Efficient pipeline for joint gravitational wave searches from individual binaries and a gravitational wave background with Hamiltonian sampling

Gabriel E. Freedman and Sarah J. Vigeland

Center for Gravitation, Cosmology and Astrophysics, University of Wisconsin–Milwaukee,
P.O. Box 413, Milwaukee, Wisconsin 53201, USA

The pulsar timing array community has recently reported the first evidence of a low-frequency stochastic gravitational wave background. With longer observational time spans we expect to be able to resolve individual gravitational wave sources in our data alongside the background signal. The statistical modeling and Bayesian searches for such individual signals is a computationally taxing task that is the focus of many different avenues of methods development. We present a pipeline for performing efficient joint searches for gravitational waves originating from individual supermassive black hole binaries as well as a gravitational wave background using a Hamiltonian Monte Carlo sampling scheme. Hamiltonian sampling proposes samples based on the gradients of the model likelihood, and can both converge faster to more complicated and high-dimensional distributions as well as efficiently explore highly covariant parameter spaces such as the joint gravitational wave background and individual binary model. We show the effectiveness of our scheme by demonstrating accurate parameter estimation for simulated datasets containing low- (6 nHz) or high- (60 nHz) frequency binary sources. Additionally we show that our method is capable at more efficiently generating skymaps for individual binary sources—maps displaying the upper limits on the gravitational wave strain of the source, h_0, as a function of sky location—by sampling over larger portions of the full sky. Comparing against results for the NANOGrav 12.5-year dataset, we find similar reconstructed upper limits on the gravitational wave strain while simultaneously reducing the number of required analyses from 72 independent binned searches down to a single run.

I. INTRODUCTION

The first evidence of a low-frequency stochastic gravitational wave (GW) signal reported [1–4] by the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [5], European Pulsar Timing Array [6], Indian Pulsar Timing Array [7], Parkes Pulsar Timing Array [8], and Chinese Pulsar Timing Array [4] has opened a new chapter in the field of GW astrophysics. Pulsar timing array (PTA) [9–11] collaborations search for nHz frequency GWs by analyzing the times-of-arrival (TOAs) of radio pulses emitted by millisecond pulsars. By regularly observing such pulsars over a decades-long time span PTAs can reach the sensitivity necessary to probe the nHz band. The recently identified stochastic GW signal displayed, to varying levels of significance, the expected Hellings-Downs (HD) [12] spatial correlation signature between pulsars that is indicative of the signal being a gravitational wave background (GWB).

One possible source describing the nHz GWB is the collective signal from the population of supermassive black hole binaries (SMBHBs) present in the observable universe [13]. All massive galaxies hold a supermassive black hole, typically of mass 10^6–$10^{10} M_\odot$, at their centers [14].

Galactic merger events consequently lead to the formation of SMBHB systems. When the component black holes reach inspiral phase, the emission of GWs becomes the dominant force behind the system's evolution. To date there have been no confirmed observations of SMBHBs at subparsec separations. With the discovery of a GWB signal, a next major step for PTA science is to search for particularly loud individual binaries that may be detected among the stochastic ensemble within the next decade [15–18]. Measurements of GWs from individual sources, colloquially referred to as continuous waves (CWs) due to their minimal frequency evolution, would provide useful constraints on the astrophysical environments of SMBHBs [19,20] and could be coupled with electromagnetic observations to study galactic evolution and further multimessenger astrophysical research (e.g. [21]).

Single source searches prove inherently more computationally taxing than a comparable GWB analysis. Modeling GWs from an individual SMBHB adds to the already large PTA parameter space, and such parameters bring covariances among themselves as well as with other red-noise processes present in the data. The computational cost additionally compounds with increased data volume, which includes longer observation span and new pulsars added to

the array. This complication is particularly apparent for efforts at combining datasets from multiple PTAs, yielding highly sensitive yet computationally overwhelming data products. Multiple techniques at exploring the CW parameter space have been developed [22–26], and recent improvements have led to a 100-fold speed up of the full analysis through the use of a tailored likelihood calculation [26].

In this paper we detail an additional procedure for achieving efficient CW searches through a Monte Carlo routine established through sample proposals based in the gradient of the model likelihood. This utilizes a Hamiltonian Monte Carlo (HMC) [27,28] sampler to replace the random-walk nature of traditional MCMC methods with a simulation of Hamiltonian dynamics on the target probability distribution. The algorithm concentrates on drawing subsequent samples at much further distances in the multidimensional parameter space, trading pure speedups of the likelihood calculation for higher sample acceptance rates and an efficient exploration of the distribution. Within the realm of PTA science, HMC was first utilized in the development of a model-independent approach to Bayesian inference with PTA data [29], and soon after to the task of outlier removal in single-pulsar data. In a previous paper [30], we demonstrated the effectiveness of using HMC to perform Bayesian GWB searches with the full marginalized PTA likelihood. Here we extend the methods and previous results to allow for joint inference of individual binary sources and common-process signals.

This paper is outlined as follows. In Sec. II we review the signal model of a single binary and describe the current Bayesian formalism for searching for such sources in the context of PTA data. We describe the Hamiltonian Monte Carlo sampling procedure and introduce a new pipeline, predicated on this algorithm, for performing the analyses in a more efficient manner. In Sec. III we validate this pipeline against a suite of simulated PTA datasets. We assess the efficiency of this new analysis prescription on the NANOGrav 12.5-year dataset in Sec. IV. Lastly, in Sec. V we summarize and discuss opportunities for future development of this work.

II. METHODOLOGY AND SOFTWARE

Here we provide a brief overview of the data, PTA signal model, and likelihood function used in this paper, as well as characterize the GW signal for an individual binary. We then describe the HMC algorithm, and introduce our code and pipeline tailored to applying this method to CW searches.

A. PTA likelihood

First we discuss pulsar timing data and the structure of the PTA likelihood. Pulsar observational data exists in the

form of pulse times-of-arrival (TOAs). After subtracting from each pulsar's TOAs a timing model comprised of parameters such as proper motion, parallax, spin period, spin period derivative, and other orbital parameters, we are left with timing residuals δt that we can characterize as a linear combination of noise sources and GW signals

$$\delta \mathbf{t} = M\boldsymbol{\varepsilon} + \mathbf{n}_{\mathrm{RN}} + \mathbf{n}_{\mathrm{CRN}} + \mathbf{n}_{\mathrm{WN}} + \mathbf{s}. \qquad (1)$$

The first term $M\boldsymbol{\varepsilon}$ represents inaccuracies originating from subtracting the linearized timing model solution. Next the term $\mathbf{n}_{\mathrm{RN}}$ denotes effects due to low-frequency "red" noise that are intrinsic to each pulsar. The following term $\mathbf{n}_{\mathrm{CRN}}$ again describes red-noise sources, but this time specifically references sources that are common among all of the pulsars, including for example a GWB. Here we model the common spectrum process with a fiducial power-law spectrum with a characteristic strain h_c and cross-power spectral density S_{ab}:

$$h_c(f) = A_{\mathrm{gw}} \left(\frac{f}{f_{yr}} \right)^{\alpha}, \qquad (2)$$

$$S_{ab}(f) = \Gamma_{ab} \frac{A_{\mathrm{gw}}^2}{12\pi^2} \left(\frac{f}{f_{\mathrm{yr}}} \right)^{-\gamma} f_{\mathrm{yr}}^{-3}, \qquad (3)$$

where the spectral index $\gamma = 3 - 2\alpha$. In the case where the common-process signal represents a background generated by the GW emission from a population of inspiraling SMBHBs in circular orbits, we have $\alpha = -2/3$ and $\gamma = 13/3$ [31]. The function Γ_{ab}, called the overlap reduction function (ORF), defines the average correlations between a set of two pulsars a and b based on their relative angular separation. For common uncorrelated red-noise (CURN) processes the ORF is equal to 1. For an isotropic, stochastic GWB it is given by the Hellings-Downs correlation function [12]

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

with $x_{ab} = (1 - \cos \xi_{ab})/2$ for an angular separation ξ_{ab} between two pulsars. Following the common-process noise signals is a term $\mathbf{n}_{\mathrm{WN}}$ encoding all high-frequency "white" noise sources present in the data, including constant multiplicative correction factors to TOA uncertainties (EFAC), additional noise added in quadrature (EQUAD), and observational epoch-correlated noise (ECORR) that is uncorrelated across separate epochs. Lastly the vector $\mathbf{s}$ represents the component of the timing residuals caused by additional deterministic signals. Here we treat $\mathbf{s}$ as the signal induced by an individual binary, and is outlined in more detail in Sec. II B.

Finally, we construct the form of the PTA likelihood for use in our Bayesian inference pipelines. First we

dramatically reduce the dimensionality of our posterior by marginalizing over the timing model parameters [29,32]. Then by constructing the total PTA covariance matrix $C = N + TBT^T$, with N the white noise covariance matrix, T the design matrix for the timing model, red noise, and ECORR signals, and B the prior covariance matrix for those three sets of parameters, we can state the multivariate Gaussian likelihood function used for the analyses in this paper:

$$L(\delta \mathbf{t}|\mathbf{q}) = \frac{\exp\left(-\frac{1}{2}(\delta \mathbf{t} - \mathbf{s})^T C^{-1}(\delta \mathbf{t} - \mathbf{s})\right)}{\sqrt{\det 2\pi C}}, \quad (5)$$

where $\mathbf{q}$ denotes the vector of varying parameters existing in our model in both deterministic signals $\mathbf{s}$ as well as the total noise covariance matrix C.

B. CW signal

We now review the signal model for GWs originating from an SMBHB and their effect on PTA residuals. The GW signal can be written as

$$s(t, \hat{\Omega}) = F^+(\hat{\Omega})\Delta s_+(t, \hat{\Omega}) + F^\times(\hat{\Omega})\Delta s_\times(t, \hat{\Omega}), \quad (6)$$

with the scripts $\{+\times\}$ denoting the plus and cross polarization modes, respectively, the two tensor polarizations allowed by general relativity, and $\hat{\Omega}$ is a unit vector pointing from the GW source to the solar system barycenter. The functions F^+ and $F^\times$ represent the antenna pattern functions that describe the response of a given pulsar to the emitting source, and are composed of the binary sky location polar and azimuthal angles θ and ϕ, respectively, and GW polarization angle ψ (for a complete description, see [33]. The terms $\Delta s_{+,\times}(t)$ account for the fact that the Earth and pulsar see the induced GW signal at different times in the binary evolution, and therefore define the difference between the "pulsar-term" and "Earth-term"

$$\Delta s_{+,\times}(t, \hat{\Omega}) = s_{+,\times}(t_p) - s_{+,\times}(t), \quad (7)$$

where t_p is the time measured at the pulsar and t the time measured at the solar system barycenter. The two times are geometrically related according to:

$$t_p = t - L(1 + \hat{\Omega} \cdot \hat{u}), \quad (8)$$

where L is the distance to the pulsar and $\hat{u}$ represents a line of sight vector to the pulsar. For the analyses present in the remainder of the paper, we focus only on searching for the Earth-term component of the full signal:

$$s_E(t, \hat{\Omega}) = F^+(\hat{\Omega})s_+(t) + F^\times(\hat{\Omega})s_\times(t). \quad (9)$$

The exact forms of $s_{+,\times}(t)$ for a circular binary are given, to zeroth post-Newtonian (0-PN) order, by

$$s_+(t) = -\frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \sin 2\Phi(t)(1 + \cos^2 \iota), \quad (10)$$

$$s_\times(t) = \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} 2 \cos 2\Phi(t) \cos \iota. \quad (11)$$

The parameter $\mathcal{M}$ represents the binary chirp mass $\mathcal{M} \equiv (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ for the component black hole masses m_1 and m_2. The parameters d_L and ι are the luminosity distance to the binary and the source inclination angle, respectively. The time-dependent angular frequency and phase functions are, for reference Earth-term frequency ω_0 and phase Φ_0

$$\omega(t) = \omega_0 \left[1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}(t - t_0)\right]^{-3/8}, \quad (12)$$

$$\Phi(t) = \Phi_0 + \frac{1}{32}\mathcal{M}^{-5/3}\left[\omega_0^{-5/3} - \omega(t)^{-5/3}\right]. \quad (13)$$

Additionally, one can define the overall strain amplitude, h_0, as

$$h_0 = \frac{2\mathcal{M}^{5/3}(\pi f_{\text{GW}})^{2/3}}{d_L}, \quad (14)$$

with the GW frequency f_{GW} related to the initial angular frequency ω_0 by $\omega_0 = \pi f_{\text{GW}}$. We note that Eq. (14) shows a degeneracy between h_0, $\mathcal{M}$, f_{GW}, and d_L, allowing us to choose three of those four quantities when constructing our complete parameter vector. In practice we typically exclude the luminosity distance in favor of h_0, $\mathcal{M}$, and f_{GW}. The full Earth-term CW source is therefore completely parametrized by $(\theta, \phi, \iota, \psi, \Phi_0, h_0, \mathcal{M}_c, f_{\text{GW}})$.

In order to further elucidate the complications in sampling joint CW and common-process models, we compute Eq. (5) for an individual binary and particular noise realization. We plot the Earth-term only likelihood surface in Fig. 1 as a function of the CW sky location parameters. The surface displays highly nontrivial structure and demonstrates some of the difficulties in efficiently sampling the full parameter space. There are numerous local extrema where a random-walk MCMC sampler is liable to get trapped and be unable to fully explore the full posterior. This highlights the need for more sophisticated sampling routines and corresponding pipelines.

C. Hamiltonian Monte Carlo sampling

The HMC algorithm [27,28], an extension of the traditional Metropolis-Hastings technique [34], tackles the problem of sampling high-dimensional and covariant state spaces by using Hamiltonian dynamics to generate proposal states that are distant relative to each other in the parameter space. Compared to a standard Gaussian proposal
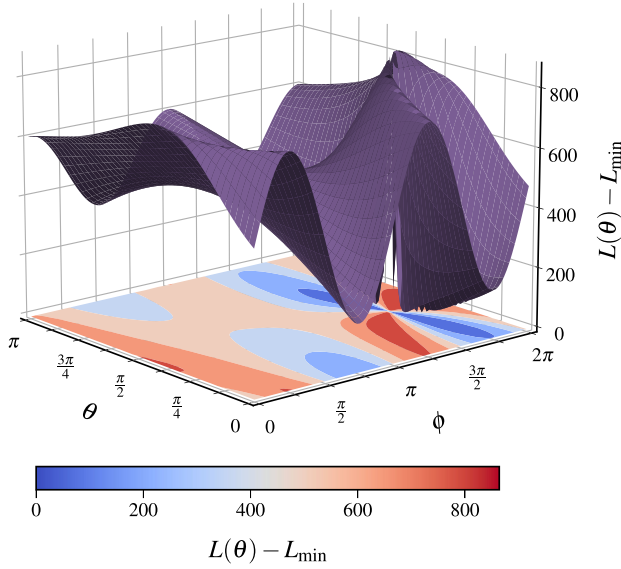
FIG. 1.   Likelihood surface for an Earth-term only CW signal as a function of sky position. The x and y axes represent θ and ϕ for the source, respectively. The z axis shows the PTA log-likelihood function evaluated at a particular (θ, ϕ), then subtracting off the minimum log-likelihood value for the grid. On the plane $z = 0$ we plot a 2D color map contour of the surface. We see that the contours of the likelihood surface have many sharp peaks and valleys, indicating difficult regions of parameter space to sample over.

scheme, this reduces the overall correlation in the Markov chain while maintaining a high sample acceptance rate. It proceeds by first introducing an auxiliary momentum vector $\mathbf{p}$ alongside the model parameters $\mathbf{q}$. The Hamiltonian to be simulated is the log of the joint density of $\mathbf{p}$ and $\mathbf{q}$, which can be separated into a potential energy term $U(\mathbf{q})$ and kinetic energy term $K(\mathbf{p})$

$$H(\mathbf{p}, \mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}) = -\mathcal{L}(\mathbf{q}) + \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}, \quad (15)$$

where $\mathcal{L}(\mathbf{q})$ is the log of the likelihood function for the target parameter distribution, and M a "mass matrix" typically taken as the identity. The evolution of this system through time can then be simulated by numerically solving Hamilton's equations

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}}, \qquad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}}. \quad (16)$$

This integration proceeds for a set number of steps L, typically accomplished through a second-order symplectic "leapfrog" integrator, and ends by proposing some final position and momentum states.

The relative performance of the HMC algorithm is then determined by two factors: the computational cost of

calculating the gradient of the log likelihood function for the target distribution, and the proper tuning of two user-defined parameters: the number of steps L and integration step size ϵ. The first factor is driven entirely by the complexity of the model in question, and whether or not the log likelihood gradient can be computed exactly or requires numerical differentiation. The second factor can be resolved by automatically tuning the two extra parameters through the use of a No-U-Turn Sampler [35]. This sampler uses a recursive doubling algorithm to build a binary tree of sample proposals, simulating Hamiltonian dynamics either forward or backward in time at random for 2^j iterations with j denoting the height of the tree. The process continues until the position-momenta pairs $\mathbf{q}^+$, $\mathbf{p}^+$ and $\mathbf{q}^-$, $\mathbf{p}^-$ of the left- and rightmost leafs of the tree satisfy the condition:

$$(\mathbf{q}^+ - \mathbf{q}^-) \cdot \mathbf{p}^- < 0 \quad \text{or} \quad (\mathbf{q}^+ - \mathbf{q}^-) \cdot \mathbf{p}^+ < 0. \quad (17)$$

In effect this monitors the trajectory of proposals and stops when it begins to double back on itself, or make a "U turn." At this point a sample is chosen at random from the tree and accepted or rejected according to the Metropolis algorithm [34].

D. Software

Our new code, freely and publicly available on GitHub under the package ETUDES,[1] includes an analysis suite capable of performing HMC sampling with PTAs. Although this paper focuses on joint searches for a CW signal and common red-noise process, the modularity of the code allows for the addition of a wide array of other GW sources of interest, such as multiple binaries [36,37], GW memory [38], eccentric binaries [39–42], or advanced pulsar noise modeling [43,44]. It can also be natively run on GPUs, drastically dropping the runtime of CW analyses to timescales of hours for simulated data and days for production data. The code is under further active development to accommodate other searches of interest.

While there are renewed efforts toward utilizing hierarchical modeling for PTAs [45], This work solely uses the marginalized PTA likelihood, meaning that we need not apply coordinate transformations, such as a decentered reparameterization, designed to deal with Markov chain mixing rates and other sampling issues commonly associated with hierarchical funneling. Instead all analyses here, and the default setup for our pipeline, use a single change of coordinates known as an interval transform. This maps all model parameters from their default prior ranges $q \in [a, b]$ to the real line $q' \in (-\infty, \infty)$ via:

$$q' = \log\left(\frac{q-a}{b-q}\right), \qquad (18)$$

$$q = \frac{(b-a)\exp(q')}{1+\exp(q')}, \qquad (19)$$

where we use the Jacobian dq'/dq and its reciprocal to convert between the original and transformed probability spaces.

Sampling with HMC necessarily requires taking derivatives of the model likelihood. We accomplish this by writing the PTA likelihood and its components computations entirely with JAX [46], allowing us to use automatic differentiation to calculate gradients. To do so we decouple the entirety of the PTA computation from NANOGrav's analysis suite ENTERPRISE [47], though we do make use of the code's data structures for holding per-pulsar TOAs, residuals, and other timing model information. We utilize the implementation of the NUTS algorithm present in the BLACKJAX [48] package. All simulated PTA datasets are created using LIBSTEMPO [49].

III. SIMULATED DATA STUDY

First in order to gauge the accuracy of our pipeline and demonstrate its consistency in parameter estimation we created and analyzed a collection of simulated datasets. All datasets comprise identical TOAs, uncertainties, and timing model solutions to the NANOGrav 12.5-year dataset [50]. This constitutes 45 pulsars in total, all with an observational baseline of at least 3 years.

Each individual dataset contains the same per-pulsar noise injections. We simulated "white-noise" signals, typically instrumental noise that dominates at high frequencies, at their maximum likelihood values obtained from separate individual pulsar noise analyses. Low-frequency "red-noise" signals, representing noise intrinsic to each pulsar, were simulated again by referencing the same individual noise runs. We injected the intrinsic pulsar noise at frequencies spanning from $1/T_{\rm psr}$ up to $30/T_{\rm psr}$, with $T_{\rm psr}$ denoting the observational time span of each pulsar.

The NANOGrav 12.5-year dataset contained a CURN process with a Bayes factor in excess of 10,000 relative to a model with only intrinsic pulsar noise [51]. Therefore, for the most accurate prescription of a realistic PTA dataset, we also include a similar process in all of our simulations. The most recent dataset reported evidence for this process containing HD correlations, though we do not consider that in this study. We inject a CURN signal characterized by an amplitude $A_{\rm CURN} = 2 \times 10^{-15}$ and spectral index $\gamma_{\rm CURN} = 4.33$, in line with the expected power and shape of the spectrum.

On top of the various noise models, we inject CW signals. We choose three instances with which to create our data: a low-frequency source, a high-frequency source, and a dataset with no source injection. In all cases we inject only the Earth-term signal. The low-frequency dataset contains a binary emitting GWs at frequency $f_{\rm GW} = 6$ nHz and an amplitude chosen to achieve a moderately high signal-to-noise ratio (SNR) of 10.8. For the case of the high-frequency dataset, we include a binary emitting GWs at $f_{\rm GW} = 60$ nHz with an SNR of 9.3. In both cases the SNR is calculated as:

$$\text{SNR} = \sqrt{(s|s)} = \sqrt{s^T C^{-1} s}, \qquad (20)$$

where s is the template waveform and C is the same noise covariance matrix present in Eq. (5). This can also be considered the expected SNR that is independent from any particular noise realization. The dataset without any CW injection allows us to verify the ability of our methods to place upper limits on source properties in the absence of a detection. For the purposes of validating our pipeline, we create 100 simulated datasets with both the 6 and 60 nHz injection properties. The high- and low-frequency source properties remain fixed across their respective simulations. This allows us to test our methods across numerous noise realizations.

Next we outline the basic procedure for setting up our models before performing Bayesian inference through our HMC pipeline. Rather than simultaneously search over the hundreds of white-noise parameters, we fix them to their maximum likelihood values used in creating the datasets, a commonplace procedure in production-level PTA analyses. We model the pulsar intrinsic red-noise with a power-law power spectral density (PSD) defined by an amplitude $\log_{10} A_{\rm red} \in U[-18, -11]$ and spectral index $\gamma_{\rm red} \in U[0, 7]$. Additionally we search over the two parameters characterizing the CURN process, using priors of $A_{\rm CURN} \in U[-18, -12]$ and $\gamma_{\rm CURN} \in U[0, 7]$. When modeling the CW signal all parameters are given uniform priors. In the case of upper limit analyses, the prior on $\log_{10} h_0$ is shifted from uniform in log space $\log_{10} h_0 \in U[-18, -12]$ to uniform in linear space $\log_{10} h_0 \in U[10^{-18}, 10^{-12}]$. The coordinate space outlined by these priors is later transformed via the procedure outlined in Sec. II D prior to beginning the inference.

Lastly we benchmark the speed and efficiency of both the ETUDES pipeline and comparable run with ENTERPRISE through a pilot inference run on one of the simulated 6 nHz injection datasets. The full joint CURN and CW search here constitutes 100 free parameters ($2N_{\rm psr}$ intrinsic red-noise parameters for 45 pulsars, 2 parameters for the CURN, and 8 describing the CW signal model). For the traditional MCMC pipeline with ENTERPRISE the average likelihood evaluation time is 200 ms on a 12-core Intel(R) Xeon(R) E5-2680 v3 processor. Using the same CPU, the average likelihood and gradient evaluation times with ETUDES is 170 ms and 3.7 s, respectively, and on an NVIDIA Tesla A100 GPU they are 17 ms and 1.6 s,

respectively. All of the above results are then scaled by the average autocorrelation lengths of the corresponding MC chains to calculate the timescales of statistically independent sample generation. This gives an estimate of 450 s to get an independent sample with ENTERPRISE compared to 52 s for runs on a CPU and 25 s on a GPU for the HMC pipeline. Overall, Hamiltonian sampling provides an increase of roughly an order of magnitude in computational efficiency.

A. Low-frequency (6 nHz) signal

Previous NANOGrav CW searches have consistently shown that PTAs are most sensitive to single sources at the lower end ($\sim$1–20 nHz) of their frequency ranges. This also happens to be where the GWB, and more generally any CURN process, is at its strongest. With evidence for a GWB now in hand, it is important for all future CW searches to be capable of dealing with the covariance between the common signal and any low-frequency single sources. We first analyzed a signal with a frequency of $f_{GW} = 6$ nHz, which places it near the peak sensitivity of NANOGrav PTA. The chirp mass $\mathcal{M} = 10^9 M_\odot$ and luminosity distance $d_L = 21.8$ Mpc of the source are chosen so that the GW amplitude gives an SNR of 10.8. Additionally we place the source close to the most sensitive sky location at $(\theta, \phi) = (2\pi/3, 3\pi/2)$. Lastly, the parameters $(\iota, \psi, \Phi_0) = (3\pi/4, \pi/3, 3\pi/2)$ define the source's inclination, polarization, and initial phase. Together, all of the above allows to fully classify our injected signal.

Taking the resulting chains from our analyses, we plot both the one- and two-dimensional posterior distributions for all eight binary parameters in Fig. 2. All parameters have their true injected values lying within their respective posteriors. Both the GW frequency and amplitude distributions are tightly constrained. The posterior for the chirp mass remains entirely unconstrained as we expect for Earth-term only searches and sources with slow frequency evolution. The sky location of the source is very well localized to its true value. The initial phase and polarization angles display a set of multimodal posteriors, which we can efficiently sample but are unable to break the multimodality.

B. High-frequency (60 nHz) signal

The long-term prospects of CW detection play crucial role in multi-messenger analyses and astrophysical interpretation of SMBHB populations and sources, and it is important that we have the ability to do accurate parameter estimation on possible binary candidates. With this in mind, we analyzed a signal with a GW frequency of $f_{GW} = 60$ nHz, chosen to closely mimic that of the potential SMBHB candidate 3C 66B [42,52,53]. The remaining parameters describing the source properties and sky location are $(\theta, \phi, \iota, \psi, \Phi_0, \mathcal{M}_c, d_L) = (2\pi/3, 3\pi/2, 3\pi/4, \pi/3, 3\pi/2, 10^9 M_\odot, 91.1$ Mpc). At frequencies this high the CURN is very weak and therefore

we do not have to worry with covariances between the common-process and binary signals.

In Fig. 3 we plot the posterior distributions for the eight binary parameters for this model, similar to Fig. 2. Again we find that we are able to efficiently sample the entire CW parameter space alongside both a CURN process as well as all intrinsic pulsar noise. We see the same structure in the nearly all of our posteriors: the source GW frequency, GW amplitude, and sky location are very tightly constrained, and the multimodal structure in the polarization angle and initial phase persist. Most importantly, all injected values once again fall squarely within their 1D posteriors. One notable difference is the emerging constraint on the binary chirp mass. High-mass binaries emitting at this frequency should show significant evolution over the 12.5-year observing window of our simulated datasets. Consequently we find across the 100 realizations of the data that we can place an upper limit on the binary chirp mass. In 30 of the realizations, the chirp mass posterior was less constrained than what the frequency evolution would predict.

C. Parameter estimation consistency

As a final test of our method's effectiveness with simulated data, we explore the capacity of its statistical coverage across many noise realizations of the same underlying data. First we create 100 iterations of our $f_{GW} = 6$ nHz dataset. Next we run standard Bayesian searches on all datasets with our HMC pipeline. Lastly, to check the consistency of parameter recovery for our pipeline, we consider across all 100 sets of posteriors whether if in $p\%$ of the realizations the injected parameter values fall within the $p\%$ credible region.

The results of this analysis, called a $p - p$ plot, are summarized in Fig. 4. We plot lines for CW sky location parameters, $\log_{10} h_0$, $\log_{10} f_{GW}$, and the CURN amplitude and spectral index. The dotted gray lines represent 1σ, 2σ, and 3σ confidence intervals. All parameters fall largely within the 3σ boundary indicating an unbiased recovery of the injected values. The chirp mass, being entirely unconstrained across all realizations due to the minimal evolution of the particular signal, was left out off this figure. The cosine of the binary inclination was also largely unconstrained across all realizations and was likewise excluded.

IV. ANALYSIS OF REAL PTA DATA

Ultimately we want to validate our methods against real data and published results. We use the full NANOGrav 12.5-year dataset [50] to benchmark our analysis, and focus on the particular challenge of creating sensitivity sky maps. Given the anisotropic distribution across the sky of the pulsars in our array, it is important to quantify how our observing limits change in different areas. The sky maps
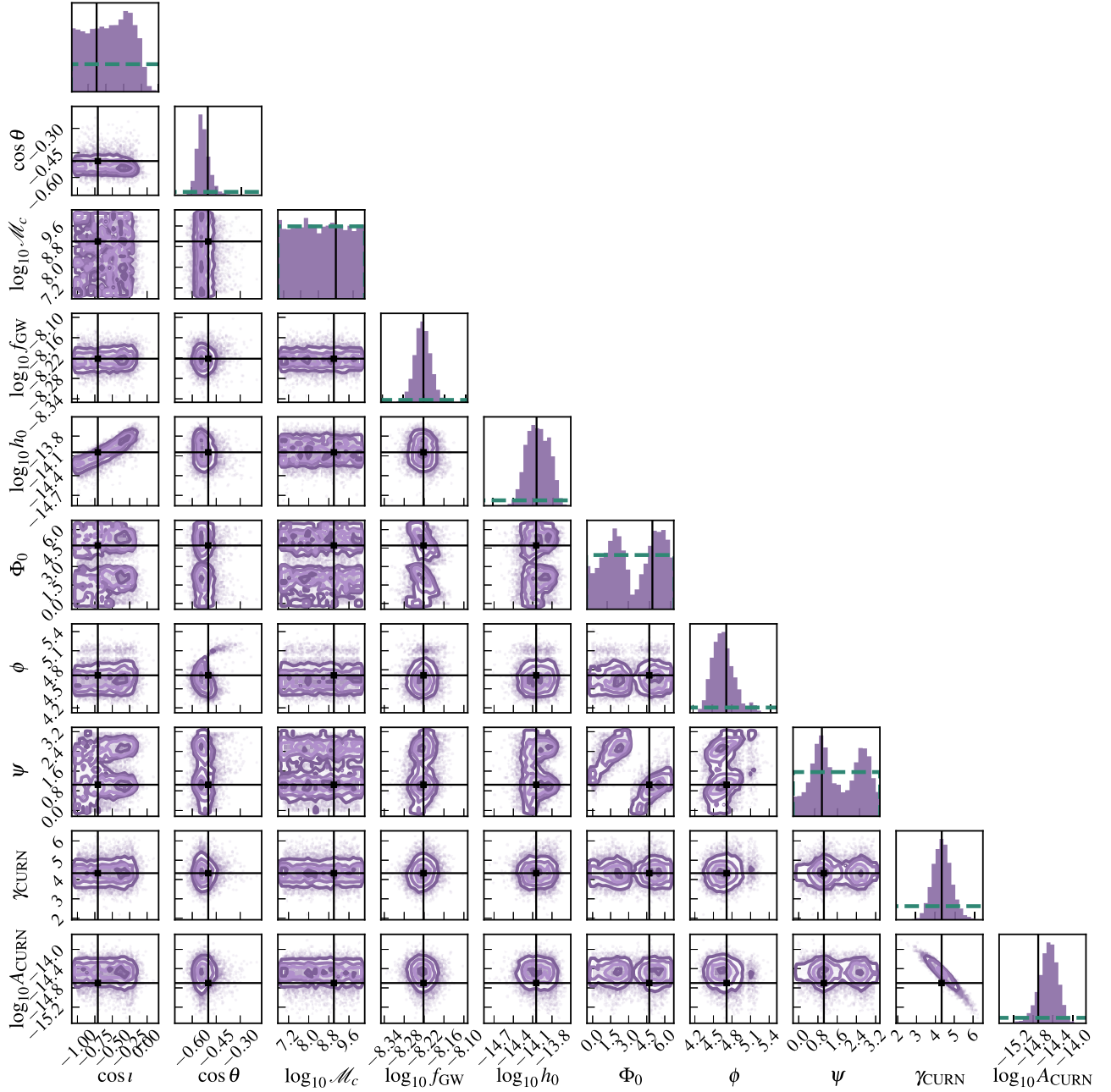
FIG. 2. 1D and 2D posterior distributions for the eight parameters describing a SMBHB signal emitting GWs at $f_{\mathrm{GW}} = 6$ nHz at an SNR of 10.8. The true values of the injected parameters are shown as solid black lines, and the priors are plotted on the 1D histograms as horizontal, green dashed lines. All true values fall within their posteriors, with the sky location, GW frequency, and GW strain parameters being tightly constrained. This demonstrates the capability of the HMC pipeline in accurate parameter estimation for full CW searches.

typically describe, for a given GW frequency, the 95% upper limits on h_0 as a function of sky location. The typical strategy for generating the plots is to bin the sky into 768 separate pixels and run an MC analysis on each individual partition. This dense pixelation is due in part to our inability to get similar number of MCMC samples across the full parameter space in an all-sky search.

We analyze a CW model including a CURN process for sky locations bounded by $\theta \in [\pi/2, 3\pi/4]$, $\phi \in [3\pi/2, 2\pi]$. The bounds were chosen so as to include the most sensitive sky location from the NANOGrav 12.5-year CW analysis [33], at an RA of $19^{\mathrm{h}}07^{\mathrm{m}}30^{\mathrm{s}}$ and a declination angle of $30°00'00''$. This range of parameter space corresponds to 72 distinct pixels, and therefore typically 72
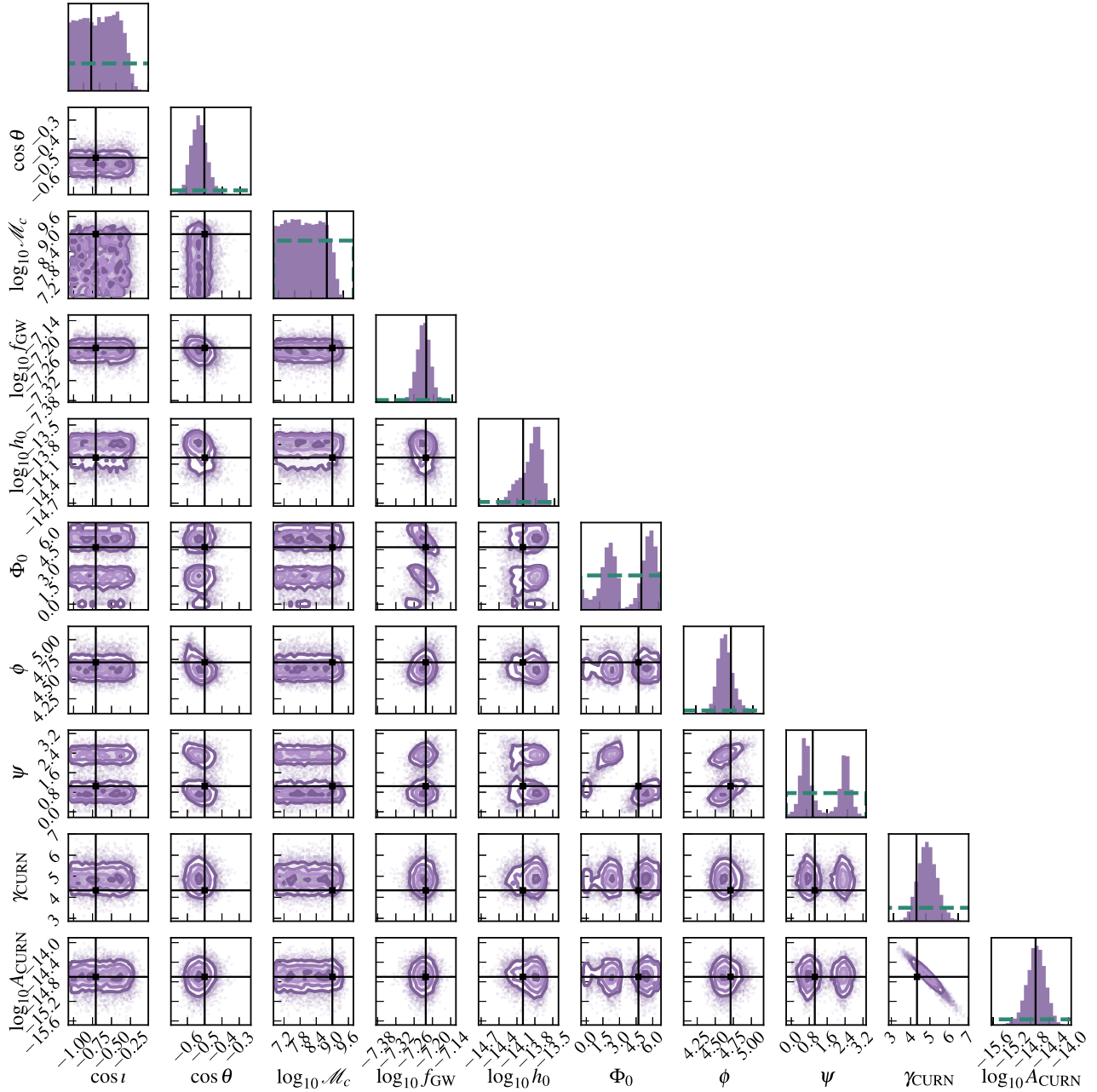
FIG. 3.   1D and 2D posterior distributions for the eight parameters describing a SMBHB signal emitting GWs at $f_{GW} = 60$ nHz at an SNR of 9.3. The true values of the injected parameters are shown as solid black lines, and the priors are plotted on the 1D histograms as horizontal, green dashed lines. Similar to the low-frequency injection analysis, all true values fall within their posteriors, with parameters such as the sky location, GW frequency, and GW strain parameters being tightly constrained. The binary chirp mass posterior now features an upper limit excluding sources that would have undergone significant frequency evolution over the data time span.

independent analyses, in the resolution of the sky map from the NANOGrav 12.5-year CW paper. The CW frequency is held fixed at $f_{GW} = 7.65 \times 10^9$ Hz, the most sensitive frequency in the NANOGrav 12.5-year dataset.

Our strategy is to run one single chain with HMC sampling and leverage the pipeline's efficiency to fully explore across the broader sky range, allowing us to compute a series of GW strain upper limits as a function of sky

location and populate the sky map in post-processing. We run one single analysis for $M = 80,000$ samples, after which we break up our chains into sky location bins consistent with the full 768-pixel map. With all autocorrelation lengths of order $\mathcal{O}(1)$, after thinning this results in between $700 - 1,200$ independent samples per reduced sky pixel.

In Fig. 5, we plot the results of our reconstructed sky map. The strain upper limit at the most sensitive sky location is
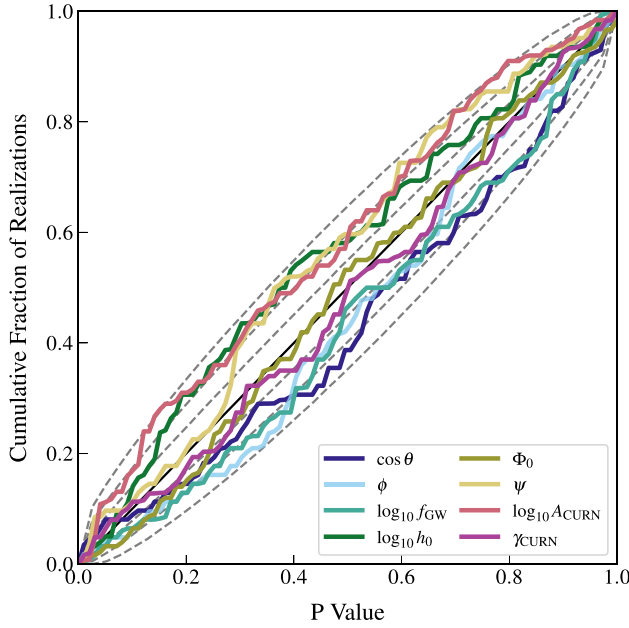
FIG. 4. $p - p$ plot displaying recovery of injected parameters across 100 simulated PTA datasets. All datasets contain a CURN process and a 6 nHz CW injection. Plotted are six lines corresponding to the CW sky location parameters, log strain, log frequency, and CURN amplitude and spectral index. The solid black line along the diagonal represents the line of perfect recovery. Dotted gray lines represent 1σ, 2σ, and 3σ confidence intervals. All plotted parameters lie within these boundaries indicating no significant bias in parameter recovery.
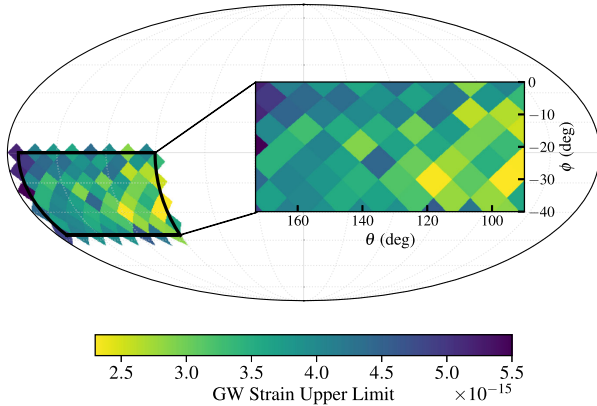

FIG. 5. Map displaying CW strain 95% upper limits for a range of sky location parameters bounded by $\theta \in [\pi/2, 3\pi/4]$, $\phi \in [3\pi/2, 2\pi]$. The data are taken from a single chain run with an HMC pipeline and pixelated to match the resolution of the analogous map for the 12.5-year data set. The analysis is run for $f_{GW} = 7.65 \times 10^9$ Hz, the most sensitive frequency searched. Pixel to pixel uncertainties range between $1.03 \times 10^{-16} < \sigma_{h_0} < 1.81 \times 10^{-15}$.

$h_0 < (2.15 \pm 0.30) \times 10^{-15}$. Its coordinates exactly match that of the most sensitive region from the full NANOGrav 12.5-year analysis, which reported a strain upper limit for

that pixel of $h_0 < (2.66 \pm 0.15) \times 10^{-15}$ [33]. We find the upper limit at the least sensitive sky location to be $h_0 < (5.45 \pm 0.36) \times 10^{-15}$. Unlike the corresponding analysis and map in the NANOGrav 12.5-year CW paper, we marginalize over the amplitude and spectral index of the CURN process instead of fixing the signal parameters to their maximum likelihood values. We also only search over the Earth term of our CW signal. Therefore we do not expect to find perfect agreement between the two results when comparing on a pixel-to-pixel basis.

By effectively sampling over larger portions of the sky, we can cut the computational cost of generating a full sky map by nearly an order of magnitude. Increasing the pixel range of our searches is also a step closer to eliminating a grid-based structure in the otherwise fully Bayesian analysis. The limiting factor in expanding the prior range is purely the computational wall time rather than specific choices on location binning as the HMC sampler can fully explore the posterior even at the least-preferred sky locations. With enough time this can develop into a single all-sky search for producing upper limit maps, and more easily enable making the maps at many different GW frequencies of interest.

V. DISCUSSION

In this paper we have presented an end-to-end pipeline for performing efficient Bayesian searches of the high dimensional and complicated parameter spaces for joint CW and common red-noise process signal analyses with PTA data. Our code employs HMC sampling to conduct accurate parameter estimation. We demonstrated the performance of this sampling routine through numerous tests across both simulated and real PTA data. We showed that by using HMC sampling we can effectively do parameter estimation for both high- and low-frequency CW signals. The methods are robust toward conducting these analyses while simultaneously marginalizing over a common-process signal and can accurately recover both GW signals.

By utilizing the HMC algorithm as our default underlying sampler, we are able to both significantly lower the autocorrelations in our MCMC chains as well as reduce the total number of samples we require per run. Our ability to evenly sample wider areas of the sky means that we are closer to removing a binning element of our otherwise completely Bayesian analysis. The sampler also scales favorably with dimensionality, a positive sign as future PTA datasets inch closer to containing $\mathcal{O}(100)$ pulsars and 100s of corresponding noise parameters.

A significant long-term advantage of this pipeline is its modularity and ability to adapt to a wide range of signal modeling choices. The code is not designed solely for the task of CW searches and can develop and grow into a general purpose analysis suite similar to the current analysis suite ENTERPRISE. For example, it can modified

to run on models considering only a GWB signal, for which previous efforts have already shown HMC sampling to be increasingly useful [30]. Further development can also add the possibility of more sophisticated pulsar noise models or additional deterministic sources of interest in the PTA band. The future of PTA GW analyses is in part defined by its potential computational pitfalls: an ever-increasing data span, noise modeling of growing complexity, and the goal of combined international datasets. These methods will prove a valuable tool alongside the range of computational developments in the PTA community toward addressing these issues before they arise, and keeping our analyses tractable to the future.

[1] G. Agazie, A. Anumarlapudi, A. M. Archibald, Z. Arzoumanian, P. T. Baker, B. Bécsy, L. Blecha, A. Brazier, P. R. Brook, S. Burke-Spolaor *et al.*, Astrophys. J. Lett. **951**, L8 (2023).

[2] J. Antoniadis, P. Arumugam, S. Arumugam, S. Babak, M. Bagchi, A. S. Bak Nielsen, C. G. Bassa, A. Bathula *et al.* (EPTA and InPTA Collaborations), Astron. Astrophys. **678**, A50 (2023).

[3] D. J. Reardon, A. Zic, R. M. Shannon, G. B. Hobbs, M. Bailes, V. Di Marco, A. Kapur, A. F. Rogers, E. Thrane, J. Askew *et al.*, Astrophys. J. Lett. **951**, L6 (2023).

[4] H. Xu, S. Chen, Y. Guo, J. Jiang, B. Wang, J. Xu, Z. Xue, R. Nicolas Caballero, J. Yuan, Y. Xu *et al.*, Res. Astron. Astrophys. **23**, 075024 (2023).

[5] S. Ransom, A. Brazier, S. Chatterjee, T. Cohen, J. M. Cordes, M. E. DeCesar, P. B. Demorest, J. S. Hazboun, M. T. Lam, R. S. Lynch *et al.*, Bull. Am. Astron. Soc. **51**, 195 (2019).

[6] M. Kramer and D. J. Champion, Classical Quantum Gravity **30**, 224009 (2013).

[7] P. Tarafdar, K. Nobleson, P. Rana, J. Singha, M. A. Krishnakumar, B. C. Joshi, A. K. Paladi, N. Kolhe, N. D. Batra, N. Agarwal *et al.*, Publ. Astron. Soc. Aust. **39**, e053 (2022).

[8] R. N. Manchester, G. Hobbs, M. Bailes, W. A. Coles, W. van Straten, M. J. Keith, R. M. Shannon, N. D. R. Bhat, A. Brown, S. G. Burke-Spolaor *et al.*, Publ. Astron. Soc. Aust. **30**, e017 (2013).

[9] M. V. Sazhin, Sov. Astron. **22**, 36 (1978).

[10] S. Detweiler, Astrophys. J. **234**, 1100 (1979).

[11] R. S. Foster and D. C. Backer, Astrophys. J. **361**, 300 (1990).

[12] R. W. Hellings and G. S. Downs, Astrophys. J. Lett. **265**, L39 (1983).

[13] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, Astrophys. J. **623**, 23 (2005).

[14] J. Kormendy and L. C. Ho, Annu. Rev. Astron. Astrophys. **51**, 511 (2013).

[15] P. A. Rosado, A. Sesana, and J. Gair, Mon. Not. R. Astron. Soc. **451**, 2417 (2015).

[16] C. M. F. Mingarelli, T. J. W. Lazio, A. Sesana, J. E. Greene, J. A. Ellis, C.-P. Ma, S. Croft, S. Burke-Spolaor, and S. R. Taylor, Nat. Astron. **1**, 886 (2017).

[17] L. Z. Kelley, L. Blecha, L. Hernquist, A. Sesana, and S. R. Taylor, Mon. Not. R. Astron. Soc. **477**, 964 (2018).

[18] B. Bécsy, N. J. Cornish, and L. Z. Kelley, Astrophys. J. **941**, 119 (2022).

[19] G. D. Quinlan, New Astron. **1**, 35 (1996).

[20] Z. Haiman, B. Kocsis, and K. Menou, Astrophys. J. **700**, 1952 (2009).

[21] M. Charisi, S. R Taylor, J. Runnoe, T. Bogdanovic, and J. R. Trump, Mon. Not. R. Astron. Soc. **510**, 5929 (2022).

[22] V. Corbin and N. J. Cornish, arXiv:1008.1782.

[23] K. J. Lee, N. Wex, M. Kramer, B. W. Stappers, C. G. Bassa, G. H. Janssen, R. Karuppusamy, and R. Smits, Mon. Not. R. Astron. Soc. **414**, 3251 (2011).

[24] J. A. Ellis, Classical Quantum Gravity **30**, 224004 (2013).

[25] S. Taylor, J. Ellis, and J. Gair, Phys. Rev. D **90**, 104028 (2014).

[26] B. Bécsy, N. J. Cornish, and M. C. Digman, Phys. Rev. D **105**, 122003 (2022).

[27] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[28] R. Neal, in *Handbook of Markov Chain Monte Carlo* (CRC, Boca Raton, 2011), pp. 113–162.

[29] L. Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren, Phys. Rev. D **87**, 104021 (2013).

[30] G. E. Freedman, A. D. Johnson, R. van Haasteren, and S. J. Vigeland, Phys. Rev. D **107**, 043013 (2023).

[31] E. S. Phinney, arXiv:astro-ph/0108028.

[32] R. van Haasteren and M. Vallisneri, Phys. Rev. D **90**, 104012 (2014).

[33] Z. Arzoumanian, P. T. Baker, L. Blecha, H. Blumer, A. Brazier, P. R. Brook, S. Burke-Spolaor, B. Bécsy, J. A. Casey-Clyde, M. Charisi *et al.*, Astrophys. J. Lett. **951**, L28 (2023).

[34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[35] M. D. Hoffman and A. Gelman, arXiv:1111.4246.

[36] S. Babak and A. Sesana, Phys. Rev. D **85,** 044034 (2012).

[37] B. Bécsy and N. J. Cornish, Classical Quantum Gravity **37,** 135011 (2020).

[38] G. Agazie, Z. Arzoumanian, P. T. Baker, B. Bécsy, L. Blecha, H. Blumer, A. Brazier, P. R. Brook, S. Burke-Spolaor, R. Burnette *et al.*, Astrophys. J. **963,** 61 (2024).

[39] S. R. Taylor, E. A. Huerta, J. R. Gair, and S. T. McWilliams, Astrophys. J. **817,** 70 (2016).

[40] A. Susobhanan, A. Gopakumar, G. Hobbs, and S. R. Taylor, Phys. Rev. D **101,** 043022 (2020).

[41] A. Susobhanan, Classical Quantum Gravity **40,** 155014 (2023).

[42] G. Agazie, Z. Arzoumanian, P. T. Baker, B. Bécsy, L. Blecha, H. Blumer, A. Brazier, P. R. Brook, S. Burke-Spolaor, J. A. Casey-Clyde *et al.*, Astrophys. J. **963,** 144 (2024).

[43] D. J. Reardon, A. Zic, R. M. Shannon, V. Di Marco, G. B. Hobbs, A. Kapur, M. E. Lower, R. Mandow, H. Middleton, M. T. Miles *et al.*, Astrophys. J. Lett. **951,** L7 (2023).

[44] B. Larsen, C. M. F. Mingarelli, J. S. Hazboun, A. Chalumeau, D. C. Good, J. Simon, G. Agazie, A. Anumarlapudi, A. M. Archibald, Z. Arzoumanian *et al.*, Astrophys. J. **972,** 49 (2024).

[45] R. van Haasteren, Astrophys. J. Suppl. Ser. **273,** 23 (2024).

[46] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne *et al.*, JAX: Composable transformations of Python+NumPy programs (2018), http://github.com/google/jax.

[47] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, Zenodo, 10.5281/zenodo.4059815 (2020).

[48] A. Cabezas, A. Corenflos, J. Lao, and R. Louf, arXiv:2402.10797.

[49] M. Vallisneri, Astrophysics Source Code Library, ascl:2002.017, https://ui.adsabs.harvard.edu/abs/2020ascl.soft02017V (2020).

[50] M. F. Alam, Z. Arzoumanian, P. T. Baker, H. Blumer, K. E. Bohler, A. Brazier, P. R. Brook, S. Burke-Spolaor, K. Caballero, R. S. Camuccio *et al.*, Astrophys. J. Suppl. Ser. **252,** 4 (2021).

[51] Z. Arzoumanian, P. T. Baker, H. Blumer, B. Bécsy, A. Brazier, P. R. Brook, S. Burke-Spolaor, S. Chatterjee, S. Chen, J. M. Cordes *et al.*, Astrophys. J. Lett. **905,** L34 (2020).

[52] F. A. Jenet, A. Lommen, S. L. Larson, and L. Wen, Astrophys. J. **606,** 799 (2004).

[53] S. Iguchi, T. Okuda, and H. Sudou, Astrophys. J. Lett. **724,** L166 (2010).