arrays (PTAs)—namely the North American Nanohertz Observatory for Gravitational Waves (NANOGrav; McLaughlin 2013), the European Pulsar Timing Array (EPTA; Kramer & Champion 2013), and the Parkes Pulsar Timing Array (PPTA; Hobbs 2013)—have now reported evidence for the presence of a spectrally common process in their latest data sets (Arzoumanian et al. 2020; Chen et al. 2021; Goncharov et al. 2021). However, the evidence for Hellings and Downs (HD) spatial cross-correlations (Hellings & Downs 1983), which are considered the definitive signature of a stochastic gravitational wave background (GWB), is not significant in any of these data sets. A spectrally common process was also reported by the International Pulsar Timing Array (IPTA; Hobbs et al. 2010) consortium in their second data release (Perera et al. 2019), which used older versions of data sets from the aforementioned PTAs. However this data set also lacked definitive evidence for the HD signature (Antoniadis et al. 2022). The detection of such a spectrally common process could be the first step toward the eventual detection of a GWB (Romano et al. 2021). Based on simulations of the NANOGrav 12.5 yr data set (Alam et al. 2021), and if the signal observed in the 12.5 yr data set is an astrophysical GWB signal, NANOGrav is expected to have sufficient evidence to report the detection of HD-consistent spatial correlations within the next few years (Pol et al. 2021). Throughout the rest of this article, unless explicitly stated

otherwise, we use the terms “correlations” and “cross-correlations” to refer to spatial cross-correlations.

If this signal is confirmed to be a GWB signal in future PTA data sets, the onus will be on characterizing the source of the GWB. An important part of this analysis will be measuring the spatial distribution of power in the GWB. For example, if the source of the GWB is a population of inspiraling supermassive black hole binaries (SMBHBs; Rajagopal & Romani 1995; Jaffe & Backer 2003; Sesana et al. 2004; Burke-Spolaor et al. 2019), then their spatial distribution might track the local matter distribution. In particular, nearby galaxy clusters that host an overabundance of SMBHBs may show up as hotspots of gravitational wave (GW) emission on the sky. For example, the Virgo cluster (Sandage 1958) has an angular diameter of $\sim 10^\circ$, and could show up as a hotspot on the GW sky at a multipole of $l_{\text{Virgo}} = 180^\circ/\theta \approx 18$. On the other hand, a single SMBHB that is louder than the GWB will show up as a point-source anisotropy, with multiple such single sources producing a pixel-level spatial distribution of GWB power. However, if the GWB is produced by a cosmological source, such as cosmic strings (e.g., Olmez et al. 2010) or primordial GWs (e.g., Grishchuk 2005), then the GWB power distribution on the sky may not display the same anisotropies as that from an SMBHB-produced GWB. Thus, the anisotropy of the GWB, or the lack of it, allows us to make inferences about the source of the GWB.

Multiple techniques have been developed to probe the anisotropy of a GWB with PTAs (e.g., Mingarelli et al. 2013; Taylor & Gair 2013; Cornish & van Haasteren 2014; Taylor et al. 2020). While these methods differ in their choice of basis for modeling anisotropy, they all take the pulsar times of arrival (TOAs) as their initial data, and employ Bayesian techniques to constrain parameters that describe anisotropy-induced deviations from the HD signature. Constraints on anisotropy were



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

first presented by the EPTA as part of the analysis of their first data release (Taylor et al. 2015), where they placed limits on the strain amplitude in $l > 0$ spherical harmonic multipoles of less than 40% of the monopole value ($l = 0$, i.e., isotropy).

As PTA data sets grow longer in timespan, add more pulsars, and become denser with higher-cadence observations, the analysis time for Bayesian methods based on TOAs is going to increase dramatically. Additionally, Bayesian model selection of anisotropy is predicated on having an appropriate hypothesis of the anisotropy, for example, a power map built from galaxy catalogs, statistically populated with inspiraling SMBHBs. In other words, Bayesian model selection always requires two models to compare: a null and a signal model. By contrast, frequentist techniques allow one to reject a null hypothesis (in this case, isotropy) if the data are in sufficient tension with it. Importantly, rejecting a null hypothesis at a probability given by some p -value is not the same as accepting a signal hypothesis with a certainty of $1 - p$. However significant tension with the assumption of isotropy is an important indicator of beyond-HD signatures in the cross-correlation data.

To overcome these challenges, we develop a frequentist framework that employs the cross-correlations between pulsar timing residuals, defined as the difference between the observed TOAs and those predicted by the timing model, across a PTA as sufficient data with which to search for anisotropy. These cross-correlations are measured as part of the standard GWB detection pipelines (e.g., Arzoumanian et al. 2020; Chen et al. 2021; Goncharov et al. 2021), using established optimal two-point correlation techniques (Allen & Romano 1999; Anholm et al. 2009; Demorest et al. 2013; Vigeland et al. 2018). Thus our framework can be easily incorporated into ongoing analysis campaigns. The data volume in our framework is also significantly lower than that in analyses starting at the TOA data level, enabling rapid estimation of anisotropy. Finally, as mentioned above, the frequentist framework allows us to infer (although not necessarily claim detection of) anisotropy via rejection of the null hypothesis of isotropy, thereby simplifying the process of constraining anisotropy.

Other frequentist techniques for searching for anisotropy with PTAs have been proposed. These include the Fisher matrix formalism developed in Ali-Haïmoud (2020, 2021), where “principal maps,” the eigenmaps of the Fisher matrix, were employed to search for anisotropy. Hotinli et al. (2019), on the other hand, decomposed the timing residual power spectrum into bipolar spherical harmonics to search for anisotropy using the correlations between pulsars in a PTA. These frequentist techniques also focus on the detection of anisotropy through rejection of the null hypothesis of isotropy. However, these methods are either based on TOAs as their data set, making them susceptible to the problem of increasing data volume described earlier, or they require a framework different from the current detection pipelines. The use of cross-correlations between detector baselines is also common when searching for anisotropic GWBs in the LIGO (e.g., Thrane et al. 2009; Abbott et al. 2021) and LISA bands (e.g., Banagiri et al. 2021).

We use our framework on simulations of idealized PTA and near-future NANOGrav data, presenting, for the first time, projections on the sensitivity of NANOGrav and other PTAs to GWB anisotropy in the mid-to-late 2020s. The paper is organized as follows: In Section 2, we describe the measurement of cross-correlations and their uncertainties, along with

the maximum likelihood framework and detection statistics that are used to constrain anisotropy, while Section 3 connects the cross-correlation uncertainty to the noise properties, cadence, and timing baseline of the PTA. Section 4 presents scaling relations for anisotropy decision thresholds of an “ideal PTA” as a function of cross-correlation uncertainty and number of pulsars in the PTA, while Section 5 presents them for a realistic PTA that is generated using the NANOGrav 12.5 yr data set. Finally, we present a discussion of results and prospects for the future in Section 6. In Appendices A and B we present derivations for computing the cross-correlations in real PTA data sets and scaling relations for the cross-correlation uncertainty with respect to PTA specifications like timing baseline, white noise, and observation cadence. The software associated with the methods developed in this work is available as a Python package on GitHub.³

2. Methods

2.1. The Optimal Cross-correlation Statistic and Overlap Reduction Function

A GWB can be uniquely identified through the cross-correlations it induces between the TOAs of pulsars in a PTA (Hellings & Downs 1983; Tiburzi et al. 2016). The spatial cross-correlations between pulsar pairs, ρ_{ab} , and their uncertainties, σ_{ab} , can be written as (Demorest et al. 2013; Siemens et al. 2013; Chamberlin et al. 2015; Vigeland et al. 2018)

$$\rho_{ab} = \frac{\delta \mathbf{t}_a^T \mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \delta \mathbf{t}_b}{\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}]},$$

$$\sigma_{ab} = (\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}])^{-1/2}, \quad (1)$$

where $\delta \mathbf{t}_a$ is a vector of timing residuals for pulsar a , $\mathbf{P}_a = \langle \delta \mathbf{t}_a \delta \mathbf{t}_a^T \rangle$ is the measured autocovariance matrix of pulsar a , and $\hat{\mathbf{S}}_{ab}$ is the template-scaled covariance matrix between pulsar a and pulsar b . This scaled covariance matrix is a template for the GWB spectral shape only, and is independent of the GWB’s amplitude and the cross-correlation signature that it induces. It is related to the full covariance matrix by $\mathbf{S}_{ab} = \langle \delta \mathbf{t}_a \delta \mathbf{t}_b^T \rangle = A_{\text{gwb}}^2 \chi_{ab} \hat{\mathbf{S}}_{ab}$, where A_{gwb} is the GWB amplitude corresponding to a given strain spectrum template, and χ_{ab} is the GWB-induced cross-correlation value for this pair of pulsars, e.g., the HD factor in the case of an isotropic GWB.

The cross-correlation statistic accounts for fitting and marginalization over the timing ephemeris of each pulsar, as well as its intrinsic white and red noise⁴ characteristics, where a power-law spectrum is usually assumed for the intrinsic red noise. Implementations of the statistic also usually assume a power-law strain spectrum template for the GWB following $f^{-2/3}$ (Phinney 2001) filtering it across the pairwise-correlated timing residuals in order to extract an optimal measurement of the GWB amplitude (Chamberlin et al. 2015; Vigeland et al. 2018). However, we note that this statistic is flexible enough to allow for different parameterized spectral templates. See Appendix A for details on how this is implemented in real PTA searches for the GWB, including how the noise weighting for each pulsar is determined, how the GWB’s spectral

³ <https://github.com/NihanPol/MAPS>

⁴ Here, white and red noise refer to power spectral densities (PSDs) with approximately equal power at all GW frequencies and higher power at lower GW frequencies, respectively.

template is set, and how the parameters of the pulsar timing ephemeris are marginalized over. In this paper we simulate and analyze data at the level of $\{\rho_{ab}, \sigma_{ab}\}$, then connect our results later to the underlying geometry and noise characteristics of the PTA (see Appendix B).

A PTA with N_{psr} pulsars has $N_{\text{cc}} = N_{\text{psr}}(N_{\text{psr}} - 1)/2$ distinct cross-correlation values. The angular dependence of these empirically measured cross-correlations can be modeled by the detector overlap reduction function (ORF)⁵ (Flanagan 1993; Mingarelli et al. 2013; Taylor & Gair 2013; Gair et al. 2014; Taylor et al. 2020) such that, for an unpolarized GWB,

$$\Gamma_{ab} \propto \int_{S^2} d^2\hat{\Omega} P(\hat{\Omega}) [\mathcal{F}^+(\hat{p}_a, \hat{\Omega}) \mathcal{F}^+(\hat{p}_b, \hat{\Omega}) + \mathcal{F}^\times(\hat{p}_a, \hat{\Omega}) \mathcal{F}^\times(\hat{p}_b, \hat{\Omega})], \quad (2)$$

where $P(\hat{\Omega})$ is the angular power of the GWB in direction $\hat{\Omega}$, normalized such that $\int_{S^2} d^2\hat{\Omega} P(\hat{\Omega}) = 1$, and $\mathcal{F}^A(\hat{p}, \hat{\Omega})$ is the antenna response pattern of a pulsar in unit vector direction \hat{p}_a to each GW polarization A

$$\mathcal{F}^A(\hat{p}, \hat{\Omega}) = \frac{1}{2} \frac{\hat{p}^i \hat{p}^j}{1 - \hat{\Omega} \cdot \hat{p}} e_{ij}^A(\hat{\Omega}), \quad (3)$$

where $e_{ij}^A(\hat{\Omega})$ denotes the polarization basis tensors, and (i, j) are the spatial indices.

We can recast the sky integral in Equation (2) as a sum over equal-area pixels (Gair et al. 2014; Taylor et al. 2020). Assuming an unpolarized GWB, and ignoring random pulsar term contributions to the cross-correlations, this can be written as

$$\Gamma_{ab} \propto \sum_k P_k [\mathcal{F}_{a,k}^+ \mathcal{F}_{b,k}^+ + \mathcal{F}_{a,k}^\times \mathcal{F}_{b,k}^\times], \quad (4)$$

where k denotes the pixel indices. Or, in a general matrix form

$$\mathbf{\Gamma} = \mathbf{R}\mathbf{P}, \quad (5)$$

where $\mathbf{\Gamma}$ is an N_{cc} vector of ORF values for all distinct pulsar pairs, \mathbf{P} is an N_{pix} vector of GWB power values at different pixel locations, and \mathbf{R} is an $(N_{\text{cc}} \times N_{\text{pix}})$ overlap response matrix given by

$$R_{ab,k} = \frac{3}{2N_{\text{pix}}} [\mathcal{F}_{a,k}^+ \mathcal{F}_{b,k}^+ + \mathcal{F}_{a,k}^\times \mathcal{F}_{b,k}^\times], \quad (6)$$

where the normalization is chosen so that the ORF matches the HD values in the case of an isotropic GWB with $P_k = 1 \forall k$.

Combining the notation introduced in Equations (1) and (2), the expected value of ρ_{ab} is such that $\langle \rho_{ab} \rangle = A^2 \Gamma_{ab}$. However, in the remainder of this paper we will deal with amplitude-scaled cross-correlation values, ρ_{ab}/A^2 , where we assume that in a real search an initial fit for A^2 will be performed on $\{\rho_{ab}\}$ under the assumption of isotropy. These amplitude-scaled cross-correlation values can then be directly compared with the ORF model, $\mathbf{\Gamma}$, to probe anisotropy. We suppress the explicit amplitude scaling in the remainder of our notation, such that ρ_{ab} henceforth implies amplitude-scaled cross-correlation values.

⁵ We note that the term ORF in ground- and space-based literature usually involves a frequency dependence. However this frequency dependence factors out in the PTA regime, such that we use the term ORF to denote only the angular dependence of the pairwise cross-correlated data (Romano & Cornish 2017).

Using the cross-correlations as our data, and assuming a stationary Gaussian distribution for the cross-correlation uncertainty, the likelihood function for the cross-correlations can be written as

$$p(\rho|\mathbf{P}) = \frac{\exp\left[-\frac{1}{2}(\rho - \mathbf{R}\mathbf{P})^T \mathbf{\Sigma}^{-1}(\rho - \mathbf{R}\mathbf{P})\right]}{\sqrt{\det(2\pi\mathbf{\Sigma})}}, \quad (7)$$

where $\mathbf{\Sigma}$ is the diagonal covariance matrix of cross-correlation uncertainties with shape $N_{\text{cc}} \times N_{\text{cc}}$. We now discuss several different bases on which to decompose the angular power vector, \mathbf{P} .

2.1.1. Pixel Basis

The GWB power can be parameterized using HEALPix sky pixelization (Gorski et al. 2005), where each equal-area pixel is independent of the pixels surrounding it:

$$P(\hat{\Omega}) = \sum_{\hat{\Omega}'} P_{\hat{\Omega}'} \delta^2(\hat{\Omega}, \hat{\Omega}'). \quad (8)$$

The number of pixels is set by $N_{\text{pix}} = 12N_{\text{side}}^2$ and N_{side} defines the tessellation of the healpix sky (Gorski et al. 2005). The rule of thumb for PTAs is to have $N_{\text{pix}} \leq N_{\text{cc}}$ (Cornish & van Haasteren 2014; Romano & Cornish 2017). This basis is well suited for detection of pixel-scale anisotropy, which can arise from individual sources of GWs, and where an isotropic GWB would be represented by equal power (within uncertainties) in each pixel on the sky.

2.1.2. Spherical Harmonic Basis

Alternatively, the GWB power can be decomposed into the spherical harmonic basis (e.g., Thrane et al. 2009), where the lowest-order multipole ($l = 0$) defines an isotropic background, while higher multipoles add anisotropy. The GWB power in this basis can be written as

$$P(\hat{\Omega}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{lm} Y_{lm}(\hat{\Omega}), \quad (9)$$

where Y_{lm} denotes the real-valued spherical harmonics and c_{lm} denotes the spherical harmonic coefficients of $P(\hat{\Omega})$. In this basis, the ORF anisotropy coefficient for the l and m components between pulsars a and b can be written as (Taylor et al. 2020)

$$\Gamma_{(lm)(ab)} = \kappa \sum_k c_{lm} Y_{lm,k} [\mathcal{F}_{a,k}^+ \mathcal{F}_{b,k}^+ + \mathcal{F}_{a,k}^\times \mathcal{F}_{b,k}^\times], \quad (10)$$

where k represents the pixel index corresponding to $\hat{\Omega}$ and the constant κ accounts for the pixel area in the healpix sky tessellation.

This basis representation, in contrast to the pixel basis, is better suited for modeling large-scale anisotropies in the GWB. Based on diffraction limit arguments, the highest-order mode, l_{max} , that can be used for modeling the anisotropy depends on the number of pulsars in the PTA, $l_{\text{max}} \sim \sqrt{N_{\text{psr}}}$ (Boyle & Pen 2012; Romano & Cornish 2017). However, Floden et al. (2022) have shown that while the diffraction limit is attuned to maximizing the significance of the detection of anisotropy, values of $l > l_{\text{max}}$ can be included in spherical harmonic

decompositions to improve the localization of any anisotropy after its detection.

The results can be expressed in terms of C_l , which is the squared angular power in each mode l :

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^l |c_{lm}|^2. \quad (11)$$

Physically, C_l represents the amplitude of statistical fluctuations in the angular power of the GWB at scales corresponding to $\theta = 180^\circ/l$. An isotropic background in this basis will contain power only in the $l=0$ multipole, thus filling the entire sky, while an anisotropic background will have power in the higher- l multipoles. On the other hand, the variance of the angular power distribution can be written as (Jenkins 2022)

$$\text{Var}[P(\hat{\Omega})] \approx \int d(\ln l) \frac{l(l+1)}{2\pi} C_l. \quad (12)$$

The quantity $l(l+1)C_l/2\pi$ thus represents the variance per logarithmic multipole bin, and is what we use to present our results in this work. As pointed out in Jenkins (2022), reporting C_l is analogous to reporting the GWB strain PSD, with $C_l = \text{constant}$ representing a white angular power spectrum, while reporting $l(l+1)C_l/2\pi$ is analogous to reporting the GWB energy density spectrum $\Omega_{\text{gw}}(f)$, with $l(l+1)C_l/2\pi = \text{constant}$ representing a scale-invariant angular power spectrum.

For these bases, since the problem is linear in the regression coefficients, the maximum likelihood solution can be derived analytically (Thrane et al. 2009; Romano & Cornish 2017; Ivezić 2019)

$$\hat{\mathbf{P}} = \mathbf{M}^{-1}\mathbf{X}, \quad (13)$$

where $\mathbf{M} \equiv \mathbf{\Sigma}^{-1}$ is the Fisher information matrix, with the uncertainties on the c_{lm} coefficients given by the diagonal elements of \mathbf{M}^{-1} , and $\mathbf{X} \equiv \mathbf{R}^T \mathbf{\Sigma}^{-1} \mathbf{\rho}$ is the “dirty map,” an inverse-noise-weighted representation of the total power on the sky as “seen” through the response of the pulsars in the array.

2.1.3. Square-root Spherical Harmonic Basis

A drawback of both the pixel and spherical harmonic bases is that they allow the GWB power to assume negative values, which is an unphysical realization of the GWB. While these tendencies can be curbed through the use of regularization techniques or rejection priors (Taylor & Gair 2013), this results in the addition of a hyperparameter to the analysis that requires further optimization or nonanalytic priors (Taylor & Gair 2013; Taylor et al. 2020). A more elegant solution is to use a basis that intrinsically conditions the GWB power to be positive over the whole sky.

Such a basis can be generated by modeling the square root of the GWB power, $P(\hat{\Omega})^{1/2}$, rather than modeling the power itself. This technique was introduced in a Bayesian context in Payne et al. (2020) for LIGO, in Banagiri et al. (2021) for LISA, and in Taylor et al. (2020) for PTAs. Decomposing the square-root power into spherical harmonics, the GWB power can be written as

$$P(\hat{\Omega}) = [P(\hat{\Omega})^{1/2}]^2 = \left[\sum_{L=0}^{\infty} \sum_{M=-L}^L b_{LM} Y_{LM} \right]^2, \quad (14)$$

where Y_{LM} denotes the real-valued spherical harmonics and b_{LM} denotes the search coefficients. Banagiri et al. (2021) showed that the search coefficients in this basis can be related to the spherical harmonic coefficients via

$$c_{lm} = \sum_{LM} \sum_{L'M'} b_{LM} b_{L'M'} \beta_{lm}^{LM, L'M'}, \quad (15)$$

where $\beta_{lm}^{LM, L'M'}$ is defined as

$$\beta_{lm}^{LM, L'M'} = \sqrt{\frac{(2L+1)(2L'+1)}{4\pi(2l+1)}} C_{LM, L'M'}^{lm} C_{L0, L'0}^{l0}, \quad (16)$$

with $C_{LM, L'M'}^{lm}$ being the Clebsch–Gordan coefficients. Taylor et al. (2020) showed that even though a full reconstruction of c_{lm} requires an infinite sum over the b_{LM} coefficients, restricting the maximum mode to $L \leq L_{\text{max}} = l_{\text{max}}$ is sufficient to produce an accurate reconstruction of the GWB power.

Since the problem in this basis is nonlinear in the regression coefficients, the likelihood in Equation (7) cannot be maximized analytically. The maximum likelihood solution thus has to be calculated through numerical optimization techniques. In this work, we use the LMFIT (Newville et al. 2021) Python package, where we use the Levenberg–Marquardt optimization algorithm (Levenberg 1944; Marquardt 1963) to determine the maximum likelihood solution. The goodness of fit is assessed through the χ^2_{dof} that is reported by LMFIT.

Figure 1 shows an example of recovering an anisotropic background using this basis. To produce the simulated cross-correlation data, we inject a GW power map corresponding to the synthesized population of SMBHBs from Taylor et al. (2020) into an “ideal PTA” consisting of 100 pulsars, with a constant cross-correlation measurement uncertainty of 0.01. Note that this is an uncertainty on cross-correlations that can assume any value between -0.2 and 0.5 , and thus represents an extremely accurate measurement of the cross-correlations between pulsars in the PTA (see Section 4.1 for our definition of an “ideal PTA”). We see that this basis is capable of reproducing the injected angular power spectrum, which translates into an accurate identification of the anisotropy in the GWB.

2.2. Detection Statistics

In addition to estimating the values of the anisotropy search coefficients (i.e., the pixels or spherical harmonic coefficients), we also need to quantify the evidence for the presence of anisotropy in the cross-correlation data. As described earlier, when searching for anisotropy, we seek to reject the null hypothesis of isotropy. In this section, we present two frequentist detection statistics that can quantify the evidence for anisotropy by measuring the (in)compatibility of the data with the null hypothesis.

2.2.1. Signal-to-noise Ratio

Confidence in the detection of anisotropy can be quantified through the signal-to-noise ratio (S/N), which is defined as the ratio of the maxima of the likelihood functions between any two models:

$$\text{S/N} = \sqrt{2 \ln[\Lambda_{\text{ML}}]}, \quad (17)$$

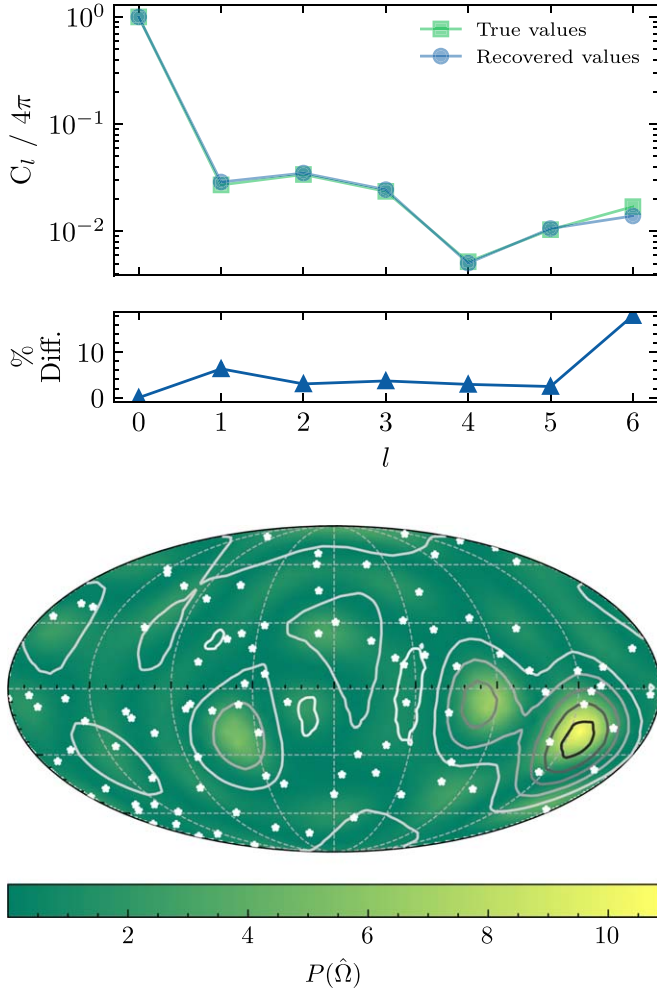


Figure 1. Example recovery of an anisotropic GWB using the square-root spherical harmonic basis described in Section 2. These simulations are based on an ideal PTA consisting of 100 pulsars, with a cross-correlation uncertainty of 0.01 across all pulsar pairs, while the anisotropy is based on a realistic population of inspiraling SMBHBs (Taylor et al. 2020). Top: The true and recovered angular power spectra, as well as the percent difference between them, where both are normalized such that the power in the $l=0$ mode is $C_0 = 4\pi$. Bottom: The sky map of the GWB power corresponding to the recovered angular power spectrum. The contours represent the true distribution of the GWB power on the sky for the anisotropic GWB, while the stars represent the positions of the simulated pulsars on the sky.

where $\Lambda_{\text{ML}} = p(\rho|\mathbf{P}_{\text{ML},1})/p(\rho|\mathbf{P}_{\text{ML},2})$ is the ratio of the maxima of the likelihood functions for the two models that are under consideration.

When searching for anisotropy, we can define three S/N statistics that together provide a complete description of the evidence for a GWB signal being present in the cross-correlation data:

1. “Total S/N”: This is defined as the ratio of the maximum likelihood value of an anisotropic model (i.e., $p(\rho|\mathbf{P}_{\text{ML}}(l_{\text{max}} > 0))$) to that of a model with only spatially uncorrelated noise (i.e., $p(\rho|\mathbf{P}_{\text{ML}} = 0)$). The total S/N quantifies the evidence for the presence of any signal in the cross-correlations.
2. “Isotropic S/N”: This is defined as the ratio of the maximum likelihood value of an isotropic model (i.e., $p(\rho|\mathbf{P}_{\text{ML}}(l_{\text{max}} = 0))$) to that of a model with noise only (i.e., $p(\rho|\mathbf{P}_{\text{ML}} = 0)$). Note that the isotropic S/N is

equivalent to the optimal S/N statistic defined in Chamberlin et al. (2015). The isotropic S/N quantifies how well the cross-correlations are described by a purely isotropic model.

3. “Anisotropic S/N”: This is defined as the ratio of the maximum likelihood value of a model with anisotropy (i.e., $p(\rho|\mathbf{P}_{\text{ML}}(l_{\text{max}} > 0))$) to that of an isotropic model (i.e., $p(\rho|\mathbf{P}_{\text{ML}}(l_{\text{max}} = 0))$). The anisotropic S/N quantifies evidence in favor of inclusion of modes $l > 0$.

2.2.2. Decision Threshold

Another method for determining the significance of possible anisotropy is to assess the certainty with which the null hypothesis of isotropy can be rejected. If we can quantify the distribution of the angular power, C_l , under the null hypothesis, then we can also quantify how (in)consistent the measured angular power is with isotropy through a test statistic like the p -value.

To calculate the null distribution of C_l under the null hypothesis, we generate many realizations of cross-correlation data, where we assume that the measurements are Gaussian-distributed around the HD curve, with the spread of the distribution given by the uncertainty on the cross-correlation values. We define the decision threshold, C_l^{th} , as the value of C_l corresponding to a p -value of 3×10^{-3} , where a measurement of angular power greater than this threshold would indicate a tension with the null hypothesis at the $\sim 3\sigma$ level.

3. Cross-correlation Uncertainties from PTA Design

The uncertainty on the cross-correlation measurements introduced in Equation (1) depends on the pulsars that are used to construct the PTA (Anholm et al. 2009; Siemens et al. 2013; Chamberlin et al. 2015). As shown in Siemens et al. (2013), the trace in Equation (1) can be written as

$$\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}] = \frac{2T}{A^4} \int_{f_l}^{f_h} df \frac{P_g^2(f)}{P_a(f)P_b(f)}, \quad (18)$$

where T is the duration in which a pulsar is observed, i.e., the timing baseline, A is the amplitude of the GWB, $P_g(f)$ is the power spectrum of the timing residuals induced by the GWB, $P_a(f)$ and $P_b(f)$ are the intrinsic power spectra of pulsars a and b , and $f_l = 1/T$ and f_h are the low- and high-frequency cutoffs used in the GWB analysis.

Using the same analysis as presented in Siemens et al. (2013) and Chamberlin et al. (2015), we can show that in the weak-signal regime, where the intrinsic pulsar noise dominates the GWB power (see Appendix B), the cross-correlation uncertainty between a given pair of pulsars scales as

$$\sigma \propto \frac{w^2 T^{-\gamma}}{c}, \quad (19)$$

while in the strong-signal regime, where the GWB power dominates the intrinsic pulsar noise,

$$\sigma \propto \frac{A^2}{\sqrt{cT}}, \quad (20)$$

where w is the white noise rms, $c = 1/\Delta t$ is the observing cadence, $\gamma = 3 - 2\alpha$ (Arzoumanian et al. 2020) is the slope of the timing residual power spectrum induced by the GWB, and

A is the amplitude of the GWB, whose characteristic strain spectrum is given by $h_c(f) = A(f/f_{\text{yr}})^\alpha$, with $\alpha = -2/3$ for an SMBHB GWB.

These scaling relations imply that the uncertainty on the cross-correlations can be reduced by observing pulsars for longer duration (timing baselines), or with a higher cadence. The effect of the timing baseline and cadence is strongest in the weak-signal regime while it becomes weaker as we move into the strong-signal regime. Similarly, in the weak-signal regime, the cross-correlation uncertainty can be reduced by decreasing the intrinsic white noise by, for example, increasing the receiver bandwidth or increasing the integration time for the pulsars. The white noise does not affect the cross-correlation uncertainty in the strong-signal regime, where the GWB signal dominates over the intrinsic noise in the pulsars at all frequencies.

4. Simulations with an Ideal PTA

4.1. Defining an Ideal PTA

In the framework described above, the rejection of isotropy in the GWB depends on three variables: the uncertainty on the cross-correlation values, the number of pulsars in the PTA (which defines the number of cross-correlation values that are measured), and the distribution of pulsars on the sky. The first two variables primarily dictate the strength of the rejection of isotropy, while the third variable is important for the characterization of the anisotropy. In this section, we examine how the cross-correlation uncertainty and the number of pulsars in the PTA affect an *ideal* PTA, which we define as a PTA that has pulsars distributed uniformly on the sky and having identical noise properties. The latter constraint implies that all measured cross-correlations have the same uncertainty. This is different from current, real PTAs, where each pulsar is unique and thus the uncertainties on the cross-correlations between each pulsar pair are different. We examine realistic PTAs in Section 5.

4.2. Scaling Relations

For a given level of anisotropy, its detection significance will depend on the number of pulsars in the array, as well as on the accuracy with which the cross-correlations between different pulsar pairs can be measured. Romano & Cornish (2017) showed that the total S/N for linear anisotropy models can be written as

$$\text{total S/N} = (\hat{P}^T [\mathbf{R}^T \mathbf{\Sigma}^{-1} \mathbf{R}] \hat{P})^{1/2}, \quad (21)$$

where \hat{P} is the maximum likelihood estimate of the GWB power. Since $\mathbf{\Sigma}$ is a diagonal matrix of the squared cross-correlation uncertainties, the total S/N $\propto \sigma^{-1}$. Similarly, the total S/N is proportional to the square root of the number of data points available for inference (assuming all cross-correlation uncertainties are the same for all pairs), i.e.,

$$\text{total S/N} \propto \sqrt{N_{\text{cc}}} = \sqrt{\frac{N_{\text{psr}}(N_{\text{psr}} - 1)}{2}}, \quad (22)$$

where for sufficiently large N_{psr} , the total S/N will scale linearly with the number of pulsars in the PTA.

We show that both of these scaling relations for the total S/N are also satisfied when using the nonlinear maximum likelihood approach described in Section 2.1.3. Figure 2 and

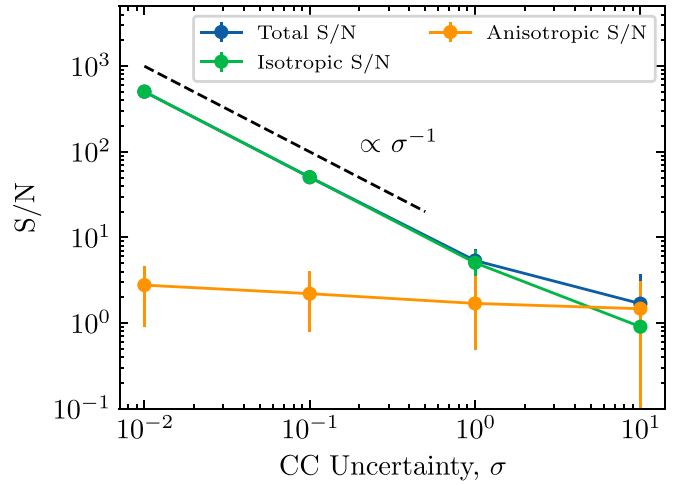


Figure 2. The evolution of the total, isotropic, and anisotropic S/N values over 10^4 noise realizations as a function of cross-correlation uncertainty for an ideal PTA with 100 pulsars and an isotropic GWB injection. The points represent the median, while the error bars represent the 95% confidence intervals on the distribution of the S/N values. A black dashed line corresponding to the scaling relation in Section 4.2 is shown for reference. For low cross-correlation uncertainties, the total and isotropic S/N values have identically large values relative to the anisotropic S/N, implying strong evidence for an isotropic GWB in the data. These S/N values decrease as the cross-correlation uncertainty increases, implying loss of confidence in the detection of a GWB as well as loss of ability to distinguish isotropy from anisotropy. The anisotropic S/N shows little evolution across uncertainties, though the better fit provided by the additional degrees of freedom in an anisotropic model prevents this S/N from being consistent with zero for small cross-correlation uncertainties.

the left panel in Figure 3 show that the total S/N scales inversely with the uncertainty on the cross-correlations, while the right-hand panel of Figure 3 shows that the total S/N scales proportionally to the number of pulsars in the PTA. Since the injected signal here is an isotropic GWB, Figure 2 shows that the total and isotropic S/N values are identically large, and decrease as the uncertainty on the cross-correlations increases. The anisotropic S/N shows little evolution across uncertainties, though the better fit provided by the additional degrees of freedom in an anisotropic model prevents it from being consistent with zero for small cross-correlation uncertainties. For large uncertainties, Figure 2 shows that we lose the ability to detect and distinguish between isotropic and anisotropic signals in the data.

We can similarly compute scaling relations for the decision threshold as a function of the cross-correlation uncertainty and the number of pulsars in the PTA. Since this is an empirically constructed detection statistic, we do not have analytical expressions for its scaling relations, though we can derive the scaling expressions computationally, as shown in Figure 4. As expected, as we increase the number of pulsars in the PTA, the decision threshold decreases across all multipoles. Similarly, as the uncertainty on the cross-correlation measurements decreases, so do the multipole-dependent decision thresholds, corresponding to an improved sensitivity to deviations from isotropy.

4.3. Sensitivity Figure of Merit

Rather than treating the number of pulsars and the cross-correlation uncertainty as separate variables, we can consider a combination that is inspired by the weighted arithmetic mean of cross-correlation measurements involved in, e.g., S/N

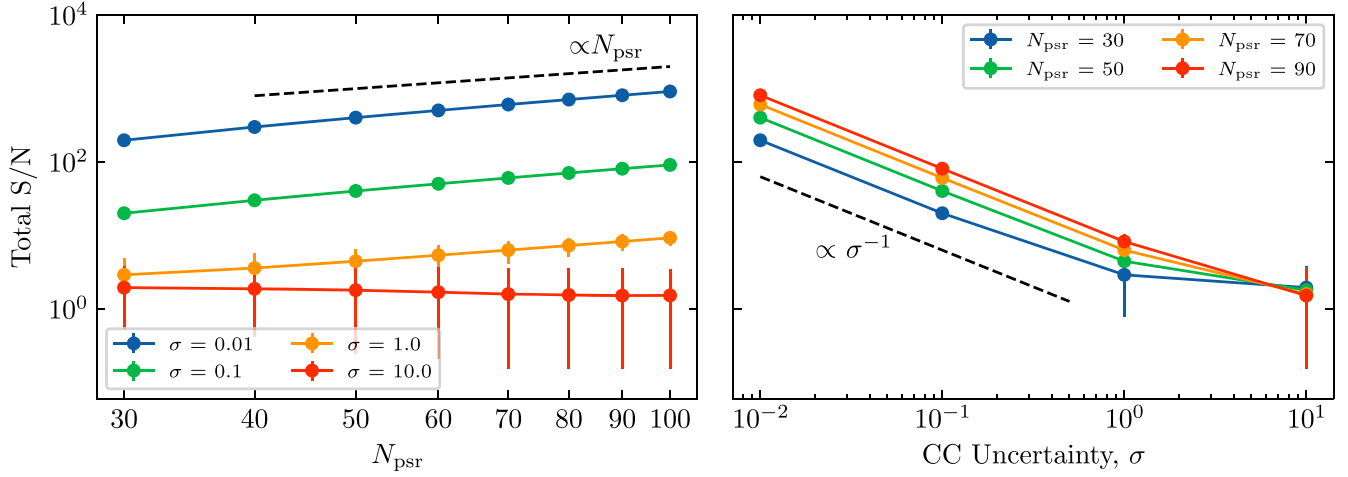


Figure 3. The evolution of the total S/N values for an ideal PTA with 100 pulsars and an isotropic injected GWB. The points represent the median values across 10^4 noise realizations, while the errors represent 95% confidence intervals. Left: Evolution of the total S/N as a function of number of pulsars in the PTA for different values of cross-correlation uncertainty. A black dashed line corresponding to the scaling relation in Section 4.2 is shown for reference. Right: Evolution of the total S/N as a function of the cross-correlation uncertainty for different numbers of pulsars in the PTA. A black dashed line corresponding to the scaling relation in Section 4.2 is shown for reference. Together, these results show that the total S/N is higher for a PTA with small cross-correlation uncertainties and a large number of pulsars.

calculations. In such calculations, operations like $\sum_{ab}(\dots)/\sigma_{ab}^2$ are proportional to $N_{\text{cc}}/\sigma^2 \propto (N_{\text{psr}}/\sigma)^2$ in the limit of equal cross-correlation measurement uncertainties, or where we can characterize the distribution of uncertainties by its mean or median over pairs. Therefore we define a sensitivity figure of merit (FOM), N_{psr}/σ , and quantify the dependence of our detection statistics with respect to this. This also allows us to quantify the trade-off between the number of pulsars in the PTA and the cross-correlation uncertainties, which, in turn, are related to the noise characteristics of the pulsars.

The relation of the total S/N and decision threshold to the sensitivity FOM is shown in Figure 5. We confirm that the total S/N is proportional to the sensitivity FOM, N_{psr}/σ , in logarithmic space. This implies that, as expected, fewer pulsars or larger uncertainties on the cross-correlation measurements result in a reduced total S/N, while larger numbers of pulsars or smaller uncertainties result in an increase in the total S/N value. Similarly, Figure 5 shows the dependence of the decision threshold for each anisotropy multipole on N_{psr}/σ . As with the total S/N, a PTA with fewer pulsars or larger uncertainties on the cross-correlation measurements will be able to reject the null hypothesis with lower significance than a PTA with more pulsars and/or smaller uncertainties on the cross-correlations.

5. Simulations with a Realistic PTA

The ideal PTA described in Section 4 is useful in discerning scaling relationships that map between the PTA design and detection statistics. However, unlike the ideal PTA in Section 4, real PTAs do not (yet) consist of pulsars distributed uniformly across the sky, nor are all the pulsars in the array identical. The latter fact implies that the cross-correlation uncertainties in a real PTA will be described by a distribution, rather than a constant value as assumed in Section 4.

To simulate a realistic PTA, we use the methods developed in Pol et al. (2021). We base our simulations on the NANOGrav 12.5 yr data set (Alam et al. 2021), and extend the data set to a 20 yr timing baseline to forecast the sensitivity of NANOGrav to anisotropies in the GWB. The TOA timestamps of the initial 12.5 yr portion are the same as those

in the real NANOGrav data set, while the radiometer uncertainties and pulse-phase jitter noise that are injected are obtained from the maximum likelihood pulsar noise analysis performed as part of the NANOGrav 12.5 yr analysis (Arzoumanian et al. 2020). The injection values for the intrinsic per-pulsar red noise are taken from a global PTA analysis that also models a common-spectrum process. This is done to isolate the intrinsic red noise in each pulsar’s data set so that it is not contaminated by the common process reported in Arzoumanian et al. (2020).

Once the 45 simulated pulsars from the NANOGrav 12.5 yr data set are generated using the above recipe, the data set is then extended into the future by generating distributions for the cadence and measurement uncertainties using the previous year’s worth of data for each pulsar. We then draw TOAs using these distributions until the data set has a maximum baseline of 20 yr. Finally, we inject 100 statistically random realizations of an isotropic GWB with an amplitude of $A_{\text{GWB}} = 2 \times 10^{-15}$ and spectral index $\alpha = -2/3$, consistent with the common process observed in Arzoumanian et al. (2020).

We then pass all 100 realizations of the data set through the standard NANOGrav detection pipeline (Arzoumanian et al. 2018, 2020; Pol et al. 2021) to calculate the cross-correlations and their uncertainties between all pairs in the 45-pulsar data set (see also Appendix A). The evolution of these cross-correlation uncertainties across the 100 realizations as a function of the timing baseline is shown in Figure 6. As we can see, the median cross-correlation uncertainty reduces from ~ 5 at 13 yr (similar to the total baseline of the 12.5 yr data set) to ~ 1 at 20 yr, implying a scaling relation $\sigma \propto T^{-7/2}$. This is shallower than the spectral index predicted for the weak-signal regime in Section 4.2, which might imply that the NANOGrav PTA is in the intermediate-signal regime (Siemens et al. 2013), also hinted at by the fact that the lowest frequencies of the PTA are dominated by a common-spectrum process as shown in Arzoumanian et al. (2020). Combining the median uncertainty with the 45 pulsars in the PTA, we obtain values for our sensitivity FOM $N_{\text{psr}}/\sigma \approx 9$ at 13 yr, and $N_{\text{psr}}/\sigma \approx 45$ at 20 yr.

We pass the cross-correlations measured from these 100 realizations through the statistical framework described in

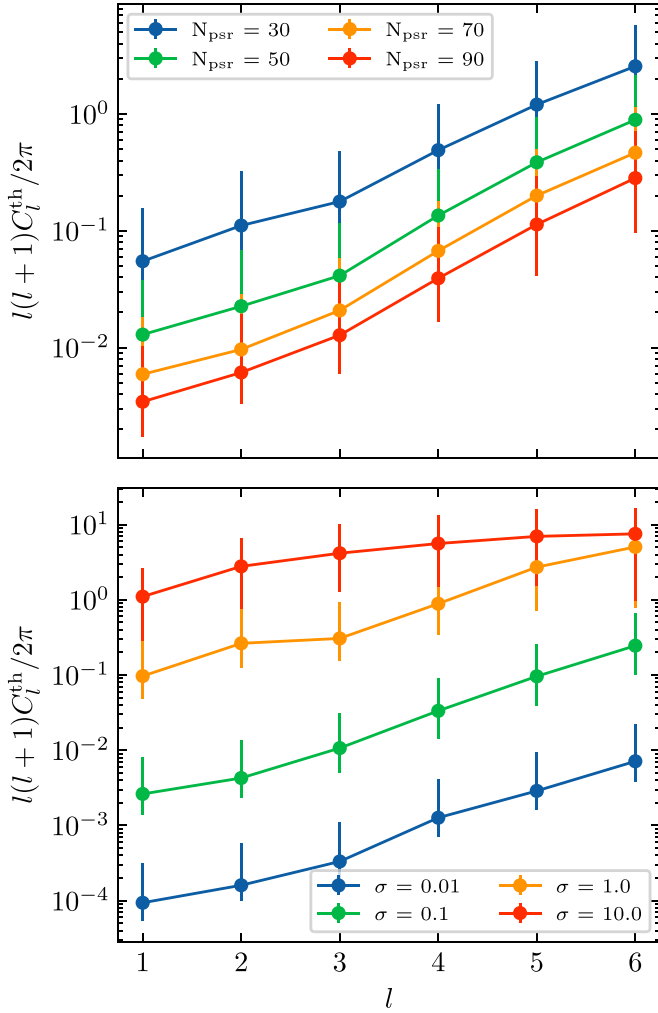


Figure 4. The evolution of the decision threshold, C_l^{th} , for an ideal PTA with an isotropic injected GWB. The points represent the median values while the errors represent the 95% confidence interval values across 10^4 noise realizations. Top: Evolution of the decision threshold per mode for different numbers of pulsars in the PTA. Bottom: Evolution of the decision threshold for different cross-correlation uncertainties. These values are generated for an ideal PTA with 100 pulsars and a cross-correlation uncertainty of 0.1. These results show that the decision threshold is lower for PTAs that have a larger number of pulsars and smaller cross-correlation uncertainties.

Section 2 to search for the presence of anisotropy under realistic PTA and data quality conditions. The evolution of the three S/N statistics as a function of time is shown in Figure 7. Since the injected GWB is isotropic, the total and isotropic S/N increase with the timing baseline. This is consistent with the reduction in uncertainties on the cross-correlations allowing for a stronger detection of the isotropic background. By contrast, the anisotropic S/N does not increase with time, and has support at S/N=0 for all baselines. Note that the total S/N seen in these realistic simulations is consistent with the prediction made in Figure 5 and Pol et al. (2021). Similarly, Figure 8 shows the evolution of the decision threshold as a function of the spherical harmonic multipole l and the timing baseline. We find that the anisotropy decision threshold is such that, in terms of the C_l values, GWB anisotropies at levels $C_{l=1}/C_{l=0} \gtrsim 0.3$ (i.e., greater than 30% of the power in the monopole) would be inconsistent with the null hypothesis of isotropy at the $p = 3 \times 10^{-3}$ level for the 20 yr baseline.

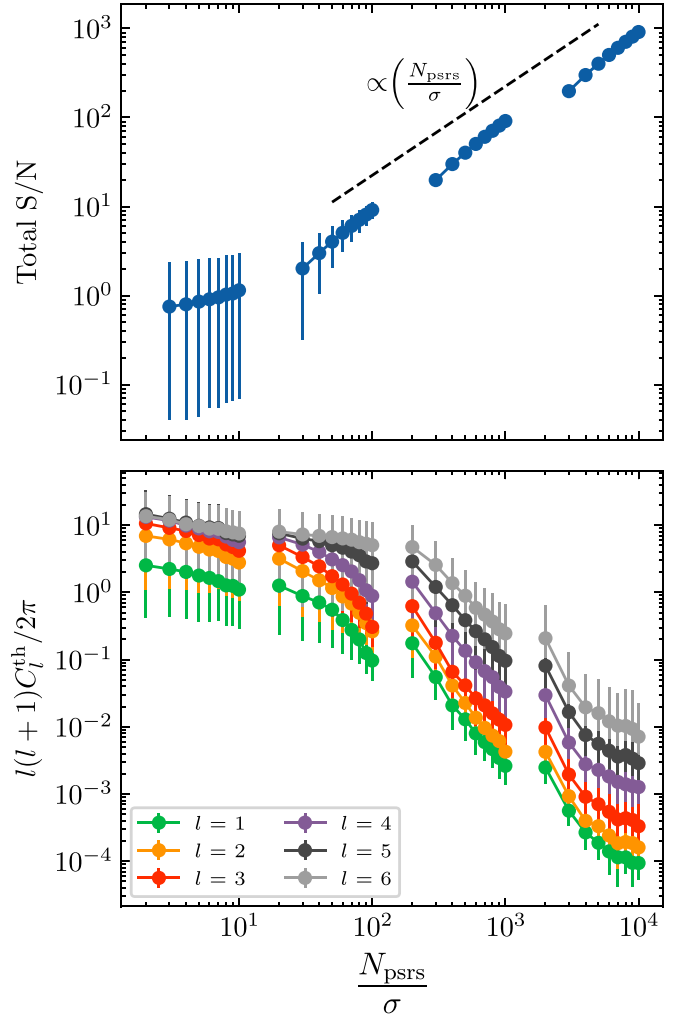


Figure 5. The evolution of detection statistics with the sensitivity FOM, N_{psr}/σ , defined in Section 4.3 for an ideal PTA with an isotropic injected GWB. The points represent the medians, while the error bars represent the 95% confidence intervals across 10^4 noise realizations. Top: Evolution of the total S/N as a function of the sensitivity FOM. This scaling relation implies that PTAs with either a large number of pulsars or small cross-correlation uncertainties (or both) will return a larger total S/N value than a PTA with fewer pulsars and/or larger cross-correlation uncertainty. Bottom: The evolution of the decision threshold as a function of the sensitivity FOM. This implies that PTAs with fewer pulsars or larger cross-correlation uncertainties will have higher decision thresholds, while PTAs with more pulsars and smaller cross-correlation uncertainties will have lower decision thresholds.

For comparison, in Figure 8 we also plot the Bayesian 95% upper limits on GWB anisotropy using six pulsars from the EPTA's first data release (Taylor et al. 2015) for a model extending to $l_{\text{max}} = 4$. This data set has a maximum baseline of 17.7 yr, which is toward the upper end of the baselines that we simulate for the NANOGrav data. However, the number of pulsars (6) in the EPTA analysis is significantly lower than the number of pulsars in our simulations (45). The longer EPTA timing baseline allows the Bayesian EPTA upper limits and NANOGrav anisotropy decision thresholds to be comparable at low multipoles until NANOGrav's timing baseline exceeds that of the EPTA. However, the larger number of pulsars in NANOGrav not only gives it higher spatial resolution (and thus access to higher multipoles), but also improves the sensitivity of NANOGrav at higher multipoles relative to the EPTA 2015 limit. As shown in Floden et al. (2022), access to these higher

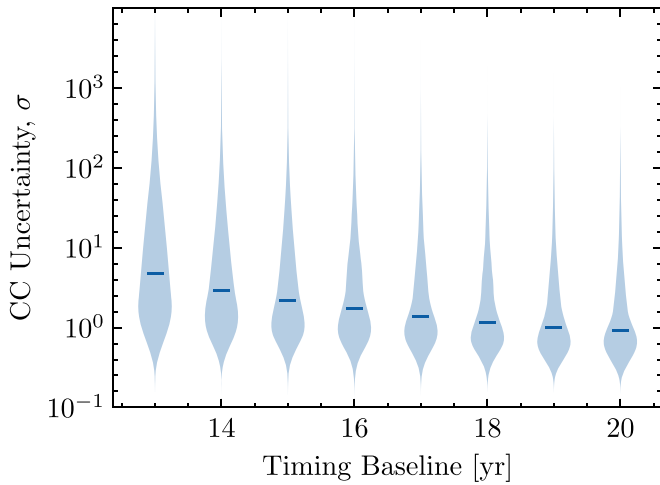


Figure 6. The evolution of the cross-correlation uncertainty across all pulsars and 100 noise realizations of the realistic PTA data set simulations described in Section 5. The evolution of the median cross-correlation uncertainty can be approximately described by $\sigma \propto T^{-7/2}$, which is shallower than the scaling law prediction of $\sigma \propto T^{-13/3}$ for the weak-signal regime in Section 4.2, but steeper than the strong-signal regime prediction of $\sigma \propto T^{-1/2}$. This implies that the NANOGrav PTA is in the intermediate-signal regime, which is corroborated by the fact that the lowest frequencies of the PTA are now dominated by a common-spectrum process (interpreted as the GWB) as shown in Arzoumanian et al. (2020).

multipoles can aid in the localization of GWB anisotropies caused by individual GW sources, or due to finiteness in the source population constituting the GWB. This highlights the importance of including more pulsars in a PTA for spatial resolution, even with the shorter timing baseline of such new additions.

6. Discussion and Conclusion

We have explored the detection and characterization of stochastic GWB anisotropy through pulsar cross-correlations in a PTA. Using a frequentist maximum likelihood approach, we can search for anisotropy by modeling GWB power in individual sky pixels or through a weighted sum of spherical harmonics. Anisotropy would then manifest in measured cross-correlations between pulsar timing residuals through a power-weighted overlap of pulsar GW antenna response functions. As a refinement on previous approaches, we prevent the GWB power from assuming unphysical negative values by adopting a model that naturally restricts this; we have referred to this as the square-root spherical harmonic basis throughout our analysis. We have also defined two detection metrics: (a) the S/N, defined as the ratio between the maximum likelihood values of a signal and a noise model, and (b) the anisotropy decision threshold, C_l^{th} , defined as the level at which the measured angular power is inconsistent with isotropy at the $p = 3 \times 10^{-3}$ ($\sim 3\sigma$) level. The S/N comes in three flavors: (i) the total S/N, which measures the strength of an anisotropic GWB model against noise alone; (ii) the isotropic S/N, which measures the strength of an isotropic GWB model against noise alone (this directly corresponds to the usual optimal statistic S/N used in PTA data analysis); and (iii) the anisotropic S/N, which measures the statistical preference for anisotropy over isotropy.

We examined the evolution of these detection statistics as a function of the uncertainty on the measured cross-correlations, as well as of the number of pulsars in the PTA. We showed that

the cross-correlation uncertainty and the number of pulsars in the PTA can be combined into a single FOM for the PTA sensitivity, N_{psr}/σ , which succinctly maps the PTA configuration and noise specifications to the detectability of anisotropy. Our scaling relations show that increasing the number of pulsars in an array, while reducing the uncertainty on the cross-correlation measurements, leads to higher total S/N and lower anisotropy decision thresholds. As shown in Section 3, the cross-correlation uncertainty scales inversely with both the timing baseline and the cadence of observation for each pulsar, with the power-law index dependent on the signal regime occupied by the PTA. The pulsar timing baseline is set to increase as PTAs continue operation into the future, and as IPTA data combinations are utilized (e.g., Perera et al. 2019). Improving observing cadence is more challenging due to constraints on the available telescope time for each PTA, though once again IPTA data combinations can help alleviate this problem. In addition to IPTA data combinations, the CHIME radio telescope (CHIME/Pulsar Collaboration et al. 2021) will offer near-daily cadence to the NANOGrav PTA (McLaughlin 2013), which is an order-of-magnitude improvement over the current near-monthly cadence employed by NANOGrav.

Finally, we examined the evolution of the S/N and anisotropy decision thresholds as a function of timing baseline using realistic NANOGrav data. Since we injected an isotropic GWB signal in these data, we found that the anisotropic S/N remains consistent with zero at all times and across all signal realizations. By contrast, the total and isotropic S/N increase with time, as expected (Siemens et al. 2013). We found that any anisotropic GWB power distribution with $C_{l=1} \gtrsim 0.3C_{l=0}$ would be in tension with an isotropic model at the $p = 3 \times 10^{-3}$ ($\sim 3\sigma$) level. We note that these simulations held the number of pulsars in the array fixed to the 45 that were included in the NANOGrav 12.5 yr data set. However, this number will increase in future NANOGrav data sets. Based on the results in Section 4.3, this will lead to larger total S/N values and lower anisotropy decision thresholds, implying improved sensitivity to any anisotropy that might be present in the real GWB. IPTA data combinations will further increase the timing baseline and number of pulsars in the array allowing further improvements on the ability to detect anisotropy. Furthermore, new instruments such as ultrawideband receivers, and new telescopes such as MeerKAT (Bailes et al. 2020), the Square Kilometre Array (Bailes et al. 2016), and DSA-2000 (Hallinan et al. 2019) will aid in the detection and characterization of anisotropy with PTAs.

The techniques and scaling relations that we have developed in this work are PTA-agnostic and can be projected onto any PTA specification, allowing for immediate usage by the broader PTA community. Yet we have made assumptions in our framework that can be generalized in the future. For example, while our techniques operate on PTA data at the level of cross-correlations rather than TOAs, in order to get to that stage we have implicitly assumed that the GWB characteristic strain spectrum is well described by a power-law model. This follows the same approach as Arzoumanian et al. (2020), where an average power-law spectrum $h_c \propto f^{-2/3}$ is assumed for the GWB. For a GWB produced by a population of inspiraling SMBHBs, this power-law representation is an approximation to the true spectrum (Phinney 2001), where there are different SMBHBs contributing to the GWB at different frequencies

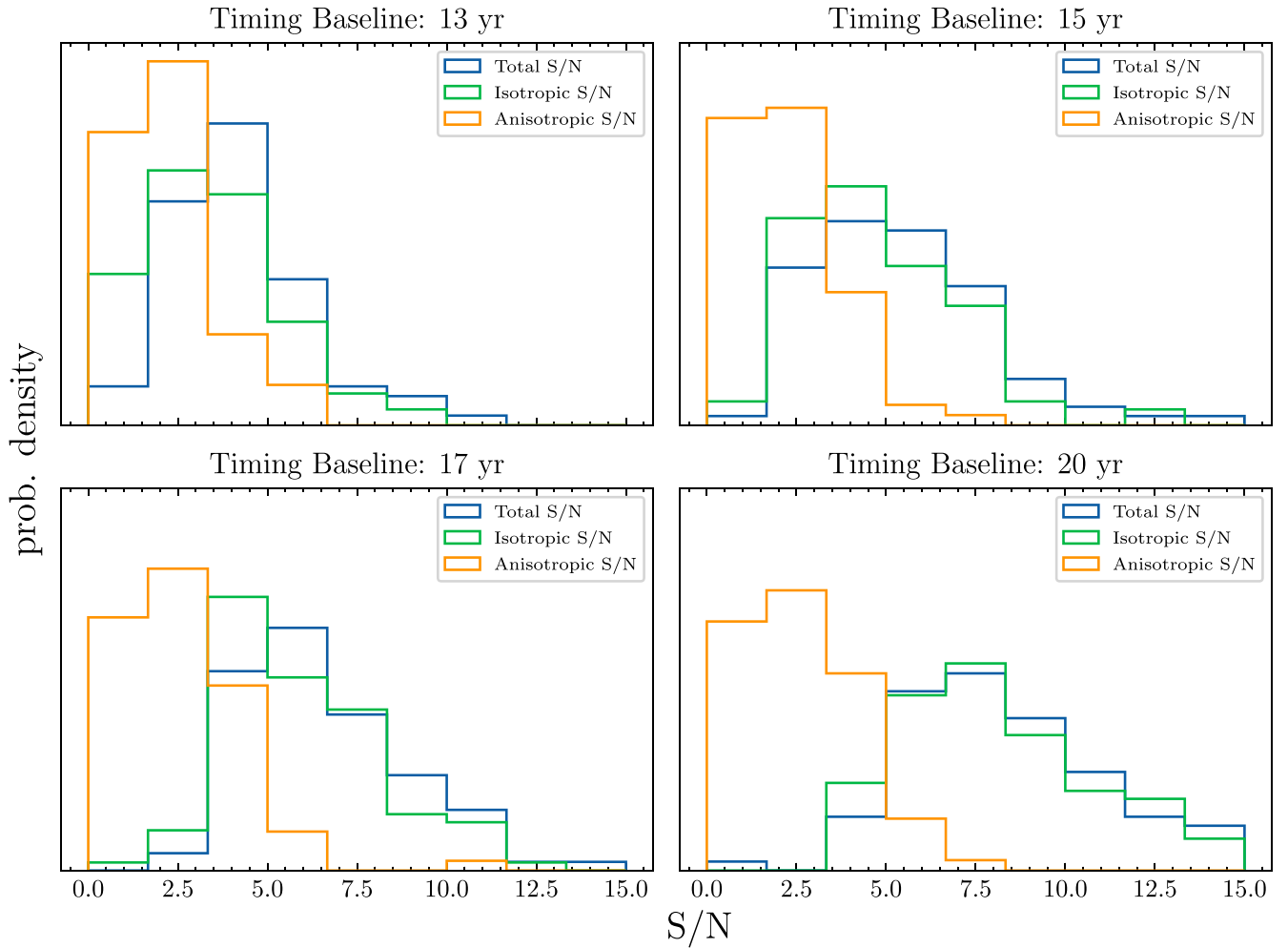


Figure 7. Evolution of the S/N values for the realistic simulations described in Section 5. The total, isotropic, and anisotropic S/N are shown by the blue, green, and orange histograms, respectively. Since the injected GWB is isotropic, we see the total and isotropic S/N values increase as a function of the timing baseline, while the anisotropic S/N stays consistent with zero for all baselines.

(Kelley et al. 2017). Thus, a more appropriate way to characterize anisotropy in the SMBHB GWB would be to measure the cross-correlations of pulsar TOAs as a function of GW frequency, rather than use our current approach of computing cross-correlations that are filtered against a power-law GWB spectral template (see Appendix A). We would then have a more general data structure that includes pulsar cross-correlations and uncertainties for each GW frequency, for which the methods developed here can be applied at each of those frequencies independently. We also note that the methods developed here can be modified to search for multiple backgrounds (Suresh et al. 2021), where an astrophysical background (e.g., from SMBHBs; Sesana et al. 2004; Kelley et al. 2017; Burke-Spolaor et al. 2019) would be expected to be anisotropic, while a cosmological background (e.g., from cosmic strings; Olmez et al. 2012) may be isotropic.

We plan to explore these improvements and generalizations in future analyses in a bid to extract as much spatial and angular information as possible from the exciting new PTA data sets now under development. As mentioned, these techniques will aid not only in the detection of GWB anisotropy, but also in its characterization for the purposes of isolating regions of excess power that may be indicative of individually resolvable GW sources, and as leverage for the

separation of potentially multiple stochastic GW signals of astrophysical and cosmological origin.

N.P. was supported by the Vanderbilt Initiative in Data Intensive Astrophysics (VIDA) Fellowship. S.R.T. acknowledges support from NSF AST-2007993, PHY-2020265, and PHY-2146016, and a Vanderbilt University College of Arts & Science Dean’s Faculty Fellowship. J.D.R. was partially supported by startup funds provided by Texas Tech University.

This work has been carried out by the NANOGrav Collaboration, which is part of the IPTA. The NANOGrav project receives support from National Science Foundation Physics Frontiers Center award Nos. 1430284 and 2020265.

This work was conducted using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

Software: We use `lmfit` (Newville et al. 2021) for the nonlinear least-squares minimization in the square-root spherical harmonic basis. We use `libstempo` (Vallisneri et al. 2021) to generate our realistic PTA data sets and to inject the pulsar noise parameters and GWB signals in these data sets. We use the software packages `Enterprise` (Ellis et al. 2020) and `enterprise_extensions` (Taylor et al. 2021) for model construction, along with `PTMCMCSampler`

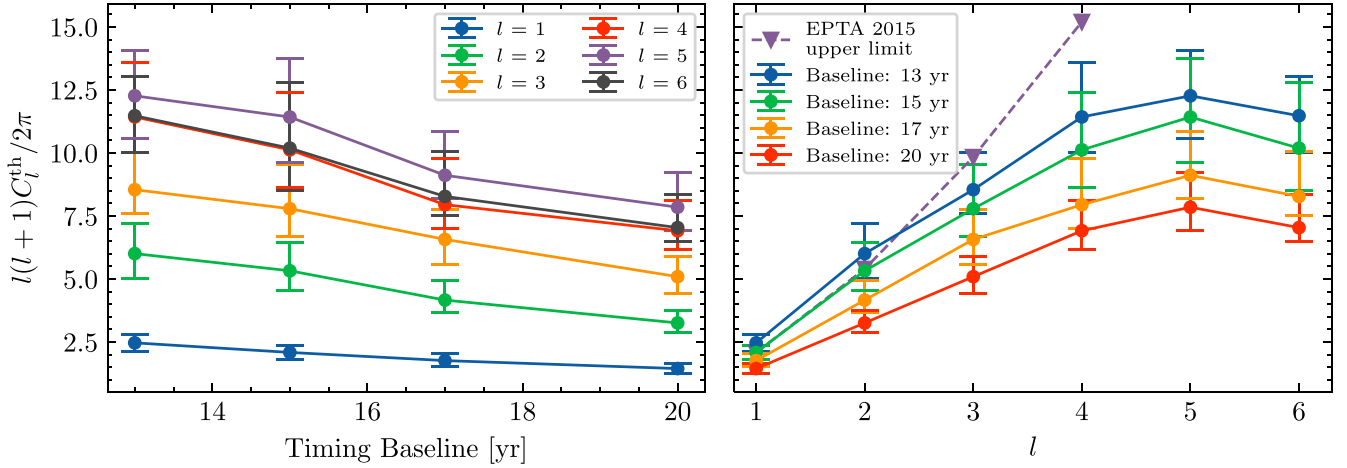


Figure 8. Evolution of decision threshold for realistic simulations described in Section 5. Left: Evolution of the decision threshold as a function of timing baseline for all spherical harmonic multipoles. The decision threshold decreases with an increase in the timing baseline, and higher spherical harmonic multipoles have a higher decision threshold than lower multipoles. Right: Evolution of the decision threshold across spherical harmonic multipoles for different timing baselines. We also plot the Bayesian 95% upper limits on anisotropy derived in Taylor et al. (2015) from EPTA Data Release 1. As these realistic simulations have 45 pulsars with different noise properties resulting in different cross-correlation uncertainties per pulsar pair, we see the sensitivity of the PTA saturate at higher multipoles.

(Ellis & van Haasteren 2017) as the Markov Chain Monte Carlo sampler for our realistic PTA Bayesian analyses. We also extensively use Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), Python (Oliphant 2007; Millman & Aivazis 2011), and SciPy (Virtanen et al. 2020).

Appendix A Computing Cross-correlations in Realistic Pulsar Timing Data Sets

Reiterating Equation (1), the cross-correlation value and its uncertainty measured between pulsar a and pulsar b are

$$\rho_{ab} = \frac{\delta \mathbf{t}_a^T \mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \delta \mathbf{t}_b}{\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}]}, \quad \sigma_{ab} = (\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}])^{-1/2}, \quad (\text{A1})$$

where $\delta \mathbf{t}_a$ is a vector of timing residuals for pulsar a , $\mathbf{P}_a = \langle \delta \mathbf{t}_a \delta \mathbf{t}_a^T \rangle$ is the measured autocovariance matrix of pulsar a , and $\hat{\mathbf{S}}_{ab}$ is the template-scaled covariance matrix between pulsar a and pulsar b . This scaled covariance matrix is a template for the spectral shape only, and is amplitude- and ORF-independent. It is related to the full covariance matrix by $\mathbf{S}_{ab} = \langle \delta \mathbf{t}_a \delta \mathbf{t}_b^T \rangle = \mathbf{A}^2 \chi_{ab} \hat{\mathbf{S}}_{ab}$.

Covariance matrices in the PTA likelihood are modeled using rank-reduced approximations, e.g., long-timescale stochastic processes are modeled on a truncated Fourier basis with a number of frequencies that is much smaller than the number of TOAs. The induced timing delays of such stochastic processes are modeled as $\mathbf{r} = \mathbf{T}\mathbf{b}$, where \mathbf{b} is a vector of amplitude coefficients for the basis design matrix $\mathbf{T} = [\mathbf{M} \mathbf{F}]$, which itself is a concatenation of the stabilized timing model design matrix \mathbf{M} and Fourier design matrix \mathbf{F} . The former is a matrix of partial derivatives of TOAs with respect to timing model parameters, which is then column-normed or stabilized by singular-value decomposition to reduce the dynamic range of the entries. The latter is a matrix of sine and cosine basis functions evaluated at all TOAs for each sampling frequency of the time series. The autocovariance matrix of pulsars is then modeled as $\mathbf{P}_a = \mathbf{N}_a + \mathbf{T}_a \mathbf{B}_{aa} \mathbf{T}_a^T$, where \mathbf{N}_a is the white noise

covariance matrix of pulsar a , and

$$\mathbf{B} = \begin{bmatrix} \infty & 0 \\ 0 & \phi \end{bmatrix} \quad (\text{A2})$$

is the covariance matrix of the \mathbf{b} coefficients, $\mathbf{B} = \langle \mathbf{b} \mathbf{b}^T \rangle$, with infinite variance in the timing model diagonal entries to approximate an unbounded uniform prior, and ϕ as the variance of the Fourier coefficients that is related to the PSD of the stochastic process. The interpulsar covariance matrix is treated similarly, except that the timing model entries can be completely ignored since they will be uncorrelated between different pulsars. Hence, $\hat{\mathbf{S}}_{ab} = \mathbf{F}_a \hat{\phi} \mathbf{F}_b^T$, where $\hat{\phi}$ is the scaled Fourier covariance matrix of the GWB (i.e., it is amplitude- and ORF-independent).

We use the Woodbury matrix lemma to perform inversions, such that

$$\begin{aligned} \mathbf{P}_a^{-1} &= (\mathbf{N}_a + \mathbf{T}_a \mathbf{B}_{aa} \mathbf{T}_a^T)^{-1} \\ &= \mathbf{N}_a^{-1} - \mathbf{N}_a^{-1} \mathbf{T}_a (\mathbf{B}_{aa}^{-1} + \mathbf{T}_a^T \mathbf{N}_a^{-1} \mathbf{T}_a)^{-1} \mathbf{T}_a^T \mathbf{N}_a^{-1} \\ &= \mathbf{N}_a^{-1} - \mathbf{N}_a^{-1} \mathbf{T}_a \Sigma_a^{-1} \mathbf{T}_a^T \mathbf{N}_a^{-1}. \end{aligned} \quad (\text{A3})$$

Hence the numerator of Equation (1) can be written as

$$\delta \mathbf{t}_a^T \mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \delta \mathbf{t}_b = \mathbf{X}_a^T \hat{\phi} \mathbf{X}_b, \quad (\text{A4})$$

where

$$\mathbf{X}_a = \mathbf{F}_a^T \mathbf{N}_a^{-1} \delta \mathbf{t}_a - \mathbf{F}_a^T \mathbf{N}_a^{-1} \mathbf{T}_a \Sigma_a^{-1} \mathbf{T}_a^T \mathbf{N}_a^{-1} \delta \mathbf{t}_a, \quad (\text{A5})$$

and the denominator is written as

$$\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}] = \text{tr}[\mathbf{Z}_a \hat{\phi} \mathbf{Z}_b \hat{\phi}], \quad (\text{A6})$$

where

$$\mathbf{Z}_a = \mathbf{F}_a^T \mathbf{N}_a^{-1} \mathbf{F}_a - \mathbf{F}_a^T \mathbf{N}_a^{-1} \mathbf{T}_a \Sigma_a^{-1} \mathbf{T}_a^T \mathbf{N}_a^{-1} \mathbf{F}_a. \quad (\text{A7})$$

Crucially, the \mathbf{X} and \mathbf{Z} matrices for each pulsar depend on the measured noise characteristics (e.g., white noise and intrinsic red noise) or fixed aspects of the modeling (e.g., timing model and Fourier design matrices). They can thus be constructed and cached for follow-up analysis. The diagonal matrix $\hat{\phi}$ acts as a spectral template that we use in a procedure

akin to matched filtering, where we assess the S/N of cross-correlated signals with, e.g., a power-law PSD that has a $-13/3$ exponent, matching expectations for a circular, GW-driven population of SMBHBs forming a stochastic GWB.

Thus, in production-level PTA searches for the stochastic GWB, the cross-correlations are computed as

$$\begin{aligned}\rho_{ab} &= \frac{\mathbf{X}_a^T \hat{\Phi} \mathbf{X}_b}{\text{tr}[\mathbf{Z}_a \hat{\Phi} \mathbf{Z}_b \hat{\Phi}]}, \\ \sigma_{ab} &= (\text{tr}[\mathbf{Z}_a \hat{\Phi} \mathbf{Z}_b \hat{\Phi}])^{-1/2}.\end{aligned}\quad (\text{A8})$$

Appendix B

Cross-correlation Measurement Uncertainty from PTA Specifications

As shown in Section 3, the trace in Equation (1) can be written as

$$\text{tr}[\mathbf{P}_a^{-1} \hat{\mathbf{S}}_{ab} \mathbf{P}_b^{-1} \hat{\mathbf{S}}_{ba}] = \frac{2T}{A^4} \int_{f_l}^{f_h} df \frac{P_g^2(f)}{P_a(f)P_b(f)}, \quad (\text{B1})$$

where T is the timing baseline, A is the amplitude of the GWB, $P_g(f)$ is the PSD of timing residuals induced by the GWB, $P_a(f)$ and $P_b(f)$ are the intrinsic PSDs of timing residuals in pulsars a and b , and $f_l = 1/T$ and f_h are the low- and high-frequency cutoffs used in the GWB analysis. For a GWB with a characteristic strain spectrum

$$h_c(f) = A \left(\frac{f}{f_{\text{yr}}} \right)^\alpha, \quad (\text{B2})$$

and using the convention from Siemens et al. (2013), the GWB spectrum can be written in the timing residual space as

$$P_g(f) = \frac{A^2}{12\pi^2} \left(\frac{f}{f_{\text{ref}}} \right)^{2\alpha} f^{-3} = bf^{-\gamma}, \quad (\text{B3})$$

where $\gamma = 3 - 2\alpha$, and f_{ref} is a reference frequency. Consequently, the uncertainty on the cross-correlations can be written as

$$\sigma_{ab} = \frac{A^2}{\sqrt{2T}} \left[\int_{f_l}^{f_h} df \frac{P_g^2(f)}{P_a(f)P_b(f)} \right]^{-1/2}. \quad (\text{B4})$$

As described in Siemens et al. (2013), assuming no intrinsic red noise in the pulsars, and that all pulsars are identical, the intrinsic power can be written as the sum of the GWB power and the white noise, w , in each pulsar, $P_a(f) = P_b(f) = P_g(f) + 2w^2\Delta t$, where the cadence of pulsar observations is given by $c = 1/\Delta t$. Substituting the intrinsic and GWB power in Equation (B4), the cross-correlation uncertainty can be written as

$$\sigma_{ab} = \frac{A^2}{\sqrt{2T}} \left[\int_{f_l}^{f_h} df \frac{b^2 f^{-2\gamma}}{(bf^{-\gamma} + 2w^2\Delta t)^2} \right]^{-1/2}. \quad (\text{B5})$$

The solution for the integral in Equation (B5) is nontrivial, but as described in Siemens et al. (2013) and Chamberlin et al. (2015), it can be solved using hypergeometric functions. We can define three regimes in which PTAs can operate: (i) the weak-signal regime, where the intrinsic pulsar white noise dominates over the GWB power, i.e., $2w^2\Delta t \gg bf^{-\gamma}$; (ii) the

strong-signal regime, where the GWB power dominates the intrinsic white noise, i.e., $2w^2\Delta t \ll bf^{-\gamma}$; and (iii) the intermediate-signal regime, where only the lowest few frequency bins in the PTA are dominated by the GWB, while at higher frequencies, the white noise dominates the GWB power. Given the recent results from the regional PTAs (Arzoumanian et al. 2020; Chen et al. 2021; Goncharov et al. 2021) and the IPTA (Antoniadis et al. 2022), current PTAs are in the intermediate-signal regime.

In the weak-signal regime, Siemens et al. (2013) showed that Equation (B5) can be written as

$$\sigma_{ab} \approx \frac{A^2}{\sqrt{2T}} \left[\frac{b^2}{(2w^2\Delta t)^2} \frac{T^{2\gamma-1}}{2\gamma-1} \right]^{-1/2} \quad (\text{B6})$$

which reduces to

$$\sigma_{ab} \approx 12\pi^2 \frac{w^2}{c} \sqrt{4\gamma-2} \times T^{-\gamma}. \quad (\text{B7})$$

Similarly, in the strong-signal regime, the same integral reduces to the equation of Siemens et al. (2013):

$$\sigma_{ab} \approx \frac{1}{2} \frac{A^2}{\sqrt{cT}}. \quad (\text{B8})$$

Note that the scaling expressions above include the GWB amplitude, while in the rest of this paper, we use amplitude-scaled cross-correlation values and uncertainties.

ORCID iDs

Nihan Pol  <https://orcid.org/0000-0002-8826-1285>
Stephen R. Taylor  <https://orcid.org/0000-0003-0264-1453>
Joseph D. Romano  <https://orcid.org/0000-0003-4915-3246>

References

- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021, *PhRvD*, **104**, 022005
- Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021, *ApJS*, **252**, 4
- Ali-Haïmoud, Y., Smith, T. L., & Mingarelli, C. M. F. 2020, *PhRvD*, **102**, 122005
- Ali-Haïmoud, Y., Smith, T. L., & Mingarelli, C. M. F. 2021, *PhRvD*, **103**, 042009
- Allen, B., & Romano, J. D. 1999, *PhRvD*, **59**, 102001
- Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, *PhRvD*, **79**, 084030
- Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, *MNRAS*, **510**, 4873
- Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, *ApJL*, **905**, L34
- Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018, *ApJ*, **859**, 47
- Bailes, M., Barr, E., Bhat, N. D. R., et al. 2016, in Proc. of MeerKAT Science: On the Pathway to the SKA (Trieste: PoS), **11**
- Bailes, M., Jameson, A., Abbate, F., et al. 2020, *PASA*, **37**, e028
- Banagiri, S., Criswell, A., Kuan, T., et al. 2021, *MNRAS*, **507**, 5451
- Boyle, L., & Pen, U.-L. 2012, *PhRvD*, **86**, 124028
- Burke-Spolaor, S., Taylor, S. R., Charisi, M., et al. 2019, *A&ARv*, **27**, 5
- Chamberlin, S. J., Creighton, J. D. E., Siemens, X., et al. 2015, *PhRvD*, **91**, 044048
- Chen, S., Caballero, R. N., Guo, Y. J., et al. 2021, *MNRAS*, **508**, 4970
- CHIME/Pulsar Collaboration, Amiri, M., Bandura, K. M., et al. 2021, *ApJS*, **255**, 5
- Cornish, N. J., & van Haasteren, R. 2014, arXiv:1406.4511
- Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, *ApJ*, **762**, 94
- Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2020, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust Pulsar Inference Suite, Zenodo, doi:10.5281/zenodo.4059815
- Ellis, J. A., & van Haasteren, R. 2017, PTMCMCSampler, Zenodo, doi:10.5281/zenodo.1037579
- Flanagan, E. E. 1993, *PhRvD*, **48**, 2389
- Floden, E., Mandic, V., Matas, A., & Tsukada, L. 2022, *PhRvD*, **106**, 023010

- Gair, J., Romano, J. D., Taylor, S., & Mingarelli, C. M. F. 2014, *PhRvD*, **90**, 082001
- Goncharov, B., Shannon, R. M., Reardon, D. J., et al. 2021, *ApJ*, **917**, 19
- Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, **622**, 759
- Grishchuk, L. P. 2005, *PhyU*, **48**, 1235
- Hallinan, G., Ravi, V., Weinreb, S., et al. 2019, *BAAS*, **51**, 255
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hellings, R. W., & Downs, G. S. 1983, *ApJL*, **265**, L39
- Hobbs, G. 2013, *CQGra*, **30**, 224007
- Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010, *CQGra*, **27**, 084013
- Hotinli, S. C., Kamionkowski, M., & Jaffe, A. H. 2019, *OJAp*, **2**, 8
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Ivezić, Ž., Connelly, A. Z., Vanderplas, J. T., & Gray, A. 2019, *Statistics, Data Mining, and Machine Learning in Astronomy* (Princeton, NJ: Princeton Univ. Press)
- Jaffe, A. H., & Backer, D. C. 2003, *ApJ*, **583**, 616
- Jenkins, A. C. 2022, arXiv:2202.05105
- Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2017, *MNRAS*, **471**, 4508
- Kramer, M., & Champion, D. J. 2013, *CQGra*, **30**, 224009
- Levenberg, K. 1944, *QApMa*, **2**, 164
- Marquardt, D. W. 1963, *SIAM*, **11**, 431
- McLaughlin, M. A. 2013, *CQGra*, **30**, 224008
- Millman, K. J., & Aivazis, M. 2011, *CSE*, **13**, 9
- Mingarelli, C. M. F., Sidery, T., Mandel, I., & Vecchio, A. 2013, *PhRvD*, **88**, 062005
- Newville, M., Otten, R., Nelson, A., et al. 2021, *lmfit/lmfit-py*: 1.0.3, Zenodo, doi:10.5281/zenodo.5570790
- Oliphant, T. E. 2007, *CSE*, **9**, 10
- Olmez, S., Mandic, V., & Siemens, X. 2010, *PhRvD*, **81**, 104028
- Olmez, S., Mandic, V., & Siemens, X. 2012, *JCAP*, **2012**, 009
- Payne, E., Banagiri, S., Lasky, P. D., & Thrane, E. 2020, *PhRvD*, **102**, 102004
- Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, *MNRAS*, **490**, 4666
- Phinney, E. S. 2001, arXiv e-prints, astro, arXiv:abs/astro-ph/0108028
- Pol, N. S., Taylor, S. R., Zoltan Kelley, L., et al. 2021, *ApJL*, **911**, L34
- Rajagopal, M., & Romani, R. W. 1995, *ApJ*, **446**, 543
- Romano, J. D., & Cornish, N. 2017, *LRR*, **20**, 2
- Romano, J. D., Hazboun, J. S., Siemens, X., & Archibald, A. M. 2021, *PhRvD*, **103**, 063027
- Sandage, A. 1958, *ApJ*, **127**, 513
- Sesana, A., Haardt, F., Madau, P., & Volonteri, M. 2004, *ApJ*, **611**, 623
- Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, *CQGra*, **30**, 224015
- Suresh, J., Agarwal, D., & Mitra, S. 2021, *PhRvD*, **104**, 102003
- Taylor, S. R., Baker, P. T., Hazboun, J. S., Simon, J., & Vigeland, S. J. 2021, *enterprise_extensions* v2.3.1, https://github.com/nanograv/enterprise_extensions Enterprise Extensions
- Taylor, S. R., & Gair, J. R. 2013, *PhRvD*, **88**, 084001
- Taylor, S. R., van Haasteren, R., & Sesana, A. 2020, *PhRvD*, **102**, 084039
- Taylor, S. R., Mingarelli, C. M. F., Gair, J. R., et al. 2015, *PhRvL*, **115**, 041101
- Thrane, E., Ballmer, S., Romano, J. D., et al. 2009, *PhRvD*, **80**, 122002
- Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, *MNRAS*, **455**, 4339
- Vallisneri, M., Ellis, J. A., van Haasteren, R., & Taylor, S. R. 2021, *libstempo*, <https://github.com/vallis/libstempo>
- Vigeland, S. J., Islo, K., Taylor, S. R., & Ellis, J. A. 2018, *PhRvD*, **98**, 044003
- Virtanen, P., Gommers, R., Burovski, E., et al. 2020, *scipy/scipy*: SciPy 1.5.3, v1.5, Zenodo, doi:10.5281/zenodo.4100507