

A parallelized Bayesian approach to accelerated gravitational-wave background characterization

Stephen R. Taylor^{1,*}, Joseph Simon,² Levi Schult^{1,3}, Nihan Pol¹, and William G. Lamb¹

¹*Department of Physics and Astronomy, Vanderbilt University,
2301 Vanderbilt Place, Nashville, Tennessee 37235, USA*

²*Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, Colorado 80309, USA*

³*Department of Astronomy, University of Virginia, 1827 University Avenue, P.O. Box 400325,
Charlottesville, VA 22904, USA*



(Received 19 February 2022; accepted 28 March 2022; published 26 April 2022)

The characterization of nanohertz-frequency gravitational waves (GWs) with pulsar-timing arrays requires a continual expansion of datasets and monitored pulsars. Whereas detection of the stochastic GW background is predicated on measuring a distinctive pattern of interpulsar correlations, characterizing the background's spectrum is driven by information encoded in the power spectra of the individual pulsars' time series. We propose a new technique for rapid Bayesian characterization of the stochastic GW background that is fully parallelized over pulsar datasets. This factorized likelihood technique empowers a modular approach to parameter estimation of the GW background, multistage model selection of a spectrally-common stochastic process and quadrupolar interpulsar correlations, and statistical cross-validation of measured signals between independent pulsar subarrays. We demonstrate the equivalence of this technique's efficacy with the full pulsar-timing array likelihood, yet at a fraction of the required time. Our technique is fast, easily implemented, and trivially allows for new data and pulsars to be combined with legacy datasets without reanalysis of the latter.

DOI: [10.1103/PhysRevD.105.084049](https://doi.org/10.1103/PhysRevD.105.084049)

I. INTRODUCTION

The precision timing of millisecond radio pulsars can be exploited over long observational campaigns to search for ultra low-frequency gravitational waves (GWs). As a GW deforms the spacetime between the Earth and a pulsar, it causes the times of arrival (TOAs) of radio pulses to deviate from flat spacetime expectations. These arrival-time deviations depend on the relative position of the GW source to the Earth-pulsar line-of-sight, and the GW metric perturbation at the time the wave front first passed the pulsar as well as when it subsequently passes the Earth [1–3]. The result is that after detailed timing ephemerides are constructed for each pulsar—depending on its rotational period, period spindown, etc.—there will be an additional unmodeled GW-induced influence.

The GW frequencies to which pulsar-timing campaigns are sensitive depend on how the time series of pulses is sampled; broadly speaking, the lower frequency is dictated by the inverse observation time (now greater than a decade, such that $f_{\text{low}} \lesssim 3$ nHz), while the accessible upper frequencies are limited by the observational cadence and timing precision to $f_{\text{high}} \sim 0.1$ μ Hz. The class of objects that are expected to dominate the GW emission in this band

of frequencies are binary systems of supermassive black holes, with masses $\gtrsim 10^8$ – 10^{10} $M_{\odot}$ [4–9]. These should be formed as a consequence of the hierarchical growth of galaxies over cosmic time, where resident supermassive black holes pair up within the common galactic merger remnant after a chain of dynamical processes. The GW signals from these supermassive black hole binaries (SMBHBs) will not all be individually resolvable with pulsar timing, instead stacking incoherently within frequency-resolution bins to form a stochastic GW background (GWB) [10–12].

In an individual pulsar's data, this GWB will manifest as a low-frequency stochastic (red) process, which could appear spectrally similar to intrinsic red noise from e.g., pulsar rotational instabilities, or pulse propagation through time-dependent ionized interstellar medium paths [13–16]. Therefore, GW searches through pulsar timing are performed by correlating data across a pulsar-timing array (PTA) [17], where the sought after evidence consists of a distinctive GWB-induced interpulsar correlation signature known as the Hellings and Downs curve [18]. There are several collaborations that have now been contributing to this goal for over a decade, including the North American Nanohertz Observatory for Gravitational waves (NANOGrav) [19], the Parkes Pulsar Timing Array (PPTA) [20,21], and the European Pulsar Timing Array (EPTA) [22]. Together with

*stephen.r.taylor@vanderbilt.edu

the Indian Pulsar Timing Array (InPTA) [23], these collaborations also constitute the International Pulsar Timing Array (IPTA) [24–26]. The Chinese Pulsar Timing Array [27], and telescope-centered efforts such as CHIME/Pulsar [28] and MeerTime [29], will be important future contributors.

These PTA collaborations have recently delivered results that, for the first time, go beyond placing upper limits on the amplitude of a putative GWB. The NANOGrav Collaboration has announced the discovery a low-frequency stochastic process in its 12.5 year dataset [30], whose spectrum is common across many of the pulsars in its array. For a fiducial model of the GWB generated by a population of SMBHBs, the median characteristic strain amplitude at a frequency of 1/yr is 1.92×10^{-15} . Likewise, the PPTA [31] and EPTA [32] have found similar processes with median strain amplitude 2.2×10^{-15} and 2.95×10^{-15} , respectively. All of these results are statistically consistent with one another, and broadly in alignment with theoretical expectations of the GWB spectrum from SMBHBs [33]. Furthermore, the IPTA has also recently announced the discovery of a similar stochastic process with amplitude 2.8×10^{-15} [26], where this result derives from the synthesis of older NANOGrav, PPTA, and EPTA datasets. However, none of these results exhibit significant interpulsar correlations, and as such these processes could still be noise or other systematic processes of non-GW origin. Nevertheless, it has been argued that the first milestone of GWB detection is the initial emergence of a common-spectrum process in PTA datasets, followed later by significant interpulsar correlations [34,35]. The latter measurement is the definitive evidence for a GWB, and requires that data from many pulsars be cross-correlated to infer the Hellings and Downs signature.

Modern PTA data-analysis frameworks model all signal and noise processes in the time domain. This is primarily for practical reasons. The signal of interest is a low-frequency stochastic process whose power-spectral density (PSD) is steeper than the PSD of a rectangular window function for existing pulsar-timing baselines; as such, a direct Fourier transform of the data without careful windowing would suffer from spectral leakage effects. Likewise, the observations are unevenly sampled, again rendering traditional Fourier transform techniques invalid (although Lomb-Scargle periodogram techniques could be used). The drawback of time-domain modeling is that all observations across all pulsars must be cross-correlated within a Gaussian likelihood function. Evaluating this likelihood requires the inversion of a dense covariance matrix, which scales with the number of timing observations as $\propto \mathcal{O}(N_{\text{obs}}^3)$. Furthermore, with Bayesian techniques now the standard statistical approach, this likelihood function must be sampled at least $\gtrsim 10^6$ times across large parameter spaces in a global search over signal and noise parameters. As such, given that the total number of observations in recent NANOGrav data releases exceeds 10^5 [36], even direct time-domain modeling is not tractable.

The remedy to this has been rank-reduced approaches [37–39], where random Gaussian processes are modeled in the time domain with basis design matrices that are much lower in dimensionality than the number of observations. The bottleneck of the likelihood remains a Cholesky covariance matrix inversion, but now of order the number of basis coefficients rather than the number of observations. The number of basis coefficients includes the number of timing ephemeris parameters, and the number of frequencies on which the GWB is modeled, which is small since we are targeting a low-frequency signal. Thus, rank-reduced modeling approaches enable PTA GWB searches to be carried out within reasonable periods of time ($\sim$ days to weeks). However, even this approach is becoming costly, and will struggle to cope in the future with the expansion of PTAs to include new pulsars and data e.g., from daily CHIME observations. Hence, in this paper we propose a new analysis scheme that parallelizes PTA data analysis over pulsars, and compresses pulsar data into sufficient statistics that can be arbitrarily and quickly combined in postprocessing. The resulting approach is fast through the trivial parallelization of the analysis, and with a modularity that can cope with the computational demands of the ever-growing PTA data volume in the future.

This paper is laid out as follows. We outline our approach for parallelization of Bayesian PTA data analysis in Sec. II. In Sec. III we carry out a sequence of tests of this new approach, including assessing its performance for parameter estimation of the GWB’s amplitude, and recovering the significance of a common-spectrum process and GWB-induced interpulsar correlations. We conclude in Sec. IV, where we also remark on future applications of this method and further generalizations that we are exploring.

II. PARALLELIZING THE PTA LIKELIHOOD

GW searches with PTAs begin with data in the form of timing residuals, i.e., a best-fit deterministic timing ephemeris has already been constructed in terms of the pulsar period, spindown, binary parameters, etc., and the ephemeris-modeled TOAs then subtracted from the observed TOAs. These residuals encapsulate noise and any phenomena that were unmodeled at the stage of the original timing ephemeris construction. The latter group includes the influence of GW signals on the TOAs. A GWB will imprint spatiotemporal correlations in PTA data; these include long-timescale correlations in the residual time series as a result of the expected low-frequency characteristic strain spectrum, as well as a distinctive pattern of interpulsar correlations known as the Hellings and Downs curve [18]. This curve is the relevant correlation pattern imparted by an isotropic, stationary, and unpolarized Gaussian GWB on PTA data, and depends only on the angular separation between pulsars.

As a result, the entire PTA dataset can be modeled with a multivariate Gaussian likelihood that has a dense

time-domain covariance matrix. The elements of this matrix encode the temporal covariance expected between different timing residuals at different times, modulated by the corresponding Hellings and Downs factor for the relevant angular separation between the pulsars. We write the PTA likelihood as

$$p(\delta\mathbf{t}|\boldsymbol{\eta}) = \frac{\exp(-\frac{1}{2}\delta\mathbf{t}^T C^{-1}\delta\mathbf{t})}{\sqrt{\det(2\pi C)}}, \quad (1)$$

where $\delta\mathbf{t}$ is the vector of concatenated timing residuals across all pulsars in the PTA, and $C \equiv C(\boldsymbol{\eta})$ is the data covariance matrix with parameters $\boldsymbol{\eta}$. This covariance matrix can be described by

$$\langle\delta\mathbf{t}_{ai}\delta\mathbf{t}_{bj}^T\rangle = C_{(ai)(bj)} = N_{a,(ij)}\delta_{ab} + C_{a,(ij)}^{\text{red}}\delta_{ab} + \Gamma_{ab}C_{(ij)}^{\text{GWB}}, \quad (2)$$

where $\delta_{(\dots)}$ is the Kronecker delta function, (a, b) index pulsars, (i, j) index timing residuals, N is a white noise covariance matrix in a given pulsar, C^{red} is the intrinsic red-noise covariance matrix in a given pulsar, C^{GWB} is the GWB covariance matrix, and Γ_{ab} is the Hellings and Downs cross-correlation factor between the relevant pair of pulsars, given by

$$\Gamma_{ab} = \frac{3}{2}x_{ab}\ln x_{ab} - \frac{x_{ab}}{4} + \frac{1}{2} + \frac{\delta_{ab}}{2}, \quad (3)$$

and $x_{ab} = (1 - \cos\theta_{ab})/2$ for pulsars with angular separation θ_{ab} . A schematic for the structure of this PTA covariance matrix is shown in Fig. 1.

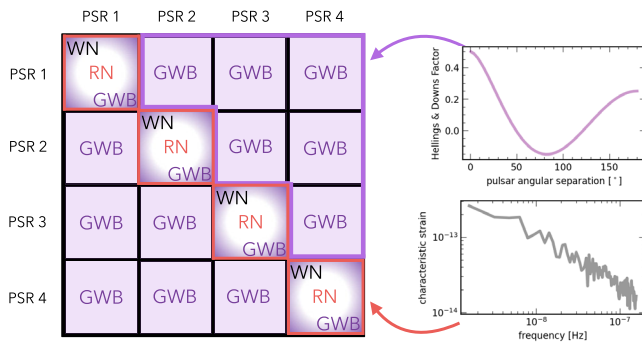


FIG. 1. A schematic representation of the PTA data covariance matrix, where the pulsar autocorrelations are influenced by intrinsic white noise (typically instrumental), intrinsic red noise (typically low-frequency pulsar rotational instabilities), and a GWB signal. By contrast, the cross-correlations are influenced only by the GWB, and encode the sufficient evidence for a GW detection claim. On the other hand, the pulsar autocorrelations will dominate the spectral characterization of the GWB, since the cross-correlations are weaker.

Recent work [34] has shown that the curation of existing PTA datasets has led to several high-quality pulsars being observed over long baselines, with the potential result that key pulsars now lie in the intermediate signal regime of GWB detection; in Siemens *et al.* [40] this is defined as the GWB signal exceeding the level of intrinsic pulsar and instrumental noise in lower frequencies in the power spectrum of pulsar timing residuals. An additional effect is that the expected GWB-induced cross-correlations are no greater than a half of the autocorrelation values (i.e., the Hellings and Downs curve has a maximum of 0.5). As such, the significance of a common-spectrum process (i.e., information encoded in the autocorrelations) should currently be much greater than the significance of GWB-induced cross-correlations. This has been shown in simple analyses [34], as well as more realistic simulations that mirror the existing cadence and sensitivity of real datasets [35]. Thus, the working explanation for the emergence of strong common-spectrum processes in the datasets of all three long-baseline PTA collaborations [30–32] is that this is an early indicator of an emerging GWB signal at low frequencies, to be followed several years later by the definitive GWB evidence in the form of significant cross-correlation measurement [35].

With the information content of GWB autocorrelations currently swamping that of cross-correlations, the PTA covariance matrix can be approximated with a block-diagonal structure. While we can not use this autocorrelation information directly to claim evidence of a GWB, we can use it for accurate, fast, and parallelizable spectral characterization. To see this, we now drop the cross-correlation blocks of the PTA covariance matrix, such that

$$\begin{aligned} p(\delta\mathbf{t}|\boldsymbol{\eta}) &\approx \frac{\exp(-\frac{1}{2}\sum_a \delta\mathbf{t}_a^T C_{aa}^{-1} \delta\mathbf{t}_a)}{\sqrt{\det(2\pi C)}}, \\ &= \prod_a \frac{\exp(-\frac{1}{2}\delta\mathbf{t}_a^T C_{aa}^{-1} \delta\mathbf{t}_a)}{\sqrt{\det(2\pi C_{aa})}}, \\ &= \prod_a p(\delta\mathbf{t}_a|\boldsymbol{\eta}). \end{aligned} \quad (4)$$

We have now factorized the PTA likelihood into a product of per-pulsar likelihoods. This then allows independent, parallel analyses to be performed in each pulsar dataset, with these intermediate results then stitched together in postprocessing to yield the full PTA result.

In this paper we consider a GWB with the fiducial power-law characteristic strain spectrum from a population of GW-emitting SMBHBs, $h_c(f) = A_{\text{GWB}}(f/1 \text{ yr}^{-1})^{-2/3}$ [e.g., [41]], which yields a timing-residual power spectral density (PSD), $S_{\text{GWB}}(f) = (A_{\text{GWB}}^2/12\pi^2)(f/1 \text{ yr}^{-1})^{-\gamma} \text{ yr}^3$, where $\gamma = 13/3$. As such, in each independent pulsar dataset we sample intrinsic pulsar and instrumental noise, in addition to the amplitude, A_{CP} , of a low-frequency common process that has a fixed PSD spectral index of $\gamma = 13/3$. This latter

common process is a proxy for the GWB in an individual pulsar dataset. Since A_{CP} is the only common search parameter across the array, we can use its marginal posterior distribution from each pulsar as a sufficient statistic to recover the marginal posterior of A_{CP} for the entire array, i.e.,

$$p(A_{\text{CP}}|\delta\mathbf{t}) \propto p(A_{\text{CP}}) \prod_a \frac{p(A_{\text{CP}}|\delta\mathbf{t}_a)}{\bar{p}(A_{\text{CP}})}, \quad (5)$$

where $\bar{p}(A_{\text{CP}})$ is the prior on the common-process amplitude applied during the analysis of each pulsar dataset, and $p(A_{\text{CP}})$ is the prior on the final, desired PTA result. The practical details of how this factorized likelihood (FL) scheme is implemented are given in the next section, along with the means by which FL results can also accelerate the search for cross-correlation significance.

III. RESULTS

We test the efficacy of our new FL approach against the full PTA likelihood function using a suite of realistic PTA simulated datasets. The format of these simulations closely follow Pol *et al.* [35]. The PTA configuration corresponds to the 45 pulsars from the NANOGrav 12.5 year dataset [36] that were searched for the influence of an isotropic stochastic GWB [30]. Observational timestamps, TOA uncertainties, noise properties, and other metadata were derived from the real pulsar data, and then used to create synthetic noise-only data realizations. Rather than restrict to a maximum of 12.5 yr of data, we extend to a near-future scenario corresponding to 15 yr. This is achieved by drawing observational cadences and uncertainties from distributions of the final year of real data, then adding new observations to each pulsar until the maximum baseline of 15 yr is reached. No new pulsars are added to the array when expanding from 12.5 to 15 yr [42].

We then inject 100 different realizations of a GWB signal into these noise-only datasets, generating a set of 100 PTA datasets, each of which has 45 pulsars. One of our key aims is to verify that the FL approach has equivalent Bayesian statistical coverage to the full PTA likelihood, i.e., the injected parameter values fall within the $p\%$ posterior credible region in $p\%$ of realizations. In Bayesian inference, this is only guaranteed when the injected parameter values are drawn from within the prior distribution that is subsequently used in the analysis. Thus, the spectral characteristics of the GWB used to generate each of the 100 injections are drawn from $\log_{10} A_{\text{GWB}} \in U[-18, -14]$, where the spectral index is fixed at $\gamma_{\text{GWB}} = 13/3$.

The basic analyses to be performed for all of the following tests are Bayesian per-pulsar noise characterization studies. Since we are ignoring interpulsar correlations, these analyses are performed entirely in parallel. The inference model of each corresponds to white noise fixed at the injected level, with intrinsic red noise characterized by a power-law PSD with search parameters

$\log_{10} A_{\text{red}} \in U[-20, -11]$ and $\gamma_{\text{red}} \in U[0, 7]$. We also model an additional power-law PSD process with fixed spectral index $\gamma_{\text{CP}} = 13/3$, and $\log_{10} A_{\text{CP}} \in U[-18, -14]$; this process acts as a modeling proxy for the presence of the GWB in each pulsar. The result of analyzing each PTA dataset is a sequence of 45×3 -dimensional Markov Chain Monte Carlo (MCMC) chains, each iterated for at least 1×10^6 steps, and subsequently thinned by a factor of 10 (i.e., only every 10th sample was used) to deliver 1×10^5 usable samples. Each chain is postprocessed into a histogram representation of the 1-dimensional marginalized posterior distribution of $\log_{10} A_{\text{CP}}$ in the respective pulsars. While there are various rules-of-thumb to deduce the optimal bin-width in such a histogram representation (e.g., Scott's rule [43]), these usually assume that the underlying density is Gaussian. A more agnostic scheme would implement the Sheather-Jones algorithm [44], which attempts to minimize the asymptotic mean integrated squared error between the estimator and the true underlying density. In practice, we used $N_{\text{bin}} = 100$ for most of our results, however this choice is varied in Sec. III C. The end result of analyzing all pulsars within a given PTA dataset is that the data has been compressed into histogram estimators of the marginal posterior of $\log_{10} A_{\text{CP}}$.

A. Parameter estimation

With our entire PTA datasets distilled down to a set of histogram probability density estimators of the common-spectrum process' amplitude, parameter estimation is remarkably simple. Given that the prior on this amplitude was uniform in $\log_{10} A_{\text{CP}}$, all histogram estimators can be multiplied together and renormalized to deduce the equivalent full PTA posterior of $\log_{10} A_{\text{CP}}$. One pragmatic refinement is that occasionally some pulsars will lack MCMC-sample coverage over the same values as others, leading to multiplication of finite values by zeros in some bins. This can result in sharp posterior declines or poor posterior recovery at higher, data-informed values of the amplitude. To remedy this, we add a small value of $\epsilon = 10^{-20}$ to all bins in all pulsar histograms.

In the left panel of Fig. 2, we use three simulations (corresponding to strong, moderate, and weak signal injections) to contrast the performance of our FL estimation of $\log_{10} A_{\text{CP}}$ with the full PTA likelihood model. Not only is the FL estimate equivalent to what is deduced through sampling with the full PTA likelihood, it is also much less inhibited by sampling limitations in low-density regions. It is able to recover the posterior density many orders of magnitude lower than what direct MCMC sampling with the PTA likelihood can achieve (compare the FL estimation as solid lines in Fig. 2 to direct MCMC sampling as shaded histograms), which results from sampling the tails of many individual pulsar posteriors that have less contrast than the full PTA posterior. We will see the enormous benefit of this for model selection in the next subsection. The right panel

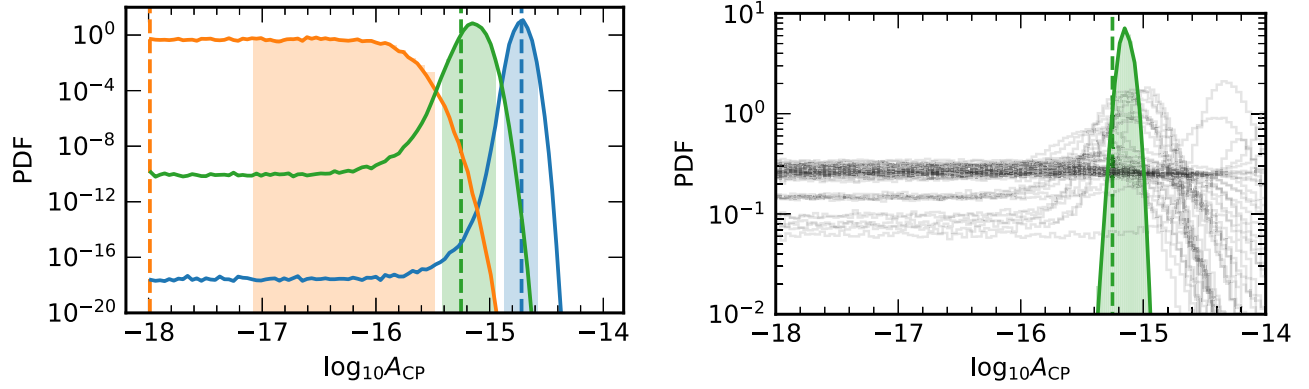


FIG. 2. Left: Example of common process amplitude posterior recovery for a weak, moderate, and strong signal injection. Faded regions are from MCMC sampling of the full PTA likelihood. Solid lines are parallelized reconstructions using the factorized likelihood technique. Dashed vertical lines are the injected values. Right: The FL amplitude posterior recovery of the moderate signal from the left panel is shown again here, along with the common process amplitude posteriors from the individual pulsar analyses (faded gray histograms). These individual pulsar posteriors are the compressed representation of the pulsar data that are used for the final PTA recovery (green).

of Fig. 2 illustrates the FL recovery of the moderate GWB signal injection in the PTA, alongside the recoveries of the amplitude of the proxy process in each individual pulsar.

While the case studies shown in Fig. 2 are very promising, we now undertake a more rigorous assessment of the parameter estimation efficacy of the FL approach. The simulated dataset generation that was discussed earlier was specifically constructed to enable tests of Bayesian coverage through $p - p$ plots. By drawing injected signal parameters from within the analysis prior, the injected values should lie within the $p\%$ credible region in $p\%$ of simulations. Rather than judge the overall efficacy of the standard PTA data analysis pipeline, the scope of our analysis is to compare the relative performance of the FL approach to that achieved with full PTA likelihood. The results of this comparison are shown in Fig. 3, where the difference in the Bayesian coverage of each approach is shown as a function of credible interval. If the FL approach perfectly matched what an analysis with the full PTA likelihood could achieve then the difference should be zero at all credible intervals. While there is some mild scatter around zero, the resulting match is excellent, indicating that the FL approach effectively reproduces the efficacy of the standard PTA analysis pipeline.

B. Model selection

The parallelized parameter estimation discussed in the previous section facilitates a two-stage model selection process for PTA GWB searches, namely: (1) emergence of a spectrally-common stochastic process across pulsars in the array, and (2) detection of Hellings and Downs interpulsar correlations, regarded as the definitive detection signature.

1. Detecting a common process

The relevant detection statistic for a spectrally-common stochastic process is the Bayes factor between models that

include this process versus one in which it is absent (i.e., per-pulsar noise alone). In our case, where the GWB spectral shape is fixed to $\gamma_{\text{CP}} = 13/3$, the amplitude parameter acts as a signal “switch” that encompasses the noise-only model at the lower prior boundary ($\log_{10} A_{\text{CP}} = -18$), and the signal + noise model for all other values. As such, computation of this Bayes factor can be addressed with the Savage-Dickey approximation, which depends only on the prior to posterior density ratio at the amplitude for which the signal is effectively zero. Hence,

$$\mathcal{B}_{\text{CP}} = \frac{p(\log_{10} A_{\text{CP}} = -18)}{p(\log_{10} A_{\text{CP}} = -18 | \{d_i\}_{N_p})}. \quad (6)$$

We evaluate this for our suite of simulations, using both the full PTA model analysis and the parallelized FL analysis. In the full PTA model analysis, the posterior

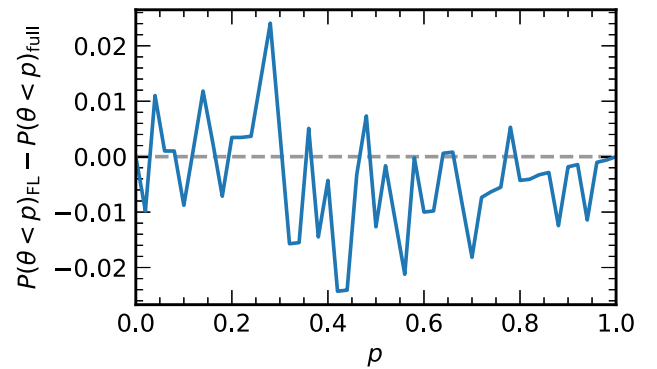


FIG. 3. Bayesian $p - p$ comparison of amplitude recovery between the FL approach and the standard full PTA likelihood. The y-axis shows the difference that these two approaches give for the fraction of simulations in which the injected value lies within the $p\%$ credible region on the x-axis. The horizontal line is for equivalent efficacy of the approaches.

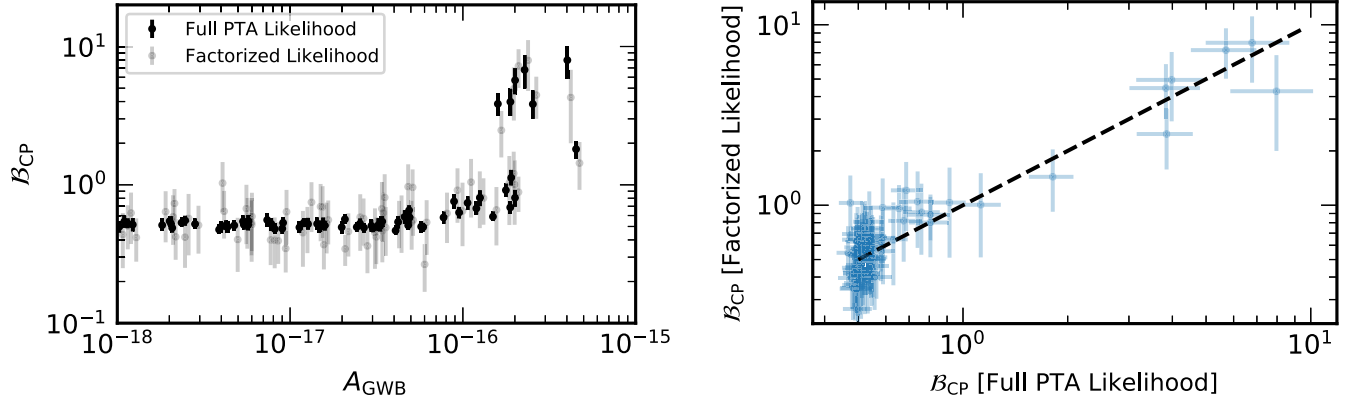


FIG. 4. Left: Bayes factors in favor of a common-spectrum process across all pulsars are shown as a function of the injected GWB amplitude. These are computed under the standard full PTA likelihood analysis, and using the parallelized FL analysis. The FL values are horizontally offset for ease of viewing. Right: The amplitude dependence from the left panel is collapsed over to more directly compare the Bayes factor computed under each approach. The black dashed line is a zero-intercept unit-gradient relation to represent perfect agreement.

density is practically evaluated using $p(\log_{10} A_{CP} = -18 | \{d_i\}_{N_p}) = f_b / \Delta_b$, where f_b is the fraction of total MCMC samples contained within $-18 \leq \log_{10} A_{CP} \leq -18 + \Delta_b$, and the choice of Δ_b is varied over 100 values between 0.01 and 0.1, with the results averaged. The bin width Δ_b is chosen to be small enough so as not to encounter any significant changes in the posterior density away from the value at the lower prior boundary. In the FL analysis, the normalized histogram estimator directly provides the posterior density. In both approaches, we estimate the uncertainty on the Bayes factors using bootstrap resampling. In the full PTA model analysis the single MCMC chain is resampled with replacement, while in the FL analysis each pulsar's MCMC chain is resampled with replacement before forming new histogram estimators that are multiplied across. This procedure is repeated 100 times for each approach, generating distributions of Bayes factors that are summarized as median values with 68% uncertainty regions. The results are shown in Fig. 4, where the left panel exhibits the growth of this common-process Bayes factor with GWB amplitude. The results deduced under each approach match well, as is shown more directly in the right panel where the amplitude dimension is collapsed over and the Bayes factors are directly plotted against one another.

2. Detecting interpulsar spatial correlations

While evidence for a common-spectrum process across the array is a necessary condition for PTA GWB detection, it is not sufficient. It is possible that statistically-independent (yet spectrally-similar) low-frequency noise may arise in millisecond pulsars, which could appear as a common-spectrum process. Likewise, solar-system ephemeris or clock systematics could manifest as low-frequency processes in pulsar timing-residual data. As such, the only

robust metric for GWB detection is via the measurement of Hellings and Downs interpulsar timing-residual correlations, corresponding to the overlap reduction function for an isotropic stochastic background composed of tensor-transverse GW polarizations. The signature is predominantly quadrupolar in pulsar angular separation, with power at higher multipoles but none at $l = 0, 1$. By contrast, intrinsic pulsar noise has zero expected cross-correlation with other distinct pulsars, while solar-system ephemeris and clock systematics are expected to produce dipolar and monopolar correlations, respectively [45].

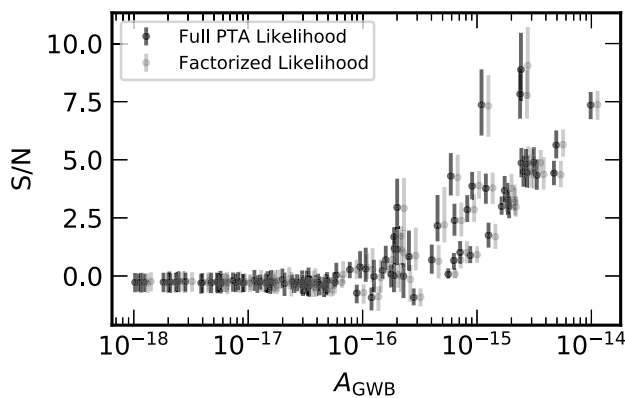
The FL approach can be used to construct a statistic that measures the significance of cross-correlations, whether they be Hellings and Downs or another template signature. This statistic is the optimal two-point correlation statistic for a GWB in PTA data—commonly referred to as the optimal statistic (OS)—and provides a signal-to-noise (S/N) ratio estimate for cross-correlations in PTA data [40,46–48]. The OS can be shown to derive from the linear term in a Taylor expansion of the PTA log-likelihood with respect to cross correlations [46,47,49], and as such it is only applicable in the weak signal regime (where signal strength here refers to the relative amplitude of cross-correlated power versus autocorrelated power). However, an approach has been developed to improve the efficacy of the OS in intermediate signal scenarios, where it is marginalized over the joint posterior distribution of intrinsic pulsar noise and autocorrelated GWB power that has been computed from a full PTA search for a common process [48,50]. Both intrinsic noise and autocorrelated GWB power feature as individual pulsar weightings in the construction of the OS; by mapping the distribution of the statistic under the posterior spread of these weighting contributions, we propagate all sources of uncertainty in the noise estimation into the final OS distribution. This approach is referred to as the noise marginalized OS

(NMOS). It is important that the calculation of the NMOS S/N ratio use Monte Carlo draws from a joint posterior derived from a full PTA search, so that the covariance between individual pulsar red noise parameters and the common process (i.e., the autocorrelation term of the GWB) is accurately reflected.

As shown previously in this paper, the FL approach gives an almost identical reproduction of the common red-noise posterior when compared to a conventional full PTA analysis. Therefore we would expect it to give an equally faithful reproduction of NMOS results. However, to do so we must reweight the posterior distributions from each of our separate pulsar analyses in order to achieve consistency with the FL-reconstructed common-process posterior. This is equivalent to using the FL-reconstructed marginal distribution of the common process amplitude as a weighting to piece together the full, joint PTA posterior distribution from our separate per-pulsar analyses. By doing so, the covariance between all red noise models across the PTA is properly incorporated just as in a conventional, full PTA Bayesian analysis.

Mathematically, this is achieved through reweighting the conditional distribution of red noise parameters from an individual pulsar analysis, $p(\{A_{\text{red}}, \gamma_{\text{red}}\}_i | A_{\text{CP}}, d_i)$, by the ratio of the FL-reconstructed common process posterior $p(A_{\text{CP}} | \{d_i\}_{N_p})$ to the common-spectrum process posterior in that individual pulsar, $p(A_{\text{CP}} | d_i)$,

$$p(\{A_{\text{red}}, \gamma_{\text{red}}\}_i | A_{\text{CP}}, \{d_i\}_{N_p}) = p(\{A_{\text{red}}, \gamma_{\text{red}}\}_i | A_{\text{CP}}, d_i) \times \frac{p(A_{\text{CP}} | \{d_i\}_{N_p})}{p(A_{\text{CP}} | d_i)}. \quad (7)$$



The calculation of the NMOS is then as follows: (i) a random draw is made from the FL posterior of A_{CP} ; (ii) the reweighted joint posterior of a pulsar's red noise parameters given this value of A_{CP} is constructed, and then used to draw values for the red noise parameters. This procedure ensures that the drawn samples capture the same covariance structure present in the posteriors from a full Bayesian analysis. Stage (ii) is repeated for every pulsar in the PTA. The entire process is repeated 1×10^4 times to give a distribution of OS values. Practically, we carry out this procedure by creating a three-dimensional empirical distribution for each pulsar from the posteriors for $(A_{\text{red}}, \gamma_{\text{red}}, A_{\text{CP}})$. The 3D distribution is then collapsed to a 2D conditional distribution for $(A_{\text{red}}, \gamma_{\text{red}})$ for each value of A_{CP} drawn from the FL posterior. The red noise parameters used in the NMOS calculation are then drawn from each individual pulsar's 2D conditional red-noise distribution.

Figure 5 shows the results of applying this FL scheme for computing cross-correlation S/N ratio values on our simulated datasets. The agreement with the conventional approach is very strong, showing virtually identical recovery in the median and spread of S/N ratio values. The benefits of the FL approach are that this never requires a full PTA Bayesian analysis, pulsars do not need to be reanalyzed repeatedly, and this parallelized scheme easily allows for new pulsars to be analyzed and incorporated into the calculations.

C. Cross-validation

Parameter estimation and model selection are important inference tasks, but are not the only means by which confidence can be built in the detection of a signal. The model selection discussed so far has all been “in sample”,

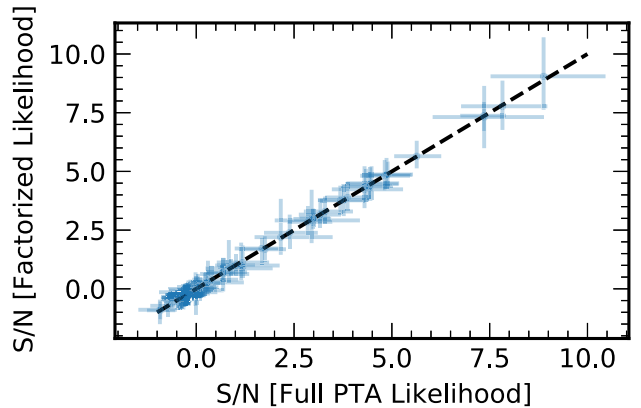


FIG. 5. Left: NMOS S/N ratio values for Hellings and Downs cross-correlations are shown as a function of the injected GWB amplitude. The conventional calculation involves marginalizing the OS values over noise and common-process samples generated from a Bayesian analysis with the full PTA likelihood. In the FL approach, noise values for each pulsar are drawn from reweighted conditional distributions based on the FL-reconstruction of the common process amplitude, thereby imprinting the covariance structure of the full PTA on individual-pulsar posterior distributions. The FL values are horizontally offset for ease of viewing. Right: The amplitude dependence from the left panel is collapsed over to more directly compare the S/N ratio values from each approach. The black dashed line is a zero-intercept unit-gradient relation to represent perfect agreement.

i.e., all data is used to compute a figure of merit under different models, then the model with the largest figure of merit (in our case the Bayesian evidence) is favored. However, “out of sample” model selection can also be important in assessing how well a model is internally self-consistent among the data (see, e.g., Wang *et al.* [51]). For example, leave-one-out cross-validation in PTA analysis consists of assessing a model’s performance by conditioning it on all but one pulsar, then projecting it into the left-out pulsar to see how well it is supported. Essentially, we are testing the predictive power of a model that has been trained on a portion of the data then tested against another portion that has been held out.

In the PTA literature, leave-one-out cross-validation is referred to as “dropout,” because each pulsar is dropped out sequentially to determine whether the model evidence for a common process increases or decreases as a result [30,31,52]. Below we provide a mathematical framework for the dropout factor to serve as a guide for developing a FL approach. The dropout factor will be presented in full in a separate study [53].

Consider two hypotheses for an array of N_p pulsars:

- (i) $\mathcal{H}_0$: intrinsic noise per pulsar, plus a common-spectrum process (with 13/3 spectral index) in all N_p pulsars.
- (ii) $\mathcal{H}_1$: intrinsic noise per pulsar, plus a common-spectrum process (with 13/3 spectral index) in $N_p - 1$ pulsars, where pulsar p is included in the analysis but does not respond to the common process.

The Bayesian evidence for $\mathcal{H}_0$ is

$$\mathcal{Z}_0 = p(\{d_i\}_{N_p} | \mathcal{H}_0) = \int p(\{d_i\}_{N_p} | \{\vec{\theta}_i\}_{N_p}, A_{\text{CP}}) \times p(\{\vec{\theta}_i\}_{N_p}) p(A_{\text{CP}}) d^{N_p} \vec{\theta} dA_{\text{CP}}, \quad (8)$$

where $p(\{d_i\}_{N_p} | \{\vec{\theta}_i\}_{N_p}, A_{\text{CP}})$ is the PTA likelihood for all N_p pulsars, $p(\{\vec{\theta}_i\}_{N_p})$ is the prior for all intrinsic noise parameters, and $p(A_{\text{CP}})$ is the prior for the common process’ amplitude. As discussed in Sec. II, without any interpulsar correlations the PTA likelihood exactly factorizes into a product over pulsars. We can simply split this into the likelihood for $(N_p - 1)$ pulsars (which is the likelihood for hypothesis $\mathcal{H}_1$) and the p th pulsar.

$$\begin{aligned} p(\{d_i\}_{N_p} | \{\vec{\theta}_i\}_{N_p}, A_{\text{CP}}) &= \prod_{i=1}^{N_p} p(d_i | \vec{\theta}_i, A_{\text{CP}}) \\ &= p(\{d_i\}_{N_p-1} | \{\vec{\theta}_i\}_{N_p-1}, A_{\text{CP}}) \\ &\quad \times p(d_p | \vec{\theta}_p, A_{\text{CP}}). \end{aligned} \quad (9)$$

Therefore, the evidence for $\mathcal{H}_1$ can be written as

$$\begin{aligned} \mathcal{Z}_1 &= \int p(\{d_i\}_{N_p-1} | \{\vec{\theta}_i\}_{N_p-1}, A_{\text{CP}}) p(\{\vec{\theta}_i\}_{N_p-1}) p(A_{\text{CP}}) \\ &\quad \times p(d_p | \vec{\theta}_p) p(\vec{\theta}_p) d^{N_p} \vec{\theta} dA_{\text{CP}}, \end{aligned} \quad (10)$$

where we have used the fact that the prior on intrinsic noise parameters is also factorizable. Using Bayes’ theorem, we can recognize the first grouping of terms before the product symbol as $p(A_{\text{CP}}, \{\vec{\theta}_i\}_{N_p-1} | \{d_i\}_{N_p-1}) \times \mathcal{Z}_*$, where $\mathcal{Z}_*$ is a normalization factor. Since this is a normalized posterior probability distribution, integrating over the common process amplitude and all intrinsic noise parameters of the $(N_p - 1)$ pulsars considered in $\mathcal{H}_1$ simply yields 1. The only remaining terms are an integral over the product of the likelihood and prior of intrinsic noise parameters in pulsar p , which is the Bayesian evidence for a noise-only model in this pulsar, $\mathcal{Z}_{p,1}$ (i.e., model $\mathcal{H}_1$). Finally, the evidence for model $\mathcal{H}_1$ is

$$\begin{aligned} \mathcal{Z}_1 &= \mathcal{Z}_* \int p(d_p | \vec{\theta}_p) p(\vec{\theta}_p) d\vec{\theta}_p \\ &= \mathcal{Z}_* \mathcal{Z}_{p,1}. \end{aligned} \quad (11)$$

We now return to $\mathcal{Z}_0$ to factorize the likelihood, but remembering that the likelihood for pulsar p depends on A_{CP} in model $\mathcal{H}_0$. We still get a grouping of terms that can be rewritten as $p(A_{\text{CP}}, \{\vec{\theta}_i\}_{N_p-1} | \{d_i\}_{N_p-1}) \times \mathcal{Z}_*$, but one can only marginalize this distribution over the intrinsic noise parameters of which pulsar p is independent. This gives the marginal distribution $p(A_{\text{CP}} | \{\vec{\theta}_i\}_{N_p-1}, \{d_i\}_{N_p-1})$ which we write compactly as $p(A | \mathcal{H}_1)$. The evidence for model $\mathcal{H}_0$ is then

$$\begin{aligned} \mathcal{Z}_0 &= \mathcal{Z}_* \int p(A_{\text{CP}} | \mathcal{H}_1) p(d_p | \vec{\theta}_p, A_{\text{CP}}) p(\vec{\theta}_p) d\vec{\theta}_p dA_{\text{CP}} \\ &= \mathcal{Z}_* \mathcal{Z}_{p,0} \int \frac{p(A_{\text{CP}} | \mathcal{H}_1)}{p(A_{\text{CP}})} p(A_{\text{CP}}, \vec{\theta}_p | d_p) d\vec{\theta}_p dA_{\text{CP}}, \end{aligned} \quad (12)$$

where on the second line we have used Bayes’ theorem to express the calculation as an integral over the posterior distribution of pulsar p , and $\mathcal{Z}_{p,0}$ is the Bayesian evidence for a model that includes a common process in this pulsar (i.e., model $\mathcal{H}_0$). Finally, the ratio of model evidences $\mathcal{Z}_0 / \mathcal{Z}_1$ —which is denoted as the dropout factor that assesses how well pulsar p supports the presence of a common process found by the other $(N_p - 1)$ pulsars—can be practically evaluated as

$$\text{Dropout Factor} = \frac{\mathcal{Z}_0}{\mathcal{Z}_1} = \frac{\mathcal{Z}_{p,0}}{\mathcal{Z}_{p,1}} \left\langle \frac{p(A_{\text{CP}} | \mathcal{H}_1)}{p(A_{\text{CP}})} \right\rangle_p, \quad (13)$$

where $\langle \rangle_p$ is an average over posterior samples from pulsar p . The right-hand side of Eq. (13) is easily computed using

FL techniques, thereby avoiding the need to perform large numbers of PTA analyses and also rendering it trivial to assess the support for a common process in newly added pulsars. The one-dimensional marginalized posterior distribution $p(A_{\text{CP}}|\mathcal{H}_1)$ can be deduced from an FL parameter estimation analysis in $(N_p - 1)$ pulsars, similar to those shown in Sec. III A. The ratio of this posterior density to the known prior density of A_{CP} is then evaluated at, and averaged over, the posterior samples of the dropped-out pulsar p . The ratio of evidences on the right-hand side, $\mathcal{Z}_{p,0}/\mathcal{Z}_{p,1}$, gives the Bayes factor for a common red process in pulsar p , acting as a prior weighting for this common process in the calculation. It can easily be computed from the MCMC chain of each single-pulsar analysis using the Savage-Dickey technique.

We now compute the dropout factors for each pulsar in one of our simulated datasets, which contains an injected GWB signal with amplitude $A_{\text{GWB}} = 1.91 \times 10^{-15}$ (broadly similar to the common process discovered in the NANOGrav 12.5 yr dataset [30]). The conventional technique for computing the dropout factor is via a Bayesian product-space sampling analysis [see, e.g., [54,55]], where an MCMC chain samples a space of parameters that is concatenated over two models; one in which a common process is present in all pulsars, and another in which the common process is present in all but a selected pulsar. The transition between models is achieved by including an additional indexing parameter with different model behavior activated in different ranges, e.g., $n_{\text{model}} \in [0, 0.5]$ activates the model likelihood with a common process in all pulsars, whereas $n_{\text{model}} \in [0.5, 1]$

activates a likelihood with one pulsar dropped out. The ratio of samples for which the chain occupies one model range over another gives the odds ratio between the models, and thus (assuming equal prior odds between the models) the dropout factor. The uncertainty on this value is derived from 1×10^3 bootstrap resamplings of the MCMC chain.

Results for this conventional dropout approach are shown as blue points and uncertainties in Fig. 6. We see that $\sim 1/2$ of pulsars support the common process found by the remainder of pulsars (i.e., dropout factors > 1), $\sim 1/3$ are insensitive to the common process and yield dropout factors of ~ 1 , and the rest disfavor the presence of a common process with dropout factors < 1 . This behavior is expected from the varying sensitivity of different pulsars; while it seems counter-intuitive that some pulsars disfavor a common process in a dataset that we know contains a GWB signal, one possible explanation is that we have ignored interpulsar correlations, which would otherwise exert greater control in driving agreement between the process amplitudes inferred from different pulsar subsets. This variant of the dropout factor—where interpulsar correlations are modeled—is still being developed and will be presented in Vigeland [53].

We also analyze the dropout factors for this dataset using the FL-ready approach in Eq. (13). As discussed earlier, a combination of the Savage-Dickey technique (for the first ratio of evidences) and FL posterior amplitude representation (for the integral under the posterior of a given pulsar) are needed. There are two metaparameters that we vary in this scheme. The first corresponds to the tolerance on the lowest acceptable number of MCMC samples in the lowest

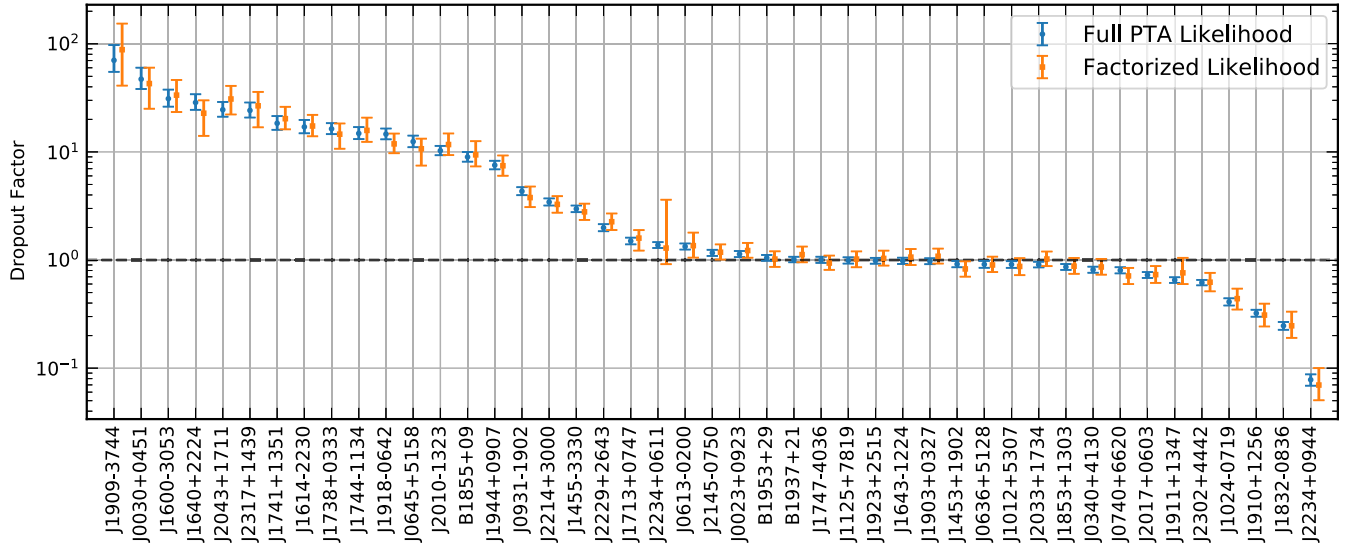


FIG. 6. The dropout factor is the ratio of Bayesian evidences between a model that includes a common process in all pulsars versus one that leaves the common process out of one pulsar. This factor assesses the support that a given pulsar has for a common process that has been found by all other pulsars, acting as a technique for performing leave-one-out cross validation. We compute the dropout factors for all pulsars in one of our simulated datasets that includes a GWB with $A_{\text{GWB}} = 1.91 \times 10^{-15}$. The conventional product-space MCMC sampling approach (blue) is compared to the new FL approach of Eq. (13) (orange), where we see excellent agreement. All uncertainties are derived through bootstrap resampling of MCMC chains.

amplitude bin for the Savage-Dickey Bayes factor calculation of $\mathcal{Z}_p/\mathcal{Z}_{p,0}$. This influences the effective widths of the lowest amplitude bin whose posterior density features in the Bayes factor calculation discussed in Sec. III B 1. We vary this across five values: 10, 25, 50, 75, and 100. The second metaparameter is the number of bins adopted for the histogram estimator of the common process posterior in the FL calculation. In our analyses thus far we found that $N_{\text{bin}} = 100$ was appropriate, however for this dropout factor analysis we allow it to be varied across $N_{\text{bin}} = [10, 25, 50, 75, 100]$. Thus, the dropout factor calculation for each pulsar was performed with 25 distinct metaparameter combinations, and for each of those combinations the MCMC chains of all pulsars were resampled with replacement 1000 times to generate 2.5×10^4 bootstrapped dropout factor estimates. These results are summarized as orange median values with 68% uncertainty regions in Fig. 6. The agreement with the conventional dropout approach is excellent. This FL dropout estimation approach is modular, all calculations are carried out in the MCMC postprocessing stage, and new pulsars can be trivially incorporated.

IV. CONCLUSIONS

We have introduced a new technique for gravitational-wave background (GWB) searches and characterization with pulsar-timing arrays. This technique is parallelized over separate pulsar datasets, making it trivial to expand the array of searched pulsars without requiring exhaustive and repetitive re-analysis of the entire PTA dataset. By leveraging the fact that spectral information about the GWB is—and may remain so even in the post-detection era—dominated by PTA autocorrelations, we can collapse the PTA data covariance matrix to a block-diagonal structure. This is equivalent to approximating the PTA likelihood as a product over individual pulsar likelihoods, each of which includes a stochastic process with a common spectrum across the array. By doing so, each pulsar dataset can be analyzed in parallel, then reduced to a set of sufficient statistical measures for the spectrum of the GWB, which are stitched together in postprocessing. We call this the factorized likelihood technique.

We have considered the case where the GWB is represented by a power-law characteristic strain spectrum, and equivalently a power-law power spectral density in the pulsar timing residuals. This power law has a fiducial spectral index based on the theoretical expectations for the GWB from a population of inspiraling supermassive binary black holes. These supermassive binary black holes are expected to synthesize the dominant GWB signal in the PTA sensitivity band, in excess of other potential cosmological signals (e.g., primordial GWBs, and backgrounds from cosmological phase transitions or cosmic strings). As such, we can fix the spectral index and vary only the amplitude of the GWB in our searches. Therefore, in the analysis of each pulsar, we model noise associated with

intrinsic pulsar and instrumental sources (with corresponding parameters), plus a stochastic process characterized by a fixed-index power law, whose amplitude is searched over. This latter process is a proxy for the GWB in each individual pulsar’s analysis.

Upon deriving the marginal posterior distributions of the GWB amplitude from each pulsar, we process these into histogram estimators that act as sufficient statistics for further analysis. All of the information about the GWB that was originally encoded in the PTA autocorrelations is now distilled into these histograms, which are de facto compressed representations of the PTA data. We found that the PTA GWB amplitude recovery using this FL approach was virtually indistinguishable from the standard PTA likelihood approach, as was the calculation of the Bayes factor for the presence of a common-spectrum process across the PTA. Furthermore, while this approach discards the PTA cross-correlations, it does facilitate the rapid and modular calculation of the PTA cross-correlation S/N ratio, which otherwise requires an additional Bayesian PTA analysis to calibrate noise weightings. The S/N ratio values derived from FL noise weightings versus a full Bayesian PTA analysis match incredibly well. Hence, the FL technique enables fast, parallelizable recovery of the amplitude and cross-correlation significance of a GWB in PTA data, thereby circumventing many of the sampling and computational limitations that PTAs will encounter as data volume grows. This technique has already seen broad uptake within the PTA community for analyzing NANOGrav [30], PPTA [31], EPTA [32], and IPTA [26] flagship datasets and other studies [56] (having been developed by the lead author), but has lacked a formal methodology until now.

There are several further applications that are straightforward when tackled with the FL technique but which would otherwise be challenging using conventional sampling with the PTA likelihood. For example, in almost all PTA analyses the white-noise characteristics of each pulsar are held fixed at values previously found in single-pulsar noise characterization. The reasons for this are that the likelihood calculation is slowed down when sampling white-noise characteristics, and there can be many white-noise parameters per pulsar that lead to a greatly expanded parameter dimensionality. However since the FL technique is parallelized over pulsars, likelihood speed and sampling efficiency remain well within reasonable levels, thereby allowing the uncertainties in these white-noise characteristics to be propagated into the final GWB results. An extension of this concept is that bespoke noise models that account for e.g., additional chromatic influences, can be constructed for each pulsar in parallel, then trivially combined in postprocessing to deliver PTA GWB constraints. Performing this same kind of analysis in a full PTA likelihood analysis is fraught with speed and sampling limitations.

The implementation of the FL technique introduced here does have some caveats that we plan to generalize in

forthcoming work. Most important is that the spectrum of the GWB is assumed to be a power law with a fixed spectral index, and parametrized only by its amplitude. Alternatively, we can model the GWB's spectrum in each pulsar with a free parameter in each frequency bin. Provided that the recovered joint posterior of this Bayesian periodogram from each pulsar can be represented with high fidelity (using e.g., optimized kernel density estimators), these then act as sufficient statistics for more generalized PTA spectral characterization of the GWB beyond the power-law assumption. With an optimally-tuned MCMC algorithm to sample the per-frequency spectral posteriors of each pulsar, this approach should be as fast and trivially parallelizable as the techniques presented in this paper. No significant additional computational time is required beyond the single-pulsar noise analyses that are already performed. This approach is currently being developed. The farthest we could push this FL concept would be to also recover the Fourier coefficients of the modeled GWB signal in each pulsar. By retaining phase information from the pulsar time-series in the form of the Fourier coefficients, we would regain the ability to perform Bayesian cross-correlation analyses, yet with a highly-compressed representation of the original data. The end goal is that the PTA GWB analysis framework would be future proofed against the torrent

of data from ongoing observation campaigns as well as newly discovered pulsars.

ACKNOWLEDGMENTS

We thank our colleagues in NANOGrav and the International Pulsar Timing Array for fruitful discussions and feedback during the development of this technique. We are particularly grateful to Michele Vallisneri for reviewing a draft of our manuscript, and to Xavier Siemens, Sarah Vigeland, and Jeff Hazboun for invaluable insight throughout this project. S. R. T. acknowledges support from NSF Grants No. AST-200793, No. PHY-2020265, No. PHY-2146016, and a Vanderbilt University College of Arts and Science Dean's Faculty Fellowship. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (Awards No. ACI-1532235 and No. ACI-1532236), the University of Colorado Boulder, and Colorado State University. N.P. acknowledges support from the Vanderbilt Initiative for Data-intensive Astrophysics (VIDA). This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, Tennessee.

-
- [1] F. B. Estabrook and H. D. Wahlquist, Response of Doppler spacecraft tracking to gravitational radiation, *Gen. Relativ. Gravit.* **6**, 439 (1975).
 - [2] M. V. Sazhin, Opportunities for detecting ultralong gravitational waves, *Sov. Astron.* **22**, 36 (1978).
 - [3] S. Detweiler, Pulsar timing measurements and the search for gravitational waves, *Astrophys. J.* **234**, 1100 (1979).
 - [4] M. C. Begelman, R. D. Blandford, and M. J. Rees, Massive black hole binaries in active galactic nuclei, *Nature (London)* **287**, 307 (1980).
 - [5] J. S. B. Wyithe and A. Loeb, Low-frequency gravitational waves from massive black hole binaries: Predictions for LISA and pulsar timing arrays, *Astrophys. J.* **590**, 691 (2003).
 - [6] A. H. Jaffe, Observing gravitational radiation with QSO proper motions and the SKA, *New Astron. Rev.* **48**, 1483 (2004).
 - [7] A. Sesana, A. Vecchio, and C. N. Colacino, The stochastic gravitational-wave background from massive black hole binary systems: Implications for observations with Pulsar Timing Arrays, *Mon. Not. R. Astron. Soc.* **390**, 192 (2008).
 - [8] S. Burke-Spolaor, S. R. Taylor, M. Charisi, T. Dolch, J. S. Hazboun, A. M. Holgado, L. Z. Kelley, T. J. W. Lazio, D. R. Madison, N. McMann, C. M. F. Mingarelli, A. Rasskazov, X. Siemens, J. J. Simon, and T. L. Smith, The astrophysics of nanohertz gravitational waves, *Astron. Astrophys. Rev.* **27**, 5 (2019).
 - [9] T. Bogdanovic, M. C. Miller, and L. Blecha, Electromagnetic counterparts to massive black hole mergers, [arXiv:2109.03262](https://arxiv.org/abs/2109.03262).
 - [10] S. Babak and A. Sesana, Resolving multiple supermassive black hole binaries with pulsar timing arrays, *Phys. Rev. D* **85**, 044034 (2012).
 - [11] A. Petiteau, S. Babak, A. Sesana, and M. de Araújo, Resolving multiple supermassive black hole binaries with pulsar timing arrays. II. Genetic algorithm implementation, *Phys. Rev. D* **87**, 064036 (2013).
 - [12] L. Boyle and U.-L. Pen, Pulsar timing arrays as imaging gravitational wave telescopes: Angular resolution and source (de)confusion, *Phys. Rev. D* **86**, 124028 (2012).
 - [13] R. M. Shannon and J. M. Cordes, Assessing the role of spin noise in the precision timing of millisecond pulsars, *Astrophys. J.* **725**, 1607 (2010).
 - [14] J. M. Cordes and R. M. Shannon, A measurement model for precision pulsar timing, [arXiv:1010.3785](https://arxiv.org/abs/1010.3785).
 - [15] M. T. Lam, T. J. W. Lazio, T. Dolch, M. L. Jones, M. A. McLaughlin, D. R. Stinebring, and M. Sumis, On frequency-dependent dispersion measures and extreme scattering events, *Astrophys. J.* **892**, 89 (2020).

- [16] M. T. Lam, J. M. Cordes, S. Chatterjee, M. L. Jones, M. A. McLaughlin, and J. W. Armstrong, Systematic and stochastic variations in pulsar dispersion measures, *Astrophys. J.* **821**, 66 (2016).
- [17] R. S. Foster and D. C. Backer, Constructing a pulsar timing array, *Astrophys. J.* **361**, 300 (1990).
- [18] 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).
- [19] M. A. McLaughlin, The North American Nanohertz Observatory for Gravitational Waves, *Classical Quantum Gravity* **30**, 224008 (2013).
- [20] R. N. Manchester, The Parkes pulsar timing array project, in *Proceedings of the 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More*, edited by C. Bassa, Z. Wang, A. Cumming, and V. M. Kaspi, American Institute of Physics Conference Series Vol. 983 (American Institute of Physics, 2008), pp. 584–592.
- [21] R. N. Manchester *et al.*, The Parkes pulsar timing array project, *Pub. Astron. Soc. Aust.* **30**, e017 (2013).
- [22] M. Kramer and D. J. Champion, The European pulsar timing array and the large European array for pulsars, *Classical Quantum Gravity* **30**, 224009 (2013).
- [23] B. C. Joshi *et al.*, Precision pulsar timing with the ORT and the GMRT and its applications in pulsar astrophysics, *J. Astrophys. Astron.* **39**, 51 (2018).
- [24] J. P. W. Verbiest *et al.*, The international pulsar timing array: First data release, *Mon. Not. R. Astron. Soc.* **458**, 1267 (2016).
- [25] B. B. P. Perera *et al.*, The international pulsar timing array: Second data release, *Mon. Not. R. Astron. Soc.* **490**, 4666 (2019).
- [26] 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).
- [27] K. J. Lee, Prospects of gravitational wave detection using pulsar timing array for Chinese future telescopes, in *Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015*, edited by L. Qain and D. Li, Astronomical Society of the Pacific Conference Series Vol. 502 (2016), p. 19.
- [28] C. Ng, Pulsar science with the CHIME telescope, *IAU Symp.* **13**, 179 (2017).
- [29] M. Bailes *et al.*, MeerTime—the MeerKAT Key science program on pulsar timing, *Proc. Sci., MeerKAT2016* (2018) 011 [arXiv:1803.07424].
- [30] Z. Arzoumanian *et al.* (Nanograv Collaboration), The NANOGrav 12.5 yr data set: Search for an isotropic stochastic gravitational-wave background, *Astrophys. J. Lett.* **905**, L34 (2020).
- [31] 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).
- [32] 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).
- [33] H. Middleton, A. Sesana, S. Chen, A. Vecchio, W. Del Pozzo, and P. A. Rosado, Massive black hole binary systems and the NANOGrav 12.5 yr results, *Mon. Not. R. Astron. Soc.* **502**, L99 (2021).
- [34] J. D. Romano, J. S. Hazboun, X. Siemens, and A. M. Archibald, Common-spectrum process versus cross-correlation for gravitational-wave searches using pulsar timing arrays, *Phys. Rev. D* **103**, 063027 (2021).
- [35] N. S. Pol *et al.* (Nanograv Collaboration), Astrophysics milestones for pulsar timing array gravitational-wave detection, *Astrophys. J. Lett.* **911**, L34 (2021).
- [36] M. F. Alam *et al.* (Nanograv Collaboration), The NANOGrav 12.5 yr data set: Observations and narrowband timing of 47 millisecond pulsars, *Astrophys. J. Suppl. Ser.* **252**, 4 (2021).
- [37] 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).
- [38] R. van Haasteren and M. Vallisneri, New advances in the Gaussian-process approach to pulsar-timing data analysis, *Phys. Rev. D* **90**, 104012 (2014).
- [39] R. van Haasteren and M. Vallisneri, Low-rank approximations for large stationary covariance matrices, as used in the Bayesian and generalized-least-squares analysis of pulsar-timing data, *Mon. Not. R. Astron. Soc.* **446**, 1170 (2015).
- [40] 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).
- [41] E. S. Phinney, A practical theorem on gravitational wave backgrounds, [arXiv:astro-ph/0108028](https://arxiv.org/abs/astro-ph/0108028).
- [42] This is a conservative choice for simulations, given that NANOGrav and other PTAs are constantly adding new pulsars. However, the addition of new pulsars is irrelevant here since we are only interested in creating a realistic set of simulations to assess the relative efficacy of the analysis approaches.
- [43] D. W. Scott, On optimal and data-based histograms, *Biometrika* **66**, 605 (1979).
- [44] S. J. Sheather and M. C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc.* **53**, 683 (1991).
- [45] C. Tiburzi, G. Hobbs, M. Kerr, W. A. Coles, S. Dai, R. N. Manchester, A. Possenti, R. M. Shannon, and X. P. You, A study of spatial correlations in pulsar timing array data, *Mon. Not. R. Astron. Soc.* **455**, 4339 (2016).
- [46] 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).
- [47] 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).
- [48] 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).
- [49] J. A. Ellis, X. Siemens, and R. van Haasteren, An efficient approximation to the likelihood for gravitational wave stochastic background detection using pulsar timing data, *Astrophys. J.* **769**, 63 (2013).
- [50] This Bayesian search for a common process, followed by computing the NMOS for evidence of cross-correlations, is still much computationally cheaper than a Bayesian search that models cross-correlations.
- [51] H. Wang, S. R. Taylor, and M. Vallisneri, Bayesian cross validation for gravitational-wave searches in pulsar-timing array data, *Mon. Not. R. Astron. Soc.* **487**, 3644 (2019).
- [52] K. Aggarwal *et al.* (NANOGrav Collaboration), The NANOGrav 11 yr data set: Limits on gravitational waves from individual supermassive black hole binaries, *Astrophys. J.* **880**, 116 (2019).
- [53] S. Vigeland (to be published).
- [54] Z. Arzoumanian *et al.* (NANOGrav Collaboration), The NANOGrav 11 year data set: Pulsar-timing constraints on the stochastic gravitational-wave background, *Astrophys. J.* **859**, 47 (2018).
- [55] S. R. Taylor, R. van Haasteren, and A. Sesana, From bright binaries to bumpy backgrounds: Mapping realistic gravitational wave skies with pulsar-timing arrays, *Phys. Rev. D* **102**, 084039 (2020).
- [56] A. D. Johnson, S. J. Vigeland, X. Siemens, and S. R. Taylor, Gravitational wave statistics for pulsar timing arrays: Examining bias from using a finite number of pulsars, [arXiv:2201.10657](https://arxiv.org/abs/2201.10657).