Exploring the capabilities of Gibbs sampling in pulsar timing arrays

Nima Laal[,1,*] William G. Lamb,[2] Joseph D. Romano,[3] Xavier Siemens,[1]
Stephen R. Taylor,[2] and Rutger van Haasteren[4]

[1]*Department of Physics, Oregon State University, 1500 SW Jefferson Way, Corvallis, Oregon 97331, USA*
[2]*Department of Physics and Astronomy, Vanderbilt University,*
2301 Vanderbilt Place, Nashville, Tennessee 37235, USA
[3]*Department of Physics, Texas Tech University, Box 41051, Lubbock, Texas 79409, USA*
[4]*Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut),*
Callinstrasse 38, D-30167, Hannover, Germany

We explore the use of Gibbs sampling in estimating the noise properties of individual pulsars and illustrate its effectiveness using the NANOGrav 11-year dataset. We find that Gibbs sampling noise modeling (GM) is more efficient than the current standard Bayesian techniques (SM) for single pulsar analyses by yielding model parameter posteriors with average effective-sample-size ratio (GM/SM) of 6 across all parameters and pulsars. Furthermore, the output of GM contains posteriors for the Fourier coefficients that can be used to characterize the underlying red noise process of any pulsar's timing residuals, which are absent in current implementations of SM. Through simulations, we demonstrate the potential for such coefficients to measure the spatial cross-correlations between pulsar pairs produced by a gravitational wave background.

I. INTRODUCTION

Pulsar timing arrays (PTAs) [1,2] are low-frequency gravitational-wave (GW) detectors that use high-precision measurements of the times-of-arrival (TOAs) of pulses produced by an array of millisecond pulsars (MSPs). MSPs have ultra-stable spin periods on the order of milliseconds, and if their TOAs are measured to sufficient accuracy using large and sensitive radio telescopes, they can be used as cosmic clocks spread throughout our galaxy. Accurate models are constructed to predict the time at which each pulse is expected to arrive, and small deviations from the expected TOAs caused by GWs can be detected by searching for quadrupolar spatial correlations in those deviations between pulsars in the PTA [3].

In recent years, multiple PTA searches for an isotropic stochastic gravitational wave background (GWB) have uncovered a common red noise process [4–7]. This process was recently observed to posses a quadrupolar correlation signature matching the predictions of Einstein's general theory of relativity with various levels of significance [8–11].

The sensitivity of PTAs to a GWB depends primarily on the number of pulsars in the array [12]. This is due to the fact that, at late times, the lowest frequencies in PTA datasets become GW-dominated, and the significance of the cross-correlations grows with the square root of the time span of the data and linearly with the number of pulsars in the array. In this regime, increasing the number of pulsars is the best way to maximize PTA sensitivity to the GWB. Currently, the International Pulsar Timing Array (IPTA) monitors 65 millisecond pulsars with 27 of such pulsars observed for more than 10 years [5]. For this reason, in each new release of a PTA dataset the number of pulsars used in GWB detection analyses is expected to grow, which in turn makes the computational cost of noise modeling and parameter estimation increase significantly. This poses a significant challenge for Bayesian inference as typical searches for a GWB involve working with a very large parameter space making the use of computationally efficient algorithms a necessity.

The standard Bayesian techniques for single and multipulsar noise modeling often result in a joint probability distribution for all of the model parameters (see Sec. II D). Despite the flexibility that this approach offers in choosing and implementing various noise models, the computational cost of parameter estimation using Markov Chain Monte Carlo (MCMC) simulations becomes prohibitive quickly. For instance, in the case of single-pulsar analyses, the number of parameters required to describe a pulsar's noise may well exceed forty (see Sec. III). This problem is more severe for the case of multipulsar analyses as even the simplest noise models require a number of parameters that is larger than twice the number of pulsars in the PTA. Hence, more computationally efficient data analysis techniques are critical for the future of PTA analyses.

063008-1

To mitigate these problems, there have been numerous efforts toward the development of more efficient Bayesian GWB detection techniques to analyze PTA datasets, such as those presented in [13–19]. In particular, the work of van Haasteren and Vallisneri [14] provides an outline for single-pulsar noise analyses in which Gibbs sampling can be used to characterize the red noise component of each pulsar's timing residuals. In this paper, we explore the capabilities of the Gibbs sampling method in single-pulsar noise analyses by applying it on the NANOGrav 11-year dataset 20]] as well as simulated datasets. We show that the Gibbs sampling method is well suited for PTA single-pulsar analyses and results in probability distribution functions for all model parameters in a significantly shorter timescale compared to those obtained via the standard MCMC methods. Furthermore, we show, via simulated datasets, that the Fourier coefficients that result from the Gibbs sampling procedure can be used to identify the shape of the underlying spatially-correlated signal in a PTA dataset.

The paper is structured as follows. In Sec. II, we review and simplify the methods presented in van Haasteren and Vallisneri [14] to outline the Gibbs sampling method and its accompanying noise modeling. Furthermore, in order to use the output of Gibbs sampling in a subsequent multi-pulsars analysis, and inspired by Anholm *et al.* [21], we introduce our version of a frequency domain optimal statistic which follows from the PTA multipulsar likelihood function. In Sec. III, we employ the outlined method in order to analyze the NANOGrav 11 year dataset and compare the results to those obtained by the standard Bayesian PTA detection techniques. Finally, in Sec. IV, we analyze PTA simulated datasets to reveal the potential of the Gibbs sampling technique in searches for a common correlated signal across an array of pulsars.

II. METHODS

We begin our review of the Gibbs sampling method [14] by writing a simple model for a pulsar's post-fit timing residuals, r, in terms of a set of Fourier coefficients a, Fourier design matrix F, linear timing model parameters ϵ, timing design matrix M, and white noise w[1]:

$$r = M\epsilon + Fa + w$$
$$= Tb + w, \tag{1}$$

where $b^T = [\epsilon, a]$ and $T = [M, F]$. Assuming Gaussian white noise, parametrized by the set of parameters n with prior $p(n)$, the above model allows for the construction of posterior probability density functions following Bayes' theorem:

$$p(\rho, b, n | r) \propto p(r | b, n) p(a) p(\epsilon) p(\rho) p(n), \tag{2}$$

where,

$$p(r | b, n) = \frac{\exp\left\{-\frac{1}{2}[(r - Tb)^T N^{-1}(r - Tb)]\right\}}{\sqrt{\det\{(2\pi)N\}}}, \tag{3}$$

$$p(a) = \frac{\exp\left\{-\frac{1}{2}[a^T \varphi^{-1} a]\right\}}{\sqrt{\det\{(2\pi)\varphi\}}}, \tag{4}$$

$$p(\rho) = \prod_{s=1}^{k} \frac{1}{\rho_s}, \tag{5}$$

for

$$\varphi = \langle aa^T \rangle, \tag{6}$$

$$B = \langle bb^T \rangle, \tag{7}$$

and ρ denoting the collective set $\{\rho_1, \rho_2, \ldots, \rho_k\}$ whose elements are used to parametrize a pulsar's power-spectral-density, frequency-bin by frequency-bin,[2] and describe the variance of the Fourier coefficients. Additionally, a log-uniform (conjugate) prior $p(\rho_k) = 1/\rho_k$ is considered as seen in Eq. (5).

Moreover, we have assumed an unbounded improper prior for the linear timing model parameters and have set $\langle \epsilon\epsilon^T \rangle = \text{diag}\{\infty\}$. Such choices for the linear timing model parameters are typical of PTA noise analyses due to the lack of physically motivated priors for all of the timing model parameters and are acceptable as long as the data is informative with respect to such parameters. Hence, we can write

$$B^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \varphi^{-1} \end{bmatrix}. \tag{8}$$

To proceed with Gibbs sampling, the posterior for each of the model parameters needs to be cast into a conditional probability distribution form where each model parameter is conditioned upon the other model parameters and the timing residuals. In the following two subsections, we derive such conditional probabilities for parameters b and ρ.

A. Conditional probability of coefficients

For the coefficients b, the conditional probability can be found by rewriting the full posterior [i.e., the product of Eqs. (3)–(5)] while ignoring all factors not depending on b

---

[1]Refer to Table I and Sec. A for more details on the definitions of the quantities used throughout this paper.

[2]Note that the total number of frequency-bins is k, but there are two Fourier coefficients per each frequency-bin. Both $a_s^{\cos}$ and $a_s^{\sin}$ have the same variance parametrized by ρ_s^2. This is reflected in Eq. (4).

coefficients explicitly. In other words, all model parameters are treated as constants and only the b coefficients are allowed to vary:

$$\ln p(\boldsymbol{b}|\boldsymbol{\rho},\boldsymbol{r},\boldsymbol{n}) \simeq -\frac{1}{2}[(\boldsymbol{r}-T\boldsymbol{b})^T N^{-1}(\boldsymbol{r}-T\boldsymbol{b}) + \boldsymbol{b}^T B^{-1}\boldsymbol{b}]$$

$$= -\frac{1}{2}[\boldsymbol{b}^T(T^T N^{-1}T + B^{-1})\boldsymbol{b}]$$

$$- \frac{1}{2}[-2\boldsymbol{b}^T T^T N^{-1}\boldsymbol{r}]. \qquad (9)$$

The above equation suggests that the b-dependence of the probability $p(\boldsymbol{b}|\boldsymbol{\rho},\boldsymbol{r},\boldsymbol{n})$ is Gaussian. Using the *maximum a posteriori* estimate of b found by maximizing Eq. (9) as an estimate of the mean of the Gaussian, one can write the conditional probability distribution of the b coefficients, Eq. (9), in the form

$$p(\boldsymbol{b}|\boldsymbol{\rho},\boldsymbol{r},\boldsymbol{n}) = \frac{\exp\{-\frac{1}{2}(\hat{\boldsymbol{\mu}}-\boldsymbol{b})^T\Sigma(\hat{\boldsymbol{\mu}}-\boldsymbol{b})\}}{\sqrt{\det\{(2\pi)\Sigma^{-1}\}}}, \qquad (10)$$

where,[3]

$$\Sigma = T^T N^{-1}T + B^{-1}, \qquad (11)$$

$$\hat{\mu} = \Sigma^{-1}T^T N^{-1}\boldsymbol{r}. \qquad (12)$$

B. Conditional probability of red noise power-spectral-density

Similar to the b coefficients, the conditional probability of the ρ parameters can be found by taking advantage of the full posterior and ignoring all the factors not depending on ρ explicitly. Additionally, we make the observation such that the relevant probability distributions can be factorized over frequency-bins:

$$p(\boldsymbol{\rho}|\boldsymbol{a},\boldsymbol{r},\boldsymbol{n}) = \prod_{s=1}^{k} p(\rho_s|a_s,\boldsymbol{r},\boldsymbol{n})$$

$$= \prod_{s=1}^{k} \frac{1}{\rho_s\sqrt{(2\pi)\rho_s^2}}\exp\left\{-\frac{1}{2}\left(\frac{a_s\cdot a_s}{\rho_s}\right)\right\}$$

$$\propto \prod_{s=1}^{k} \frac{1}{\rho_s^2}\exp\left\{-\left(\frac{\frac{a_s\cdot a_s}{2}}{\rho_s}\right)\right\}$$

$$= \prod_{s=1}^{k} \mathrm{InvGamma}\left(\alpha = 1, \beta = \frac{a_s\cdot a_s}{2}\right). \qquad (13)$$

In the above, the dot-product denotes the sum of the square of the cosine and sine Fourier coefficients for each frequency-bin that is $a_s\cdot a_s = (a_s^{\mathrm{cos}})^2 + (a_s^{\mathrm{sin}})^2$. Furthermore, despite

---

[3]The definition of Σ in Eq. (11) is chosen so that this paper's Σ represents the same quantity as the Σ defined in the PTA GWB detection literature.

the analytic form for the dependence of ρ on the Fourier coefficients $\boldsymbol{a}$, the lower and the upper bounds of the inverse-gamma distribution extending to zero and infinity would lead to astrophysically and statistically incorrect assumptions as such bounds need to be finite and constrained to avoid the implicit use of improper priors in the modeling of red noise processes. Thus, a truncated version of the derived inverse-gamma distribution needs to be considered. In Sec. B, we show how to obtain such a truncated distribution.

C. Conditional probability of white noise parameters

In contrast to b and ρ, the white noise parameters cannot be written in terms of standard statistical distributions. This is mainly due to the dependence of the white noise parameters to various radio telescope receivers (i.e., each backend of each radio telescope needs its own white noise parameters). Solving the full-likelihood for the white noise parameters, collectively denoted by $\boldsymbol{n}$, results in

$$\ln p(\boldsymbol{n}|\boldsymbol{\rho},\boldsymbol{b},\boldsymbol{r}) = -\frac{1}{2}\sum_{i=1}^{p}\{(\boldsymbol{r}-T\boldsymbol{b})^T N^{-1}(\boldsymbol{r}-T\boldsymbol{b})\}$$

$$- \frac{1}{2}\sum_{i=1}^{p}\ln(\det\{2\pi N\}), \qquad (14)$$

where the sum is over the TOAs. Since Eq. (14) cannot be simplified further in any useful way, we have no choice but to utilize a non-Gibbs MCMC procedure to sample the posterior.

D. Standard method of single-pulsar analyses

The standard method of single-pulsar analyses involves an analytical marginalization of the product of Eqs. (3) and (4) over the coefficients b. The result is

$$p(\boldsymbol{r}|\boldsymbol{\rho}) = \frac{1}{\sqrt{\det\{(2\pi)C\}}}\exp\left\{-\frac{1}{2}\boldsymbol{r}^T C^{-1}\boldsymbol{r}\right\}, \qquad (15)$$

$$C = N + TBT^T, \qquad (16)$$

$$C^{-1} = N^{-1} - N^{-1}T\Sigma^{-1}T^T N^{-1}, \qquad (17)$$

where in the last line, we have used the *Woodbury identity*:

$$(X + UYV)^{-1} = X^{-1} - X^{-1}U(Y^{-1} + VX^{-1}U)^{-1}VX^{-1}, \qquad (18)$$

and Σ is defined in Eq. (11). The dependence of Eq. (15) on the red noise parameters ρ_k is through the elements of the matrix Σ^{-1}. Once Eq. (15) is multiplied by the appropriate priors of the model parameters, the resulting joint probability distribution of $p(\boldsymbol{\rho}|\boldsymbol{r})$ is ready to be given to a non-Gibbs MCMC algorithm for parameter estimation.

E. Gibbs sampling

Gibbs sampling [22] is a MCMC algorithm designed to take advantage of the conditional probability distributions of all model parameters in order to perform parameter estimation. It is often used in statistical inferences where a joint probability distribution of all parameters is difficult to sample, yet each model parameter's probability distribution can be written in terms of the rest of the parameters and the data. Gibbs sampling allows for random draws from the conditional probability distributions of model parameters whose analytic functional form must be found prior to the start of the sampling process as we have done for the case of single-pulsar noise analyses by deriving Eqs. (10) and (13). Due to the existence of analytic forms for the probabilities, the concept of rejection of random states, an integral part of the other MCMC algorithms, does not belong to the Gibbs sampling as all draws are considered accepted. Nevertheless, Gibbs sampling is still a MCMC algorithm as it possesses features such as no long-term-memory and the need for burn-in of the final Markovian chain. We will outline a step-by-step implementation of Gibbs sampling for a single-pulsar noise analysis in the remaining part of this section.

Knowing the conditional probabilities of our model parameters, ρ, b, and n, it is simple to implement Gibbs sampling in the following way:

(i) *Step 1*: Make initial guesses of ρ and n denoted by ρ_0 and n_0.

(ii) *Step 2*: Using Eq. (10), find an estimate of b_0 given ρ_0 and n_0.

(iii) *Step 3*: To start the first iteration, find an estimate of ρ_1 given b_0 and n_0 using Eq. (13).

(iv) *Step 4*: Continuing the first iteration, find an estimate of n_1 given b_0 and ρ_1 with a very short MCMC procedure sampling Eq. (14).

(v) *Step 5*: To end the first iteration, find an estimate of b_1 given ρ_1 and n_1 using Eq. (10).

Figure 1 provides an illustration of the explained procedure. The above steps can be repeated until all the model parameters reach satisfactory convergence. Due to the analytical draws of the ρ and the b coefficients, convergence will be reached quickly compared to the fully non-Gibbs MCMC algorithms. This is one of the most desirable features of Gibbs sampling as the overall run-time of the PTA single-pulsar noise analyses will be reduced significantly.

F. Frequency domain multipulsar likelihood

The outlined Gibbs sampling procedure is an efficient Bayesian scheme capable of estimating each pulsar's power-spectral-density as well as the Fourier coefficients required to describe the total red noise (i.e., GWB plus spatially-uncorrelated intrinsic red noise process) component of the timing residuals. However, the information required in characterizing a GWB requires subsequent multipulsar analyses. As will be demonstrated in this
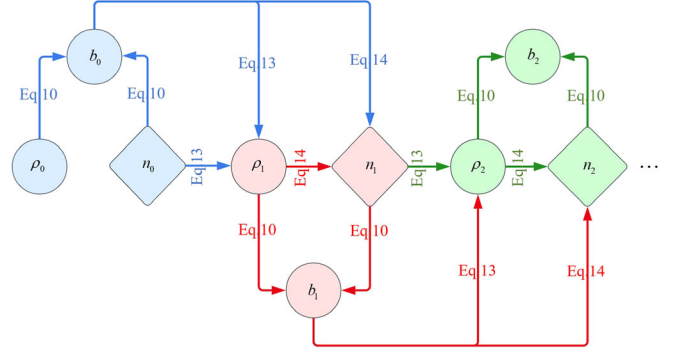

FIG. 1. A schematic representation of the first three steps of the outlined Gibbs sampling procedure. The first step of the sampling process (blue) starts by guesses of the ρ and the white noise parameters and results in an estimate of the coefficients b following Equation (10) using the previously guessed values. The second (red) and the third (green) steps of the sampling continue the sequence by estimating the next remaining model parameter given the most recent estimates of the other two parameters using the conditional probability distributions of Eqs. (10), (13), and (14).

section, the output of Gibbs sampling provides enough information to perform multipulsar analyses aiming at detecting a GWB.

Using only the Fourier coefficients a, one can construct a factorized likelihood in the frequency domain in the following way:

$$p(A|\Phi) = \frac{\exp\left\{-\frac{1}{2}\left(a_{k;I}^T \Phi_{ks}^{-1} a_{s;J}\right)\right\}}{\sqrt{\det\{(2\pi)\Phi\}}} \tag{19}$$

$$= \prod_s \frac{\exp\left\{-\frac{1}{2}\left(a_{s;I}^T \Phi_{ss}^{-1} a_{s;J}\right)\right\}}{\sqrt{\det\{(2\pi)\Phi_{ss}\}}}, \tag{20}$$

for

$$\Phi = \begin{bmatrix} \Phi_{11} & 0 & \cdots & 0 \\ 0 & \Phi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_{kk} \end{bmatrix}, \tag{21}$$

$$\Phi_{kk} = \begin{bmatrix} \varphi_{k,1} & \Gamma_{k;12}\rho_{k;g}^2 & \cdots & \Gamma_{k;1m}\rho_{k;g}^2 \\ \Gamma_{k;21}\rho_{k;g}^2 & \varphi_{k,2} & \cdots & \Gamma_{k;2m}\rho_{k;g}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{k;m1}\rho_{k;g}^2 & \Gamma_{k;m2}\rho_{k;g}^2 & \cdots & \varphi_{k,m} \end{bmatrix}, \tag{22}$$

where A denotes the collection of Fourier coefficients, across all pulsars and all frequencies, (i.e., $A = \{a_{k;m}\}$),

$\rho_{k;g}$ parametrizes the common power-spectral-density of the GWB (indicated by the subscript g) observed across the entire pulsar array at frequency k, and Γ_{IJ} represents the functional form of the cross correlations (e.g., Hellings and Downs curve).

One can use this equation to derive an optimal estimator of the signal-to-noise analogous to those presented in Anholm *et al.* [21] and Chamberlin *et al.* [23]. We leave the details of the derivation to our future project [24] where we explore the use of the Fourier coefficients in GWB characterization in great detail. Here, we simply report the results in the form of the optimal estimators of the cross-correlations λ_{IJ} and their uncertainty σ_{IJ}:

$$\lambda_{IJ} = \frac{\sum_s a_{s;I} \cdot a_{s;J} \frac{\hat{P}_g}{\varphi_{s;I}\varphi_{s;J}}}{\sum_s \frac{\hat{P}_g^2}{\varphi_{s;I}\varphi_{s;J}}}, \qquad (23)$$

$$\sigma_{IJ} = \left[\sum_s \frac{\hat{P}_g^2}{\varphi_{s;I}\varphi_{s;J}} \right]^{-\frac{1}{2}}. \qquad (24)$$

Without a need for a detailed derivation, Eqs. (23) and (24) can be understood by following a very simple rational. The numerator is the weighted product of $a_{s;I} \cdot a_{s;J}$. The weights associated with such product, $1/\varphi_I$ and $1/\varphi_J$, have the role of suppressing the contributions from pulsars whose total non-GWB noise power is substantial (i.e., dominant spatially-uncorrelated intrinsic red noise). Moreover, the choice for the normalization in the denominator ensures that the estimated correlations would yield $A_g^2 \Gamma_{IJ}$ if averaged over many realizations of GWB as is shown in Sec. C.

Additionally, estimates of the amplitude, the uncertainty of the estimated amplitude, and the signal-to-noise-ratio (SNR) can be made from Eqs. (23) and (24) by minimizing a weighted-chi-squared statistic of the form

$$\chi^2 = \sum_{IJ} \frac{(\lambda_{IJ} - A_g^2 \Gamma_{IJ})^2}{\sigma_{IJ}^2}, \qquad (25)$$

with respect to A_g^2 which results in

$$\hat{A}_g^2 = \frac{\sum_{IJ;I \neq J} \sum_s a_{s;I} \cdot a_{s;J} \Gamma_{IJ} \frac{\hat{P}_g}{\varphi_{s;I}\varphi_{J;s}}}{\sum_{IJ;I \neq J} \sum_s \Gamma_{IJ}^2 \frac{\hat{P}_g^2}{\varphi_{I;s}\varphi_{s;J}}}, \qquad (26)$$

$$\sigma_g = \left[\sum_{IJ;I \neq J} \sum_s \Gamma_{IJ}^2 \frac{\hat{P}_g^2}{\varphi_{I;s}\varphi_{J;s}} \right]^{-\frac{1}{2}}, \qquad (27)$$

$$\text{SNR} = \frac{\hat{A}_g^2}{\sigma_g}. \qquad (28)$$

When estimating the optimal correlations using Eq. (23), one has a few options to select from for the choice of a_I

and a_J. The trivial option is to draw randomly from the multivariate probability distribution of each pulsar's a (the output of Gibbs sampling) and obtain the cross product of such random draws for each pulsar pair. Another option is to construct posteriors of the mean, $\hat{\mu}$, following Eq. (10), and draw randomly from such posteriors. Similar to the previous case, the cross product of the random draws can be used in Eq. (23) and with the difference that the normalization factor in the denominator of Eq. (23) should be recalculated (see Appendix C for more details). Lastly, for the choice of φ_I, we use the total red noise power P_I.

As a final note, it is important to recognize the limitations of the presented technique as well as the *optimal statistic* in general. In practice, optimal statistic results in biased estimates of the GWB amplitude and the signal-to-noise ratio if one does not have separate estimates for the spatially-uncorrelated as well as the common red noise power. In other words, if one uses the red noise power estimates from the single-pulsar analyses instead of obtaining separate estimates for a common red noise signal and intrinsic red noise signal, one cannot characterize a common correlated signal correctly. This has been explored in depth in [25].

III. ANALYSIS OF THE NANOGrav 11 YEAR DATASET

To test the capabilities of the outlined single-pulsar data analysis technique, we analyze the NANOGrav 11 year dataset [20] using Gibbs sampling. The results are then compared to the ones obtained via standard Bayesian modeling detection routine used by the NANOGrav collaboration in their most recent work [4]. To ensure the fairness of the convergence comparisons, we allow each technique to sample the dataset for two hours for each pulsar. After the two hours time-limit, we compare the posteriors' effective-sample-size (ESS) and rank-normalized-split R-hat ($\hat{r}$) values using the diagnostic tools provided by Kumar *et al.* [26].

A. Details of the Bayesian modeling

The Gibbs sampling implementation used for the 11 year dataset models the data as outlined in Sec. II. This Bayesian modeling together with Gibbs sampling is referred to as Gibbs Method (GM) from hereon. Moreover, the competing method of analyzing the NANOGrav 11 year dataset follows the standard single-pulsar analyses currently implemented in the most recent GWB searches [8–10] and explained in Sec. II D. The PTMCMC sampling package [27] as well as the structure of the Bayesian modeling accompanying this sampling is referred to as Standard Method (SM) from hereon.

For both SM and GM, we have allowed each pulsar's set of red noise parameters, ρ, to follow a 30 frequency free-power-spectral-density model with frequencies ranging
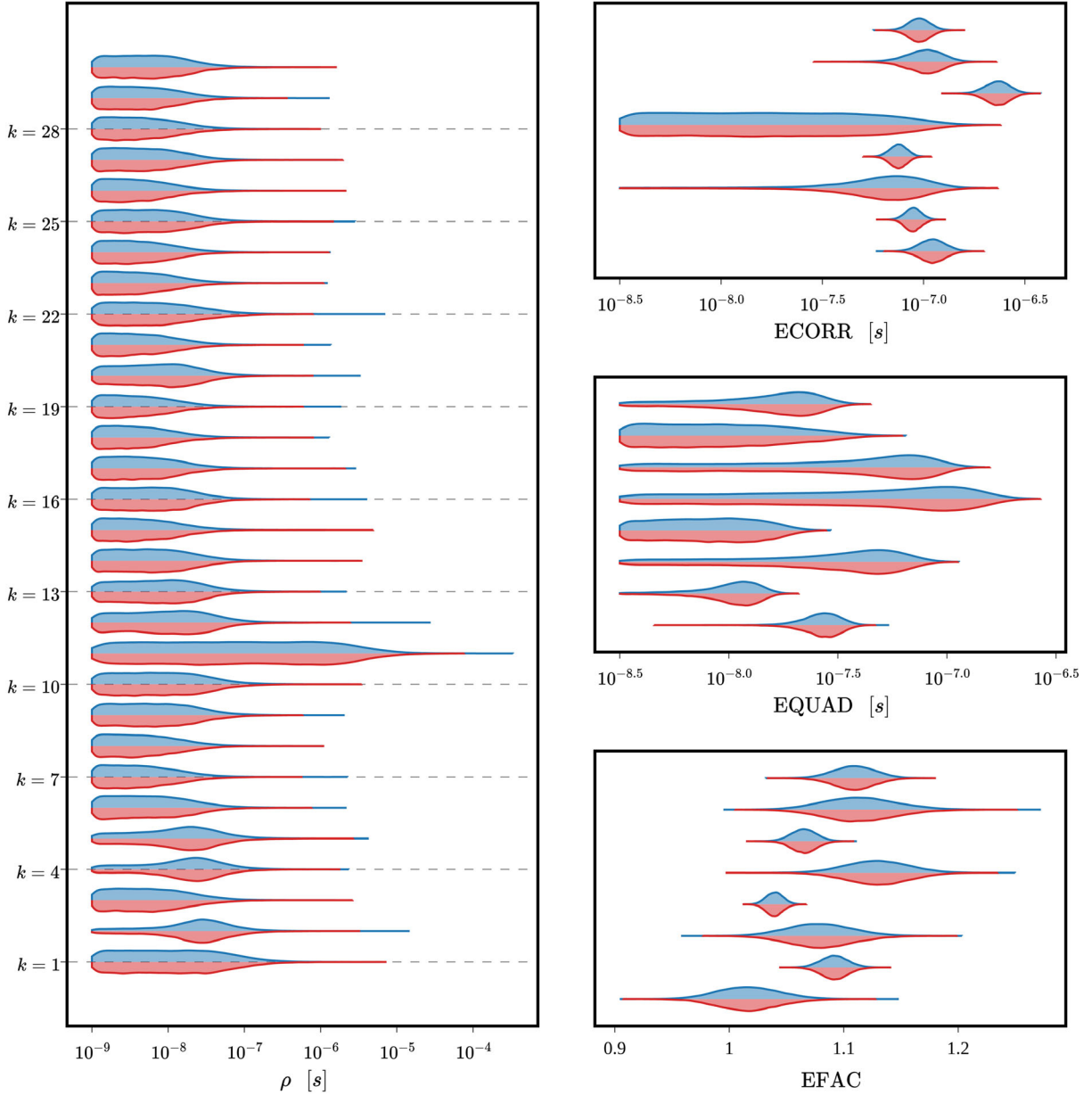
FIG. 2.   A comparison of posteriors for all model parameters of PSR J1713 + 0747 for the NANOGrav 11 year dataset obtained via GM (blue) and SM (red). The posteriors on the left column belong to the red noise model parameters, collectively referred to as $\boldsymbol{\rho}$, whereas the posteriors on the right column belong to the white noise parameters EFAC, EQUAD, and ECORR. There is one ρ_k parameter for each frequency ($k = 30$ frequencies in total) and three white noise parameters for each receiver (8 receivers in total). To obtain the plots via GM, 30 steps of a Metropolis Hasting algorithm within each step of Gibbs sampling has been implemented for the white noise parameters. The above plots show a great level of consistency in extracting the posteriors between the two methods.

from $1/T_{\mathrm{obs}}$ to $30/T_{\mathrm{obs}}$ in which T_{obs} denotes the observational baseline of each considered pulsar. The choice of prior for the model parameters are listed below. For each pulsar, the white noise parameters are per receiver/backend system while the $\boldsymbol{\rho}$ parameters are per frequency:

$$\boldsymbol{\rho}[s] \sim \text{log-Uniform}(-9, -4), \qquad (29)$$

$$\text{EQUAD}[s] \sim \text{log-Uniform}(-8.5, -5), \qquad (30)$$

$$\text{ECORR}[s] \sim \text{log-Uniform}(-8.5, -5), \qquad (31)$$

$$\text{EFAC} \sim \text{Uniform}(0.01, 10), \qquad (32)$$

for $[s]$ denoting the unit of the quantities, which is seconds.

B. Comparison of posteriors

For the sake of brevity, out of the thirty four pulsars of the NANOGrav 11 year dataset, we have chosen to feature
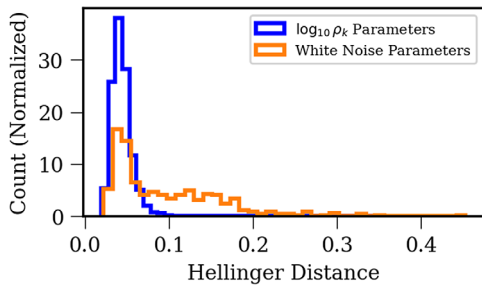
FIG. 3. A histogram showcasing the distribution of the Hellinger distance values for the $\log_{10}\rho$ posteriors (blue) and the white noise parameters (orange) obtained by comparing the outputs of GM and SM. The histogram contains the Hellinger distances of model parameters across all frequencies and pulsars. As evident by the distribution, GM and SM result in sufficiently similar distributions with a few exceptions whose inconsistencies can be attributed to the differences in the level of convergence of posteriors resulting from GM and SM even though we have allowed sufficient time for SM to converge (i.e., more than two hours). GM posteriors follow the general shape of SM posteriors but are significantly more converged.

a GM vs SM posterior comparison plot for only PSR J1713 + 0747 as this pulsar has the longest observational baseline as well as the largest number of TOAs making it the most computationally expensive pulsar to analyze. As shown in Fig. 2, the two techniques yield consistent posteriors for both the red noise and the white noise model parameters for PSR J1713 + 0747 showcasing the robustness and the capability of GM to be implemented on real PTA datasets. The same consistency is also observed in all the remaining thirty three pulsars. For a quantification of the degree of consistency between the two sets of posteriors, refer to Fig. 3 which highlights the differences in the output of GM and SM in the form of a histogram of *Hellinger distance* [28][4] values across all pulsars. With the exception of a few white noise parameters, the Hellinger distances are concentrated between 0 and 0.2 indicating an adequate degree of consistency between the GM and the SM posteriors. We attribute the higher Hellinger distance values of some model parameters (especially the white noise parameters) to the differences in the level of convergence of the posteriors as GM is more successful at yielding converged posteriors than SM. Refer to Sec. III D for a more detailed discussion.

---

[4]Hellinger distance is a measure of similarity between two probability distributions ranging from 0 (identical distributions) to 1 (disagreeing distributions). For two discrete probability distributions p and q, the Hellinger distance H is defined as $H = \frac{1}{\sqrt{2}}\sqrt{\sum_i \left(\sqrt{p_i} - \sqrt{q_i}\right)^2}$, where i ranges over the binned quantities of interest whose probability distribution is described by p and q.
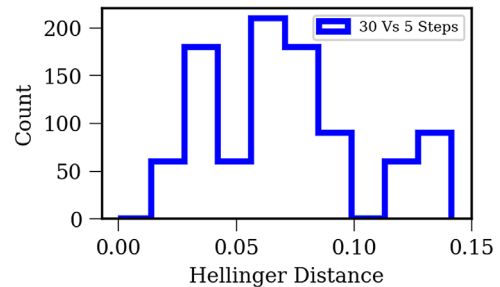

FIG. 4. A histogram showcasing the distribution of the Hellinger distance values between the $\log_{10}\rho$ posteriors obtained using GM with two different number of MCMC steps (30 and 5 steps) for each step of Gibbs sampling for the white noise parameters. The figure is made by combining the Hellinger distance values across all of the pulsars and all of the frequencies. As evident by the distribution, choosing a much lower number of MCMC steps for each step of Gibbs sampling for the white noise parameters does not change the shape of the target $\log_{10}\rho$ posteriors significantly. This effect can be attributed to the knowledge of GM about the analytical shape of the $\log_{10}\rho$ parameters prior to the start of the sampling.

C. The effect of using different number of MCMC steps in GM

To obtain the white noise posteriors of Fig. 2, 30 steps of a Metropolis Hasting algorithm for each step of Gibbs sampling has been implemented. The choice for the number of MCMC steps for each step of the Gibbs sampling depends on factors such as the number of TOAs, one's threshold and preferred measure of convergence for the posteriors as well as the efficiency of the type of MCMC algorithm used in the white noise parameter estimation. However, the red noise parameters' posteriors are not overly sensitive to this choice as the target distributions for ρ_k parameters are all analytically determined prior to the start of sampling. To test the sensitivity of the red noise parameters to the choice for the number of MCMC steps for each step of Gibbs sampling, we have applied GM on all of the NANOGrav 11 year pulsars using only 5 steps of MCMC. As shown in Fig. 4, the estimated Hellinger distance values between the two sets of posteriors of $\log_{10}\rho$ parameters are sufficiency low suggesting a weak degree of correlation between the red noise parameters' posteriors to the white noise parameters' if analyzed via GM. Nevertheless, our current implementation of GM is adequately optimized to handle large number of MCMC steps without much of a sacrifice in the overall run-time of a single-pulsar analysis.

D. Comparison of convergence levels

Despite resulting in consistent posteriors, SM and GM differ significantly in their state of convergence of the model parameters, especially those pertaining to the effective-sample-size (ESS). Figure 5 shows the spread of the ratio of ESS values (GM divided by SM) across all of the
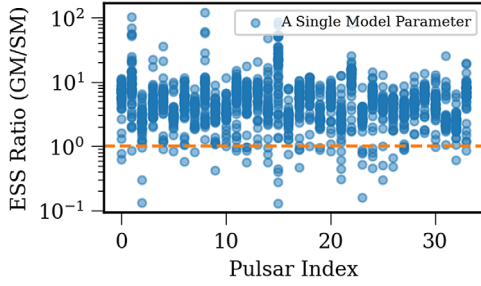
FIG. 5.   A scatter-plot showcasing the differences in the spread of the effective-sample-size (ESS) values for the $\log_{10}\rho$ and the white noise parameters expressed in the form of the ratio of GMs' ESS over SMs' ESS (blue circles). For each pulsar, there is one $\log_{10}\rho_k$ for each frequency (30 frequencies in total) and three white noise parameters for each receiver. Across all of the pulsars, GM is more capable at yielding posteriors with significantly higher ESS levels given the two hour time limit. Considering all model parameters, the average ESS ratio is 6. The values of ESS are found using the functionalities provided in Kumar *et al.* [26].
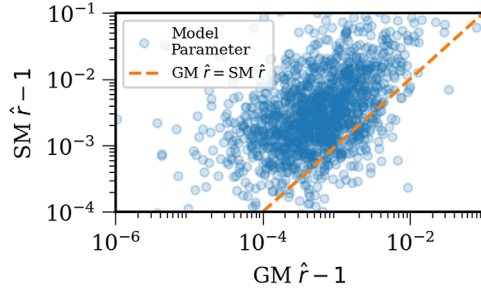

FIG. 6.   A scatter-plot showcasing the differences in the spread of rank-normalized-split-R-hat, $\hat{r}$, values obtained from the $\log_{10}\rho$ and the white noise parameters analyzed by GM and SM. Each blue circle represents a single model parameter and the figure is obtained by combining the $\hat{r}-1$ values of all model parameters for all pulsars and across all frequencies. As evident by the figure, GM is more capable at resulting in posteriors with a lower $\hat{r}$ level given the two hour time limit. The $\hat{r}$ values are estimated by dividing each Markov chain into two sub-chains and applying the rank-normalized-split-R-hat test [26] on it.

model parameters for every pulsar. As evident by Fig. 5, a significant majority of each pulsar's model parameters have higher ESS values when analyzed using GM as compared to SM. The average ESS ratio across all parameters and pulsars is 6. Figure 5 proves our claim about the high efficiency of GM. Additionally, the same observation can be made about the rank-normalized-split R-hat ($\hat{r}$) values calculated for both GM and SM posteriors for each pulsar. Figure 6 points toward the higher state of convergence of a significant majority of the model parameters that were analyzed by GM.

IV. SIMULATIONS

Despite the successful implementation of GM on the NANOGrav 11 year dataset, we have not tried to analyze

the correlation content of the dataset using the concepts discussed in Sec. II F as the 11 year dataset lacks a common correlated signal across pulsar pairs [20]. For studying the correlations, we will dedicate future projects to the analysis of the NANOGrav 15 year [8] and the upcoming IPTA's DR3 datasets. Meanwhile, to explore the capability of the Fourier coefficients a in characterizing a common spatially correlated signal, we make use of simulated PTA datasets.

A. Details of the simulations

We have chosen two types of simulated datasets, referred to as SIM0 and SIM1, with 300 realizations for each type, to analyze in order to explore the capability of the Fourier coefficients a to characterize a common correlated signal. The two simulated datasets are identical in every aspect except the content of their spatially-uncorrelated intrinsic red noise: for SIM0, the log of the amplitude of the spatially-uncorrelated intrinsic red noise of each pulsar is randomly chosen from a uniform distribution between 10^{-16} and 10^{-14} while for SIM1 this range is between 10^{-14} and 10^{-13}. For both datasets' pulsars, the spectral index of the spatially-uncorrelated intrinsic red noise follows a uniform distribution with lower and upper bounds of 0 and 7 respectively. Additionally, each dataset has 90 pulsars uniformly scattered across the sky timed for 20 years with random timing cadences between 14 to 30 days. Furthermore, each dataset contains 10 microseconds of white Gaussian noise for each pulsar as well as a unique realization of a GWB with amplitude of $A_g = 2 \times 10^{-15}$ and spectral index of $\gamma_g = 13/3$. Lastly, to employ GM on each dataset, we keep the white noise parameters constant and use the same range of frequency-bins for all pulsars which is $\{1/20 \text{ yrs}, 2/20 \text{ yrs}, 3/20 \text{ yrs}, 4/20 \text{ yrs}, 5/20 \text{ yrs}\}$.

It is worth mentioning that our intention is not about simulating realistic datasets and analyzing it with GM. We have already shown the capability of GM in single-pulsar analyses of real datasets. Our intention is to highlight what the Fourier coefficients can potentially reveal about an existing GWB signal, hence the reason behind our choices for the specific parameters of the two simulated datasets. Nonetheless, we have introduced very high levels of spatially-uncorrelated intrinsic red noise in the SIM1 dataset (higher than what is observed in the real PTA datasets) as dealing with such processes is an extremely challenging part of GWB searches using PTAs whose impact on the correlation recovery using the Fourier coefficients is nontrivial.

B. Reconstruction of red noise signal using Fourier coefficients

The a coefficients are capable of reconstructing the red component of the timing residulas as suggested by Equation (1). The reconstructed signal is *pre-fit* and
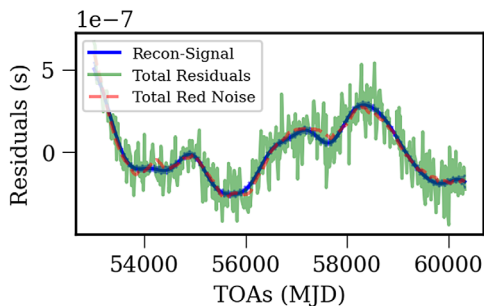
FIG. 7. A comparison between postfit time series reconstruction using the Fourier coefficients obtained from GM (blue), the injected red noise time series (red), and the total residuals (green) for one of SIM0's pulsars. The reconstructed residuals are made by considering the entire posterior probability distribution of the recovered Fourier coefficients. As evident by the figure, the reconstructed post-fit red noise signal matches the underlying red noise signal closely.

white-noise-free. Once the reconstructed signal obtained by Fa is fitted for the timing model parameters, it mirrors the underlying total *post-fit* red noise signal in the dataset closely. Figure 7 highlights this case for one of SIM0's pulsars. As suggested by the figure, the Fourier coefficients are capable of reconstructing the underlying red noise process of the total timing residuals. This fact allows the Fourier coefficients to be adequate replacement for the timing residuals in the frequency domain with the added benefit that one no longer needs to take into account a white noise process or be concerned with the complications of the timing model parameters when using the a coefficients in a subsequent analysis. In fact, the effects of the timing model parameters and the white noise levels are implicit in the posteriors for the Fourier coefficients obtained via GM.

C. Searching for correlations using Fourier coefficients

To characterize the GWB signal in each of the realizations of SIM0 and SIM1, we use Eqs. (23) and (24) with $\hat{\mu}_I$ as the quantity representing the Fourier coefficients of each
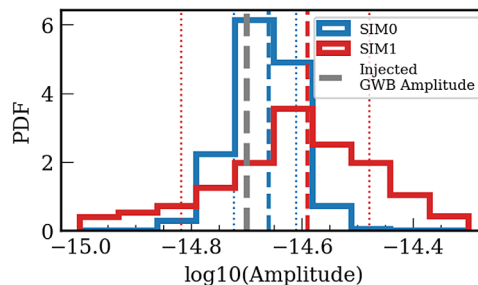

FIG. 9. Two histograms comparing the distributions of the recovered common correlated signal between SIM0 (blue) and SIM1 (red) dataset using the method provided in Sec. II F. The blue and the red vertical lines indicate the 16th and the 84th percentiles (dotted lines) as well as the mean (dashed line) of each distribution. Each distribution is obtained by combining the estimates of the amplitude [Eq. (26)] of the cross-correlated signal over 300 realizations. The injected GWB signal is indicated with a vertical dashed gray line. The figure suggests that the Fourier coefficients contain the right amount of information about the amplitude of the cross-correlated signal in the case of SIM0. In the case of SIM1, due to the existence of significantly higher non-GWB red noise power, the recovered GWB amplitude is more scattered.

pulsar [see Eq. (10)]. Furthermore, since our goal is to showcase the potential of the Fourier coefficients in revealing information about the GWB signal rather than outlining a complete and practical pipeline capable of fully characterizing a GWB signal, the weights φ_I are set to the total red noise power that was used to generate the simulated datasets.

The shape of the correlation recovery is depicted in Fig. 8 for both simulated datasets. This shape is obtained by dividing the pulsar pairs of each realization into 15 different angular separation bins such that all bins have 267 pulsars pairs in them. Additionally, the average, 16th, and 84th percentiles (over the 300 realizations) of the correlations for each angular separation bin is computed and indicated in Fig. 8. Furthermore, the histogram of the estimated
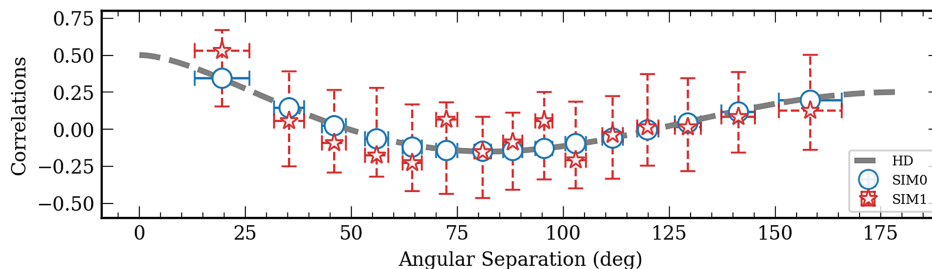

FIG. 8. A plot depicting the reconstruction of the Hellings and Downs correlation (gray dashed curve) using GM's estimates of the Fourier coefficients obtained for both SIM0 (blue circles) and SIM1 (red stars) datasets. The reconstructions are the average over 300 realizations of both datasets. The error-bar of each point indicates the range between the 16th and the 84th percentiles over the 300 realizations. Remarkably, the recovery of the shape of the correlations is not affected significantly by the introduction of extremely high levels of spatially-uncorrelated red noise to each pulsar in SIM1.
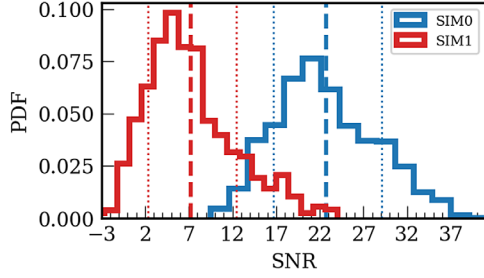
FIG. 10. Two histograms comparing the distributions of the signal-to-noise ratio (SNR) between SIM0 (blue) and SIM1 (red) dataset using the method provided in Sec. II F. The blue and the red vertical lines indicate the 16th and the 84th percentiles (dotted lines) as well as the mean (dashed line) of each distribution. Each distribution is obtained by combining the estimates of the SNR (Equation (28) of the cross-correlated signal over 300 realizations. As expected, SIM1 dataset exhibits a lower SNR due to containing a significantly higher non-GWB red noise power than the GWB power.

amplitude $\hat{A}_g$ and signal-to-noise ratio of all the 300 realizations of each dataset are stacked on top of each other (i.e., no averaging is performed) and presented in Figs. 9 and 10 respectively. The impact of introducing extreme levels of intrinsic spatially-uncorrelated red noise to the dataset manifests itself in the form of lowering the signal-to-noise ratio and more scattered amplitude recovery. However, the shape of the correlations recovery remains remarkably close to the Hellings and Downs curve over many realizations.

V. DISCUSSION AND FUTURE WORK

In this paper, we have shown that the Gibbs method (GM) is an efficient single pulsar Bayesian noise analysis technique capable of producing posteriors for the single-pulsar free-power-spectral-density and the white noise model parameters with convergence properties that are superior to those obtained using standard Bayesian methods (SM). GM is a robust and computationally efficient alternative to SM for future PTA noise analyses. Additionally, we have shown that the Fourier coefficients resulting directly from GM contain adequate information about the shape the cross-correlations signal through the use of simulations. In effect, GM produces the frequency domain representation of each pulsar's red noise signal, free of white noise and timing model parameters, hence providing all the necessary information to start performing subsequent GWB detection analyses exclusively in the frequency domain.

GM results in raw information in the frequency domain which may need to be processed further depending on the needs of the subsequent analyses. For instance, the astrophysical interpretation of a pulsar's red noise signal will require a more constrained model of the power spectral density than the free-spectrum model which could be achieved by fitting for the parameters of such a model

using the output of GM (e.g., a power-law fit to the free-spectrum model) [19]. Combined with the fitting utilities provided by Lamb *et al.* [19], GM can become a powerful and efficient tool for use in the future PTA GWB detection analyses.

A. Software

The GM code takes advantage of the functionalities provided by ENTERPRISE [29] and ENTERPRISE-extensions [30], and PTMCMC sampler [27]. The package ArviZ [26] has been used for diagnosing MCMC chains. Python packages Matplotlib [31] and plotly [32] have been used for generating the figures in this paper.

APPENDIX A: GWB DETECTION TERMINOLOGY

Most of the PTA noise analysis concepts have been developed over many years and scattered over many papers [12–14,21,23,33–37]. To help readers better understand the methods used in this paper, we define the necessary PTA noise analysis quantities and concepts in this section. Additionally, refer to Table I for a short description of the mathematical symbols used throughout this paper.

1. Basis matrices and their coefficients

To model the contribution of any red noise process to the timing model residuals of a given pulsar, $\boldsymbol{r}_{\text{Red}}$, we employ a Fourier basis matrix and a vector of coefficients such that

$$\boldsymbol{r}_{\text{Red}} = F\boldsymbol{a}, \tag{A1}$$

TABLE I. A table listing the symbols most commonly used throughout this paper and a short description of what they represent. Refer to Sec. A for more details on the definition of some of the quantities.

Symbol	Description
$T_{\rm obs}$	Observational baseline
t	Time
f	Frequency
I, J	Pulsar indices
k, s	Indices for the frequency bins
m	Index for the number of pulsars in the array
p	Number of TOAs for a given pulsar
q	Number of timing model parameters
$\boldsymbol{r}$	Timing residual
F	Fourier design matrix
M	Timing model design matrix
T	Combination of F and M such that $T = [M, F]$
N	White noise covariance matrix
B	Covariance matrix for the linear timing model parameters and the Fourier coefficients (i.e., $\langle \boldsymbol{bb}^T \rangle$)
Γ	Hellings and Downs cross correlation matrix
$\boldsymbol{A}$	Collection of fourier coefficients across pulsars and frequencies $\{a_{k;m}\}$,
φ	Single-pulsar red process covariance matrix
Φ	Multi-pulsar red process covariance matrix
$\boldsymbol{a}$	Fourier sin-cos coefficients
λ	Estimated cross correlations
$\boldsymbol{\epsilon}$	Linear timing model parameters
$\boldsymbol{b}$	Combination of $\boldsymbol{a}$ and $\boldsymbol{\epsilon}$ such that $\boldsymbol{b}^T = [\boldsymbol{\epsilon}, \boldsymbol{a}]$
$\boldsymbol{w}$	White noise time series
$\boldsymbol{n}$	Collection of all white noise parameters
ρ	Free-spectrum parameter used in describing power-spectral-density ($\rho^2 = \langle \boldsymbol{aa}^T \rangle$)
A	Amplitude of a red noise process
P	Power-spectral-density of a red noise process
$\hat{P}$	Spectral shape of a red noise process obtained by P/A^2

$$
F = \begin{pmatrix}
\sin(2\pi f_1 t_1) & \cos(2\pi f_1 t_1) & \cdots & \sin(2\pi f_k t_1) & \cos(2\pi f_k t_1) \\
\sin(2\pi f_1 t_2) & \cos(2\pi f_1 t_2) & \cdots & \sin(2\pi f_k t_2) & \cos(2\pi f_k t_2) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\sin(2\pi f_1 t_p) & \cos(2\pi f_1 t_p) & \cdots & \sin(2\pi f_k t_p) & \cos(2\pi f_k t_p)
\end{pmatrix},
\tag{A2}
$$

$$
\boldsymbol{a}^T = (a_1^{\rm sin}, a_1^{\rm cos}, \ldots, a_k^{\rm sin}, a_k^{\rm cos}),
\tag{A3}
$$

for t_p denoting the last measured TOA, f_k denoting the kth considered frequency-bin, and $a^{\rm sin}$ and $a^{\rm cos}$ referring to the coefficients of sin and cos elements of the F matrix respectively.

To model the contribution of any linear timing model parameter to the timing residuals, $\boldsymbol{r}_{\rm T}$, we use a basis matrix known as the *timing-design-matrix* such that

$$
\boldsymbol{r}_T = M\boldsymbol{\epsilon},
\tag{A4}
$$

$$
M = \begin{pmatrix}
1 & t_1 & t_1^2 & \cdots \\
1 & t_2 & t_2^2 & \cdots \\
\vdots & \vdots & \vdots & \cdots \\
1 & t_p & t_p^2 & \cdots
\end{pmatrix}.
\tag{A5}
$$

While the first three columns of the design matrix models the quadratic spin down of all millisecond pulsars, the unspecified columns of the matrix are populated with various timing model contributions specific to each pulsar. Moreover, it is often convenient to project the residuals

onto a subspace orthogonal to the timing model parameters, or in other words, to create *fitted* timing residuals. The so-called G matrix is a useful matrix obtained via singular-value-decomposition of the design-matrix constructed to perform the fitting:

$$M = USV^T$$
$$G_{xy} = U_{xy}, \tag{A6}$$

where x ranges from 1 to p (the number of TOAs) while y ranges from q to p for q being the total number of the linear timing model parameters.

To model the contribution of the white noise to the timing model residuals, r_w, we consider a $m \times m$ identity matrix as the basis with the coefficients n such that

$$r_w = w, \tag{A7}$$

$$w_i \sim \text{Normal} \left(\text{mean} = 0, \text{scale} = \sigma_{w_i} \right), \tag{A8}$$

$$\sigma_{w_i} = \text{ef}_i \sqrt{\sigma_i^2 + \text{eq}_i^2}, \tag{A9}$$

for σ_i being the TOA error of observation i, and ef and eq being the usual EFAC and EQUAD parameters [20]. Note that the Gaussianity of the white noise is an assumption included in our all of our models.

2. Noise power-spectral-density modeling

In this paper, we only consider one-sided power-spectral-densities (PSD). Most commonly for PTA noise analyses, the PSD is expressed in two ways:

(i) *Power law*: assuming the PSD to follow a simple power-law relation with amplitude A and spectral index γ as well as a reference frequency f_{ref} across all frequency-bins

$$P(f) = \frac{A^2}{12\pi^2 f^3} \left(\frac{f}{f_{\text{ref}}} \right)^{3-\gamma}, \tag{A10}$$

$$\hat{P}(f) = \frac{P(f)}{A^2}. \tag{A11}$$

The quantity $\hat{P}$ describes the shape of the spectrum and is used in Sec. IV C.

(ii) *Free-spectrum*: allowing the PSD to have independent amplitude in each frequency-bin with normalization constant T_{obs} equal to a fixed observation time. The observation time can either be the baseline of each pulsar or the baseline of the total PTA experiment.

$$P(f_k) = T_{\text{obs}} \rho_k^2. \tag{A12}$$

3. Covariance matrices

The white noise covariance matrix N plays a key role in posterior probability calculation of all model parameters. This matrix is modeled as

$$N = \text{diag}(\sigma_{w_1}, ..., \sigma_{w_p}). \tag{A13}$$

Note that the introduction of ECORR white noise parameter will complicate this picture. See chapter 7 of Taylor [35] for more details. Furthermore, the red process covariance matrix is obtained via the discretized form of the Wiener-Khinchin theorem

$$\langle r_{\text{red}}(t_i) r_{\text{red}}(t_j) \rangle = [F \varphi F^T]_{ij}, \tag{A14}$$

$$\varphi = \text{diag}\{P_{\text{red}}(f_1), P_{\text{red}}(f_1),$$
$$..., P_{\text{red}}(f_k), P_{\text{red}}(f_k)\}, \tag{A15}$$

where P_{red} is the one-sided PSD of a red noise process and the diagonal matrix φ is the matrix representation of that PSD.

APPENDIX B: TRUNCATED INVERSE-GAMMA DISTRIBUTION

To obtain a truncated inverse-gamma distribution, we take advantage of *inverse-transform sampling* method. However, first, we need to find a normalization factor, Norm, for the truncated inverse-gamma distribution defined between the lower bound ρ_{min} and the upper bound ρ_{max}:

$$\beta_k = \frac{(a_k \cdot a_k)}{2}, \tag{B1}$$

$$\text{Norm} = \left[\int_{\rho_{\text{min}}}^{\rho_{\text{max}}} d\rho_k \left\{ \frac{\beta_k}{\rho_k^2} \exp\left(-\frac{\beta_k}{\rho_k} \right) \right\} \right]^{-1}$$
$$= \frac{\beta_k}{\exp\left(-\frac{\beta_k}{\rho_{\text{max}}} \right) - \exp\left(-\frac{\beta_k}{\rho_{\text{min}}} \right)}. \tag{B2}$$

Note that the above process can be repeated for all frequency bins. Equation (B2) allows for calculation of the cumulative distribution function (CDF), which in turn can be used to find a distribution for ρ_k given a uniform random number U defined between 0 and 1 based on inverse-transform sampling method. This yields the following as the target distribution for the ρ_k parameters:

$$p(\rho_k|a_k, r, n) = -\frac{\beta_k}{\ln\left\{\exp\left(-\frac{\beta_k}{\rho_{\min}}\right)U(0,1)\left[\exp\left(-\frac{\beta_k}{\rho_{\max}}\right) - \exp\left(-\frac{\beta_k}{\rho_{\min}}\right)\right]\right\}}. \tag{B3}$$

APPENDIX C: DERIVATION OF THE NORMALIZATION FACTOR IN EQ. (23)

The choice of normalization in the denominator of Eq. (23) enforces the condition that the estimated cross correlations must yield GWB amplitude if averaged over many realizations as is shown below:

$$\lambda_{IJ} = \frac{p_{\text{top}}}{\text{Norm}}, \tag{C1}$$

$$p_{\text{top}} = \sum_k \frac{a_I}{\varphi_I}\frac{a_J}{\varphi_J}\hat{P}_g, \tag{C2}$$

$$\langle p_{\text{top}}\rangle = \left\langle \sum_k \frac{a_I}{\varphi_I}\frac{a_J}{\varphi_J}\hat{P}_g \right\rangle \tag{C3}$$

$$= \sum_k \left\langle \frac{a_I}{\varphi_I}\frac{a_J}{\varphi_J} \right\rangle \hat{P}_g \tag{C4}$$

$$= \sum_k \frac{\langle a_I a_J \rangle \hat{P}_g}{\varphi_I \varphi_J} \tag{C5}$$

$$= \sum_k \frac{(\Gamma_{IJ}A_g^2\hat{P}_g)\hat{P}_g}{\varphi_I\varphi_J} \tag{C6}$$

$$= \Gamma_{IJ}A_g^2\sum_k \frac{\hat{P}_g^2}{\varphi_I\varphi_J}, \tag{C7}$$

which makes Norm $= \sum_k \frac{\hat{P}_g^2}{\varphi_I\varphi_J}$ consequently.

Furthermore, when the quantity $\hat{\mu}$ of Eq. (12) is used in estimating the correlations following Eq. (23), the normalization need to be reestimated since $\langle a_I a_J^T \rangle \neq \langle \hat{\mu}_I \hat{\mu}_J^T \rangle$ for an average over many GWB realizations. The new normalization factor is found to be

$$\langle \hat{\mu}_I \hat{\mu}_J^T \rangle = \langle \Sigma_I F_I^T D_I^{-1} F_I a_I a_J^T F_J^T (D_J^{-1})^T F_J \Sigma_J^T \rangle$$
$$= \Sigma_I F_I^T D_I^{-1} F \langle a_I a_J^T \rangle F_J^T (D_J^{-1})^T F_J \Sigma_J^T. \tag{C8}$$

[1] M. V. Sazhin, Opportunities for detecting ultralong gravitational waves, Sov. Phys. J. **22**, 36 (1978), https://ui.adsabs.harvard.edu/abs/1978SvA....22...36S/abstract.

[2] S. Detweiler, Pulsar timing measurements and the search for gravitational waves, Astrophys. J. Lett. **234**, 1100 (1979).

[3] R. W. Hellings and G. S. Downs, Upper limits on the isotropic gravitational radiation background from pulsar timing analysis, Astrophys. J. Lett. **265**, L39 (1983).

[4] Z. Arzoumanian et al. (The NANOGrav Collaboration), The nanograv 12.5 yr dataset: Search for an isotropic stochastic gravitational-wave background, Astrophys. J. Lett. **905**, L34 (2020).

[5] J. Antoniadis et al., The international pulsar timing array second data release: Search for an isotropic gravitational wave background, Mon. Not. R. Astron. Soc. **510**, 4873 (2022).

[6] S. Chen et al., Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: Inferences in the stochastic gravitational-wave background search, Mon. Not. R. Astron. Soc. **508**, 4970 (2021).

[7] B. Goncharov et al., On the evidence for a common-spectrum process in the search for the nanohertz gravitational-wave background with the parkes pulsar timing array, Astrophys. J. Lett. **917**, L19 (2021).

[8] G. Agazie et al. (NANOGrav Collaboration), The NANO-Grav 15 yr data set: Evidence for a gravitational-wave background, Astrophys. J. Lett. **951**, L8 (2023).

[9] J. Antoniadis et al., The second data release from the European pulsar timing array III. Search for gravitational wave signals, arXiv:2306.16214.

[10] D. J. Reardon et al., Search for an isotropic gravitational-wave background with the parkes pulsar timing array, Astrophys. J. Lett. **951**, L6 (2023).

[11] H. Xu et al., Searching for the nano-hertz stochastic gravitational wave background with the Chinese pulsar timing array data release I, Res. Astron. Astrophys. **23**, 075024 (2023).

[12] X. Siemens, J. Ellis, F. Jenet, and J. D. Romano, The stochastic background: Scaling laws and time to detection for pulsar timing arrays, Classical Quantum Gravity **30**, 224015 (2013).

[13] L. Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren, Hyper-efficient model-independent Bayesian method for the analysis of pulsar timing data, Phys. Rev. D **87**, 104021 (2013).

[14] R. van Haasteren and M. Vallisneri, New advances in the Gaussian-process approach to pulsar-timing data analysis, Phys. Rev. D **90**, 104012 (2014).

[15] G. E. Freedman, A. D. Johnson, R. van Haasteren, and S. J. Vigeland, Efficient gravitational wave searches with pulsar timing arrays using Hamiltonian Monte Carlo, Phys. Rev. D **107**, 043013 (2023).

[16] S. Hourihane, P. Meyers, A. Johnson, K. Chatziioannou, and M. Vallisneri, Accurate characterization of the stochastic gravitational-wave background with pulsar timing arrays by likelihood reweighting, Phys. Rev. D **107**, 084045 (2023).

[17] J. Sun, P. T. Baker, A. D. Johnson, D. R. Madison, and X. Siemens, Implementation of an efficient Bayesian search for gravitational wave bursts with memory in pulsar timing array data, Astrophys. J. **951**, 121 (2023).

[18] S. R. Taylor, J. Simon, L. Schult, N. Pol, and W. G. Lamb, A parallelized Bayesian approach to accelerated gravitational-wave background characterization, Phys. Rev. D **105**, 084049 (2022).

[19] W. G. Lamb, S. R. Taylor, and R. van Haasteren, The need for speed: Rapid refitting techniques for Bayesian spectral characterization of the gravitational wave background using PTAs, arXiv:2303.15442.

[20] J. S. Hazboun *et al.*, The NANOGrav 11 yr data set: Evolution of gravitational-wave background statistics, Astrophys. J. **890**, 108 (2020).

[21] M. Anholm, S. Ballmer, J. D. E. Creighton, L. R. Price, and X. Siemens, Optimal strategies for gravitational wave stochastic background searches in pulsar timing data, Phys. Rev. D **79**, 084030 (2009).

[22] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[23] S. J. Chamberlin, J. D. E. Creighton, X. Siemens, P. Demorest, J. Ellis, L. R. Price, and J. D. Romano, Time-domain implementation of the optimal cross-correlation statistic for stochastic gravitational-wave background searches in pulsar timing data, Phys. Rev. D **91**, 044048 (2015).

[24] R. Case, N. Laal, J. D. Romano, and X. Siemens (to be published).

[25] S. J. Vigeland, K. Islo, S. R. Taylor, and J. A. Ellis, Noise-marginalized optimal statistic: A robust hybrid frequentist-Bayesian statistic for the stochastic gravitational-wave background in pulsar timing arrays, Phys. Rev. D **98**, 044003 (2018).

[26] R. Kumar, C. Carroll, A. Hartikainen, and O. Martin, ArviZ a unified library for exploratory analysis of Bayesian models in python, J. Open Source Softwaare **4**, 1143 (2019).

[27] J. Ellis and R. van Haasteren, jellis18/ptmcmcsampler: Official release (2017), https://doi.org/10.5281/zenodo.1037579.

[28] E. Hellinger, Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen., J. Reine Angew. Math. **1909**, 210 (1909).

[29] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite, Zenodo (2023), https://doi.org/10.5281/zenodo.4059815.

[30] S. R. Taylor, P. T. Baker, J. S. Hazboun, J. Simon, and S. J. Vigeland, enterprise-extensions (2021), v2.4.3, https://github.com/nanograv/enterprise_extensions.

[31] J. D. Hunter, Matplotlib: A 2d graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[32] P. T. Inc., Collaborative data science (2015), https://plotly.com/python/.

[33] Z. Arzoumanian *et al.* (NANOGrav Collaboration), The NANOGrav nine-year data set: Limits on the isotropic stochastic gravitational wave background, Astrophys. J. Lett. **821**, 13 (2016).

[34] J. S. Hazboun, J. D. Romano, and T. L. Smith, Realistic sensitivity curves for pulsar timing arrays, Phys. Rev. D **100**, 104028 (2019).

[35] S. R. Taylor, The nanohertz gravitational wave astronomer, arXiv:2105.13270.

[36] D. J. Reardon *et al.*, The gravitational-wave background null hypothesis: Characterizing noise in millisecond pulsar arrival times with the parkes pulsar timing array, Astrophys. J. Lett. **951**, L7 (2023).

[37] B. Goncharov, D. J. Reardon, R. M. Shannon, X.-J. Zhu, E. Thrane, M. Bailes, N. D. R. Bhat, S. Dai, G. Hobbs, M. Kerr, R. N. Manchester, S. Osłowski, A. Parthasarathy, C. J. Russell, R. Spiewak, N. Thyagarajan, and J. B. Wang, Identifying and mitigating noise sources in precision pulsar timing data sets, Mon. Not. R. Astron. Soc. **502**, 478 (2020).