

Fast Bayesian analysis of individual binaries in pulsar timing array data

Bence Bécsey¹,* Neil J. Cornish¹, and Matthew C. Digman¹

*eXtreme Gravity Institute, Department of Physics, Montana State University,
Bozeman, Montana 59717, USA*



(Received 14 April 2022; accepted 25 May 2022; published 17 June 2022)

Searching for gravitational waves in pulsar timing array data is computationally intensive. The data is unevenly sampled, and the noise is heteroscedastic, necessitating the use of a time-domain likelihood function with attendant expensive matrix operations. The computational cost is exacerbated when searching for individual supermassive black hole binaries, which have a large parameter space due to the additional pulsar distance, phase offset and noise model parameters needed for each pulsar. We introduce a new formulation of the likelihood function which can be used to make the Bayesian analysis significantly faster. We divide the parameters into *projection* and *shape* parameters. We then accelerate the exploration of the projection parameters by more than four orders of magnitude by precomputing the expensive inner products for each set of shape parameters. The projection parameters include nuisance parameters such as the gravitational wave phase offset at each pulsar. In the new scheme, these troublesome nuisance parameters are efficiently marginalized over using multiple-try Markov chain Monte Carlo sampling as part of a Metropolis-within-Gibbs scheme. The acceleration provided by our method will become increasingly important as pulsar timing datasets rapidly grow. Our method also makes sophisticated analyses more tractable, such as searches for multiple binaries, or binaries with non-negligible eccentricities.

DOI: [10.1103/PhysRevD.105.122003](https://doi.org/10.1103/PhysRevD.105.122003)

I. INTRODUCTION

Gravitational waves (GWs) with nHz frequencies can be probed by monitoring the time-of-arrival (TOA) of radio pulses emitted by millisecond pulsars. The primary targets of these pulsar timing arrays (PTAs) are GWs from supermassive black hole binaries (SMBHBs). These can potentially be recovered from the data individually, or one can detect a stochastic GW background (GWB) emerging from the ensemble of all SMBHBs in the observable Universe (for a review, see e.g., [1]).

All three major PTA experiments recently detected a common red noise process: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [2]; the European Pulsar Timing Array (EPTA) [3]; and the Parkes Pulsar Timing Array (PPTA) [4]. The International Pulsar Timing Array (IPTA) also found strong support for such a signal in the combined dataset from these regional PTAs [5]. A common red noise process is expected to be the first sign of the GWB [6,7]. There remain alternative explanations, which can be ruled out if we observe the Hellings-Downs correlations characteristic of a GWB. Assuming the signal is due to the GWB, we expect to see a clear sign of these correlations within the next few years [7].

The detection of GWs from the brightest individual SMBHB is expected to happen not long after the confident detection of the GWB (see e.g., [8,9]). Searching for GWs from individual SMBHBs is especially important since they will provide direct evidence of SMBHBs emitting GWs, unlike the GWB, which in principle could come from other GW sources (e.g. cosmological phase transitions [10]). An individually resolved SMBHB would also provide a unique opportunity for joint electromagnetic and GW observations (see e.g., [11]). Several searches for individual SMBHBs have been carried out in the past, resulting in upper limits on the amplitude of GWs from SMBHBs [12–14].

Numerous analysis techniques have been developed to search for and characterize GWs from individual SMBHBs (see e.g. [15–18]). Fully Bayesian methods tend to be computationally intensive due to the large parameter space that needs to be explored, and because the unevenly sampled datasets and heteroscedastic (nonstationary) noise rule out the fast Fourier domain methods used in the analysis of ground-based GW detectors (see [19] and references therein). This problem will further intensify as our datasets get larger, and as we try to incorporate more complicated signal models, like eccentric binaries [20,21] or multiple binaries [22]. In this paper we present a method that can significantly speed up such Bayesian analyses by separating *shape parameters*, which determine the morphology of the GW signal, from *projection parameters*,

*bencebecsy@montana.edu

which only affect how the signal is projected onto the line of sight of each pulsar. Our new technique enables exploration of projection parameters (most of which are nuisance parameters) at practically zero cost. Due to the large number of projection parameters, this results in a significant speedup of the overall analysis. This method is implemented in the `QuickCW` software package [23]. Note that this approach is similar to the F-statistic method introduced in Ref. [24], but instead of analytically maximizing over certain parameters, we numerically marginalize over them, thus keeping the analysis fully Bayesian. There are also similarities with the techniques presented in Ref. [18], where the likelihood function is either maximized or marginalized over the pulsar phase parameters.

The paper is organized as follows. In Sec. II we review the traditional formulation of the individual source signal model and introduce the alternative formulation allowing for the rapid exploration of projection parameters. In Sec. III we describe sampling methods that can be used to maximize the advantage brought about by the new likelihood. We validate these methods by analyzing various simulated datasets (Sec. IV) and the NANOGrav 11-year dataset [12,25] (Sec. V). We conclude and outline possible future directions in Sec. VI. Throughout this paper we use units where $G = c = 1$.

II. FAST LIKELIHOOD

In this section we review the effect of GWs from a circular SMBHB on PTA residuals, and we describe an alternative formulation of the signal model allowing for the separation of shape and projection parameters. The latter shows several similarities with the F-statistic (see e.g., [24,26–29]). However, we keep all parameters of the signal free, instead of maximizing over some of them as is done in the F-statistic analysis.

The emitted GW signal can be written as [30]

$$h_{ab}(t, \hat{\Omega}) = e_{ab}^+(\hat{\Omega})h_+(t, \hat{\Omega}) + e_{ab}^\times(\hat{\Omega})h_\times(t, \hat{\Omega}), \quad (1)$$

where $\hat{\Omega}$ is a unit vector from the GW source to the Solar System barycenter (SSB), $h_{+, \times}$ are the polarization amplitudes, and $e_{ab}^{+, \times}$ are the polarization tensors. The polarization tensors can be written in the SSB frame as

$$e_{ab}^+(\hat{\Omega}) = \hat{m}_a \hat{m}_b - \hat{n}_a \hat{n}_b, \quad (2)$$

$$e_{ab}^\times(\hat{\Omega}) = \hat{m}_a \hat{n}_b + \hat{n}_a \hat{m}_b, \quad (3)$$

where $\hat{\Omega}$, $\hat{m}$, and $\hat{n}$ are orthonormal vectors defined as

$$\hat{\Omega} = -\sin \theta \cos \phi \hat{x} - \sin \theta \sin \phi \hat{y} - \cos \theta \hat{z}, \quad (4)$$

$$\hat{m} = \sin \phi \hat{x} - \cos \phi \hat{y}, \quad (5)$$

$$\hat{n} = -\cos \theta \cos \phi \hat{x} - \cos \theta \sin \phi \hat{y} + \sin \theta \hat{z}. \quad (6)$$

The response of a pulsar to the source is described by the antenna pattern functions F^+ and $F^\times$,

$$F^+(\hat{\Omega}) = \frac{1}{2} \frac{(\hat{m} \cdot \hat{p})^2 - (\hat{n} \cdot \hat{p})^2}{1 + \hat{\Omega} \cdot \hat{p}}, \quad (7)$$

$$F^\times(\hat{\Omega}) = \frac{(\hat{m} \cdot \hat{p})(\hat{n} \cdot \hat{p})}{1 + \hat{\Omega} \cdot \hat{p}}, \quad (8)$$

where $\hat{p}$ is a unit vector pointing from the SSB to the pulsar. The effect of a GW on a pulsar's TOAs can be written as

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

where $\Delta s_{+, \times}$ is the difference between the signal induced at the pulsar and at the Earth (the so-called “pulsar term” and “Earth term”),

$$\Delta s_{+, \times}(t) = s_{+, \times}(t_p) - s_{+, \times}(t), \quad (10)$$

where t is the time measured at the SSB and t_p is the corresponding time at pulsar p . From geometry, we can relate t and t_p by

$$t_p = t - L_p(1 + \hat{\Omega} \cdot \hat{p}), \quad (11)$$

where L_p is the distance to the pulsar.

For a circular binary, at zeroth post-Newtonian (0-PN) order, $s_{+, \times}$ is given by

$$s_+(t) = \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} [\sin 2\Phi(t)(1 + \cos^2 \iota) \cos 2\psi + 2 \cos 2\Phi(t) \cos \iota \sin 2\psi], \quad (12)$$

$$s_\times(t) = \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} [-\sin 2\Phi(t)(1 + \cos^2 \iota) \sin 2\psi + 2 \cos 2\Phi(t) \cos \iota \cos 2\psi], \quad (13)$$

where ι is the inclination angle of the SMBHB, ψ is the GW polarization angle, d_L is the luminosity distance to the source, and $\mathcal{M} \equiv (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$ is a combination of the black hole masses m_1 and m_2 called the “chirp mass”.

The evolution of the frequency in the $\Phi(t)$ phase terms can be kept fully general, or Taylor expanded to leading order in the first time derivative of the orbital angular frequency, $\dot{\omega}$. The signal in pulsar p_i can be rewritten as [cf. Eq. (9)]

$$s_i(t, \hat{\Omega}) = \sum_{j=1}^4 \sigma_{4(i-1)+j}(\hat{\Omega}, A_e, \iota, \psi, \Phi_0, \Phi_i) S^{4(i-1)+j}(t), \quad (14)$$

where the filter functions, $S^{4(i-1)+j}(t)$, are defined as

$$\begin{aligned} S^{4(i-1)+1}(t) &= [\omega_0/\omega(t)]^{1/3} \cos 2\Phi(t), \\ S^{4(i-1)+2}(t) &= [\omega_0/\omega(t)]^{1/3} \sin 2\Phi(t), \\ S^{4(i-1)+3}(t) &= [\omega_i/\omega(t_{p_i})]^{1/3} \cos 2\Phi(t_{p_i}), \\ S^{4(i-1)+4}(t) &= [\omega_i/\omega(t_{p_i})]^{1/3} \sin 2\Phi(t_{p_i}), \end{aligned} \quad (15)$$

where the prefactors in front of the trigonometric functions take into account the slight change in amplitude due to the frequency changing within the observational time span. The reference frequency of the Earth term is denoted as $\omega_0 = \omega(t_0)$, while $\omega_i = \omega(t_0 - L_{p_i}(1 + \hat{\Omega} \cdot \hat{p}_i))$ is the reference frequency of the pulsar term in pulsar p_i . The time-dependent angular frequency is given by

$$\begin{aligned} \sigma_{4(i-1)+1} &= -A_e \omega_0^{-1} [\cos 2\Phi_0 (1 + \cos^2 \iota) (\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) - 2 \sin 2\Phi_0 \cos \iota (\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)], \\ \sigma_{4(i-1)+2} &= -A_e \omega_0^{-1} [\sin 2\Phi_0 (1 + \cos^2 \iota) (\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) + 2 \cos 2\Phi_0 \cos \iota (\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)], \\ \sigma_{4(i-1)+3} &= A_i \omega_0^{-1} [\cos 2\Phi_i (1 + \cos^2 \iota) (\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) - 2 \sin 2\Phi_i \cos \iota (\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)], \\ \sigma_{4(i-1)+4} &= A_i \omega_0^{-1} [\sin 2\Phi_i (1 + \cos^2 \iota) (\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) + 2 \cos 2\Phi_i \cos \iota (\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)], \end{aligned} \quad (18)$$

where $A_e = \mathcal{M}^{5/3} d_L^{-1} \omega_0^{2/3}$ and $A_i = \mathcal{M}^{5/3} d_L^{-1} \omega_i^{2/3}$. Note that A_i are not independent parameters, since they are uniquely determined by A_e , ω_0 , and ω_i . The log-likelihood can be written as

$$\log L = -\frac{1}{2}(\delta t - s|\delta t - s) - \frac{1}{2} \log \det(2\pi C), \quad (19)$$

where

$$(a|b) = a^T C^{-1} b. \quad (20)$$

Here $C = N + TBT^T$, where N is the white noise covariance matrix, T is the design matrix for the timing model, red noise, and jitter noise, and B is the prior matrix for the hyperparameters of those (see e.g., [1]). Using the Woodbury matrix identity, we can express the inverse of C as

$$C^{-1} = (N + TBT^T)^{-1} = N^{-1} - N^{-1}T\Sigma^{-1}T^TN^{-1}, \quad (21)$$

where $\Sigma^{-1} = B^{-1} - T^TN^{-1}T$. Using the four filters we have

$$\omega(t) = \omega_0 \left[1 - \frac{256}{5} \mathcal{M}^{5/3} \omega_0^{8/3} (t - t_0) \right]^{-3/8}. \quad (16)$$

The phases in Eq. (15) are given by

$$\begin{aligned} \Phi(t) &= \Phi_0 + \frac{1}{32} \mathcal{M}^{-5/3} [\omega_0^{-5/3} - \omega(t)^{-5/3}], \\ \Phi(t_{p_i}) &= \Phi_i + \frac{1}{32} \mathcal{M}^{-5/3} [\omega_i^{-5/3} - \omega(t_{p_i})^{-5/3}]. \end{aligned} \quad (17)$$

Note that while the initial phases Φ_i and Φ_0 are not independent parameters, we treat them as independent since the current uncertainties in the pulsar distances are so large that it is practically impossible to phase connect the Earth and pulsar terms. The first two filters are for the Earth term, and are the same for all pulsars (but they are sampled at different discrete times due to the different observing schedules for each pulsar). The coefficients σ_k are given by

$$\begin{aligned} \log L &= -\frac{1}{2}(\delta t|\delta t) - \frac{1}{2} \log \det(2\pi C) \\ &+ \sum_{k=1}^{4N_p} \sigma_k N^k - \frac{1}{2} \sum_{k=1}^{4N_p} \sum_{l=1}^{4N_p} \sigma_k \sigma_l M^{kl}, \end{aligned} \quad (22)$$

where $N^k = (\delta t|S^k)$, $M^{kl} = (S^k|S^l)$, and N_p is the number of pulsars in the array. Computing these inner products is the expensive step. The inner products have to be recomputed each time the noise model or the shape parameters of the signal (see Table I) are updated. Note that while the vast majority of the off-diagonal terms in M^{kl} are zero, the per-pulsar quadratures and pulsar-Earth cross terms will not vanish since the data is unevenly sampled and the orbital periods are generally not integer submultiples of the observation time. The band-diagonal structure of the M^{kl} means that it can be computed and stored in an array of size $10N_p$,

TABLE I. List of shape and projection parameters and their numbers as a function of the number of pulsars (N_p).

Shape parameters	Projection parameters
$(4 + 3 \times N_p)$	$(4 + N_p)$
$\theta, \phi, f_{\text{GW}}, \mathcal{M}, L_{p_i}, \gamma_{p_i}, A_{RN,p_i}$	$\iota, A_e, \Phi_0, \Psi, \Phi_i$

rather than the naive $(4N_p)^2$. Note that in the presence of correlated noise between different pulsars (e.g., a stochastic GW background), inner products will include cross terms between pulsars. Appendix A describes how the cross terms can be avoided, thereby maintaining the factorized form of the likelihood and the band-diagonal structure of the M^{kl} matrix.

For a fixed set of noise parameters and shape parameters, we can compute the likelihood for *any* set of projection parameters (see Table I) in essentially zero time. See details of how we can take full advantage of the speedup by an implementation relying on Numba in Appendix B. This increased speed allows us to fully marginalize over these projection parameters by performing a large number of MCMC updates of just these parameters.

If we are only interested in the Earth term, the sky location could also be marginalized over without needing to recompute the inner products. The pulsar terms ruin this separation of variables since the pulsar time t_p from Eq. (11) depends on the sky location.

We can see from Eq. (16) that the initial angular frequency at each pulsar is given by

$$\omega_i = \omega_0 \left(1 + \frac{256}{5} \mathcal{M}^{5/3} \omega_0^{8/3} L_{p_i} (1 + \hat{\Omega} \cdot \hat{p}_i) \right)^{-3/8}. \quad (23)$$

Note that the Earth term and pulsar term can be quite different since

$$\begin{aligned} \frac{256}{5} \mathcal{M}^{5/3} \omega_0^{8/3} L_{p_i} &= 10.16 \left(\frac{\mathcal{M}}{10^9 M_\odot} \right)^{5/3} \\ &\times \left(\frac{\omega_0}{2\pi/\text{yr}} \right)^{8/3} \left(\frac{L_{p_i}}{1 \text{ kpc}} \right). \end{aligned} \quad (24)$$

On the other hand, the Earth term frequency only changes by a small amount during an $\mathcal{O}(10)$ year observing span so long as the systems are not very heavy or the current frequency is not too high since

$$\begin{aligned} \frac{256}{5} \mathcal{M}^{5/3} \omega_0^{8/3} t &= 0.031 \left(\frac{\mathcal{M}}{10^9 M_\odot} \right)^{5/3} \\ &\times \left(\frac{\omega_0}{2\pi/\text{yr}} \right)^{8/3} \left(\frac{t}{10 \text{ yr}} \right). \end{aligned} \quad (25)$$

III. METROPOLIS-WITHIN-GIBBS SAMPLING

With the new formulation of the likelihood in Eq. (22) the computational burden is highly dependent on which parameters we try to update. Since calculating the likelihood at new projection parameter values is significantly faster than calculating it for a new set of shape parameters, we propose more updates in projection parameters.

We achieve that by employing a Metropolis-within-Gibbs sampler (see [31] and references therein), where the sampler completes a block of projection parameter updates [in this case typically $\mathcal{O}(10^3)$] before attempting a shape parameter update. The Metropolis-within-Gibbs sampler allows us to sample projection parameters extremely well with practically no additional cost.

In order to optimize the mixing in shape parameters, we also use a technique called multiple-try MCMC (MTMCMC) [32,33]. The idea of MTMCMC is that one can propose N different points, select one of them based on some importance weights, and accept or reject the new sample based on an acceptance probability which depends on the likelihood at all N proposed points. With $N = 1$, one recovers the Metropolis-Hastings algorithm, while as $N \rightarrow \infty$, we draw independent samples from the posterior. Variants of MTMCMC where the trials can be drawn from different distributions or can be correlated were introduced in [34,35].

We apply MTMCMC for the shape parameter updates in our Metropolis-within-Gibbs sampler as follows. Let us denote the parameters of the MCMC chain at the i th iteration as

$$\boldsymbol{\theta}_i = \{\boldsymbol{\theta}_i^{(s)}, \boldsymbol{\theta}_i^{(p)}\}, \quad (26)$$

where $\boldsymbol{\theta}_i^{(s)}$ are the shape parameters and $\boldsymbol{\theta}_i^{(p)}$ are the projection parameters at the i th iteration. We determine the next sample, $\boldsymbol{\theta}_{i+1}$, using the following algorithm:

- (1) Draw a set of new shape parameters $\boldsymbol{\theta}^{(s)}$ from the proposal $q^{(s)}(\boldsymbol{\theta}^{(s)}|\boldsymbol{\theta}_i^{(s)})$.
- (2) Draw N different sets of projection parameters, $\boldsymbol{\theta}_{(1)}^{(p)}, \dots, \boldsymbol{\theta}_{(k)}^{(p)}, \dots, \boldsymbol{\theta}_{(N)}^{(p)}$, from the proposals $q_k^{(p)}(\boldsymbol{\theta}_{(k)}^{(p)}|\boldsymbol{\theta}_i^{(p)})$.
- (3) Randomly select $\boldsymbol{\theta}_{(j)} = \{\boldsymbol{\theta}^{(s)}, \boldsymbol{\theta}_{(j)}^{(p)}\}$, according to the probability mass function,

$$p(\boldsymbol{\theta}_{(j)}) = \frac{L(\boldsymbol{\theta}_{(j)})}{\sum_{k=1}^N L(\boldsymbol{\theta}_{(k)})}. \quad (27)$$

- (4) Form N auxiliary samples,

$$\boldsymbol{\Theta}_{(k)} = \begin{cases} \{\boldsymbol{\theta}_i^{(s)}, \boldsymbol{\theta}_{(k)}^{(p)}\} & \text{if } k \neq j \\ \boldsymbol{\theta}_i & \text{if } k = j. \end{cases} \quad (28)$$

- (5) Set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_{(j)}$ with probability,

$$\min \left(1, \frac{\sum_{k=1}^N L(\boldsymbol{\theta}_{(k)})}{\sum_{k=1}^N L(\boldsymbol{\Theta}_{(k)})} \right), \quad (29)$$

otherwise, set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$.

Note that Eqs. (27) and (29) only depend on the likelihood values and not on the proposal densities, because we only employ symmetric proposal distributions. We can see that if $N = 1$, Eq. (29) reduces to the usual Metropolis-Hastings acceptance probability. Also note that since we are selecting the proposed projection parameters, $\theta_{(j)}^{(p)}$, according to Eq. (27), as $N \rightarrow \infty$ we are drawing independent samples from the conditional likelihood at the new shape parameters, $L(\theta_{(j)}^{(p)}|\theta^{(s)})$. At this large N limit, we can also interpret the acceptance probability in Eq. (29) as comparing the likelihood at the new and old shape parameters marginalized over the projection parameters.

To ensure the new set of shape parameters is not rejected due to the lack of an appropriate set of projection parameters, we use a fairly large number of trials (typically $N = 10,000$). A large number of projection parameter trials increases acceptance of shape parameter proposals with little additional cost due to the comparatively cheap evaluation of the likelihood at different projection parameters. For the shape parameter proposals, we use a mix of Fisher proposals, differential evolution proposals, and prior draws. At a given step we only update a specific set of parameters: (i) the four common parameters (sky location, chirp mass, frequency); (ii) pulsar distances; (iii) red noise parameters. For the projection parameter proposals, we always have one trial keep its original projection parameters. This gives a good chance of accepting the new shape parameters even if for some reason the large number of trials with perturbed projection parameters would land on low-likelihood places. For the rest, we draw each projection parameter independently using an optimal jump scale determined by the second derivative of the likelihood in that direction. If the optimal scale is larger than a threshold, we do a draw from the prior instead. Switching to prior draws ensures that when a parameter's value is not well determined we do a prior draw resulting in good exploration. Details about the implementation of the new likelihood function and the sampler can be found in Appendix B.

In general, MTMCMC algorithms provide better mixing at the cost of additional likelihood evaluations needed in Eqs. (27) and (29). MTMCMC methods are particularly well suited to our new method, because the extremely cheap evaluations of the likelihood for different

projection parameters give us the benefits at little additional cost.

IV. RESULTS WITH SIMULATED DATA

To test and illustrate the performance of the new methods described above, we analyzed several simulated datasets. To gauge expected runtimes realistically, all our datasets are made to resemble the latest publicly available dataset of the NANOGrav collaboration, the NANOGrav 12.5-year dataset [36]. They all contain the same 45 pulsars with the same timing solution and same observation properties (epochs, observing frequencies, TOA errors) as the real dataset. We simulated white and red noise in all pulsars according to the best-fit parameters found for the real observations. The typical runtime on a dataset of this size and complexity was a few days on an AMD Ryzen Threadripper 3970X 32-core processor.

In addition to white and red noise, we added three different signals to our dataset to test our analysis in different interesting scenarios: a slowly evolving signal which has comparable frequencies in the Earth term and the pulsar terms (see Dataset 1 in Table II and Sec. IV A 1); a rapidly evolving signal where the Earth and pulsar terms have significantly different frequencies (see Dataset 2 in Table II and Sec. IV A 2); and a low-SNR marginally detectable signal at the most sensitive sky location and frequency, which is meant to represent the kind of signal we are most likely to detect first (see Dataset 3 in Table II and Sec. IV A 3). We also analyzed a dataset without any signals to test how the new pipeline can produce upper limits if no significant GW sources are found (see Sec. IV B).

A. Detection analyses

1. Slowly evolving signal

We first analyzed a signal with a relatively low chirp mass and frequency (see Dataset 1 in Table II) resulting in only slightly different Earth term and pulsar term frequencies [cf. Eq. (24)]. We chose the amplitude of the signal to achieve a moderate SNR of 10.4. Fig. 1 shows how the square of the SNR (a good proxy for detectability) is distributed among the 45 pulsars in the array. The highly

TABLE II. List of parameter values for simulated datasets. Red noise parameters and pulsar distances were set to the official values from the NANOGrav 12.5-year dataset [36]. The GW phase at each pulsar was determined from Φ_0 , the light travel time to each pulsar, and f_{GW} .

	θ	ι	ϕ	f_{GW} (nHz)	A_e	$\mathcal{M}$ ($M_\odot$)	Φ_0	Ψ	d_L (Mpc) ^a
Dataset 1	$\pi/3$	1.0	4.5	8	5×10^{-15}	5×10^8	1.0	1.0	7.5
Dataset 2	$\pi/3$	1.0	4.5	20	1×10^{-14}	5×10^9	1.0	1.0	320
Dataset 3	$2\pi/3$	0.5	4.5	8	2×10^{-15}	1×10^9	2.0	1.5	60

^aNote that d_L is not an independent parameter we fit for. It is completely determined by A_e , f_{GW} , and $\mathcal{M}$.

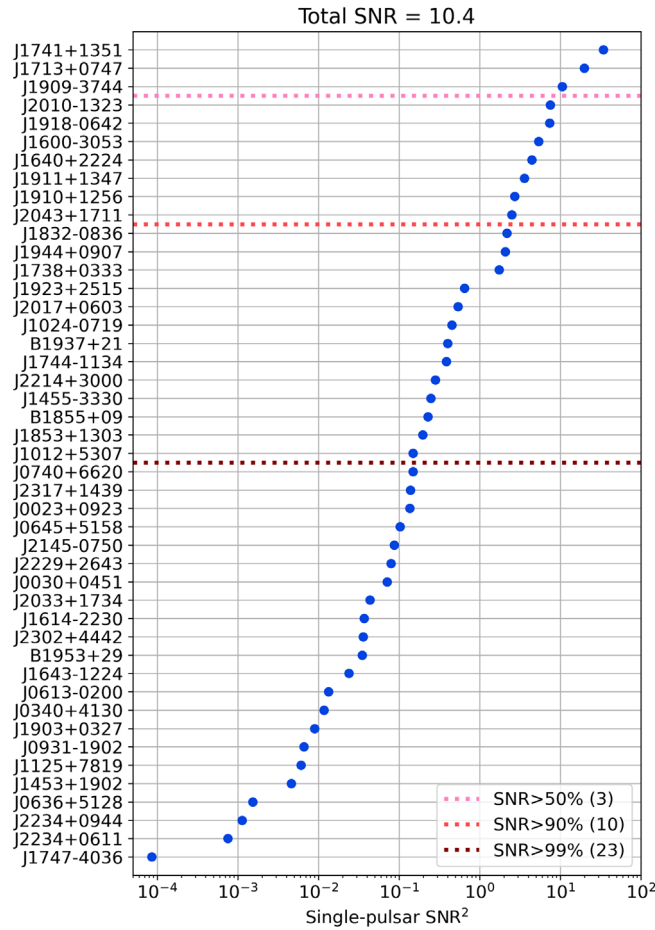


FIG. 1. SNR^2 values in datastreams of each simulated pulsar for slowly a evolving signal (Dataset 1 in Table II). Horizontal dashed lines indicate the minimum number of pulsars needed to reach 50%/90%/99% of the total SNR^2 .

heterogeneous distribution means that only 3/10/23 pulsars are responsible for 50%/90%/99% of the total SNR^2 . There are two reasons for this: (i) the pulsars in the array have a wide range of observation timespans and timing precisions, since they correspond to the real pulsars in the NANOGrav 12.5-year dataset; (ii) pulsars in favorable sky locations can incur significantly higher SNRs compared to pulsars in “bad” sky locations. This imbalance in SNR might suggest that one can arrive at similar results by only using a small subset of the pulsars. However, even pulsars with negligible SNRs can contribute to parameter estimation, since they can rule out some parts of the parameter space where they would be able to reach a higher SNR. Also, one cannot know *a priori* which pulsars could be neglected, since even a less-precisely timed pulsar can have a relatively high SNR if it happens to be in a favorable sky location for a particular source.

Fig. 2 shows the one-dimensional and two-dimensional marginal distributions of the eight signal parameters common to all pulsars. Blue lines and dots indicate the

true values of parameters, while green horizontal lines show the prior distribution for each parameter. The true values of parameters lie within the bulk of the posterior for all parameters. As expected given the slow frequency evolution of the signal, the chirp mass distribution is largely unconstrained, with only the largest values being ruled out as they would have resulted in a detectable frequency evolution. The amplitude is highly correlated with a number of nuisance parameters, most notably the inclination angle (i), emphasizing the importance of effectively sampling these. Some parameters show a highly complex, multimodal posterior, which makes the sampling of those parameters particularly challenging. Some of the intricate structure (especially in sky location) is due to the uneven distribution of pulsars on the sky, which results in highly variable sensitivity over the sky.

It is also interesting to examine the posterior distribution of some of the pulsar-specific parameters, and their correlations with some of the common parameters. Figure 3 shows the pulsar distance and pulsar phase distributions for four selected pulsars, and their correlations with the four common shape parameters. We selected PSR J1713 + 0747 and PSR J1909 – 3744, as these are two of the most precise pulsars in the array and they exhibit high SNRs for this signal (4.45 and 3.25, respectively). We also show posteriors for PSR J1918 – 0642, which is a good example of a pulsar with a moderate SNR of 2.72, and for PSR J2234 + 0944 which has a negligible SNR of 0.03. We can see that accordingly, the posterior of the GW phase at PSR J1713 + 0747 and PSR J1909 – 3744 are highly peaked, while they are less informative for PSR J1918 – 0642 and PSR J2234 + 0944. Note that some of the pulsar term phases are correlated with the sky location parameters (θ and ϕ). We can also see that at high values of $\mathcal{M}$, that also shows a correlation with the pulsar phases. These are due to the fact that we parametrize the initial phase at each pulsar as the sum of the Earth term phase, the phase collected during propagation from Earth to the pulsar, and the corresponding Φ_i parameter. Since the projected distance to the pulsar changes with sky location, the pulsar phases must be corrected as we change the sky location. Similarly, changing the chirp mass changes the phase accumulated between the Earth and the pulsar, so Φ_i needs to be adjusted.

2. Fast evolving signal

The next signal we analyzed has higher frequency and chirp mass than the previous one (see Dataset 2 in Table II), and thus shows a significant frequency evolution between the Earth and pulsars, and it even starts to show a non-negligible evolution within the 12.5-year observing timespan [see Eq. (25)]. The total SNR is 13, with a highly heterogeneous distribution within pulsars. 50%/90%/99% of the total SNR^2 comes from just 2/13/28 pulsars. We show the distribution of the eight parameters common to all

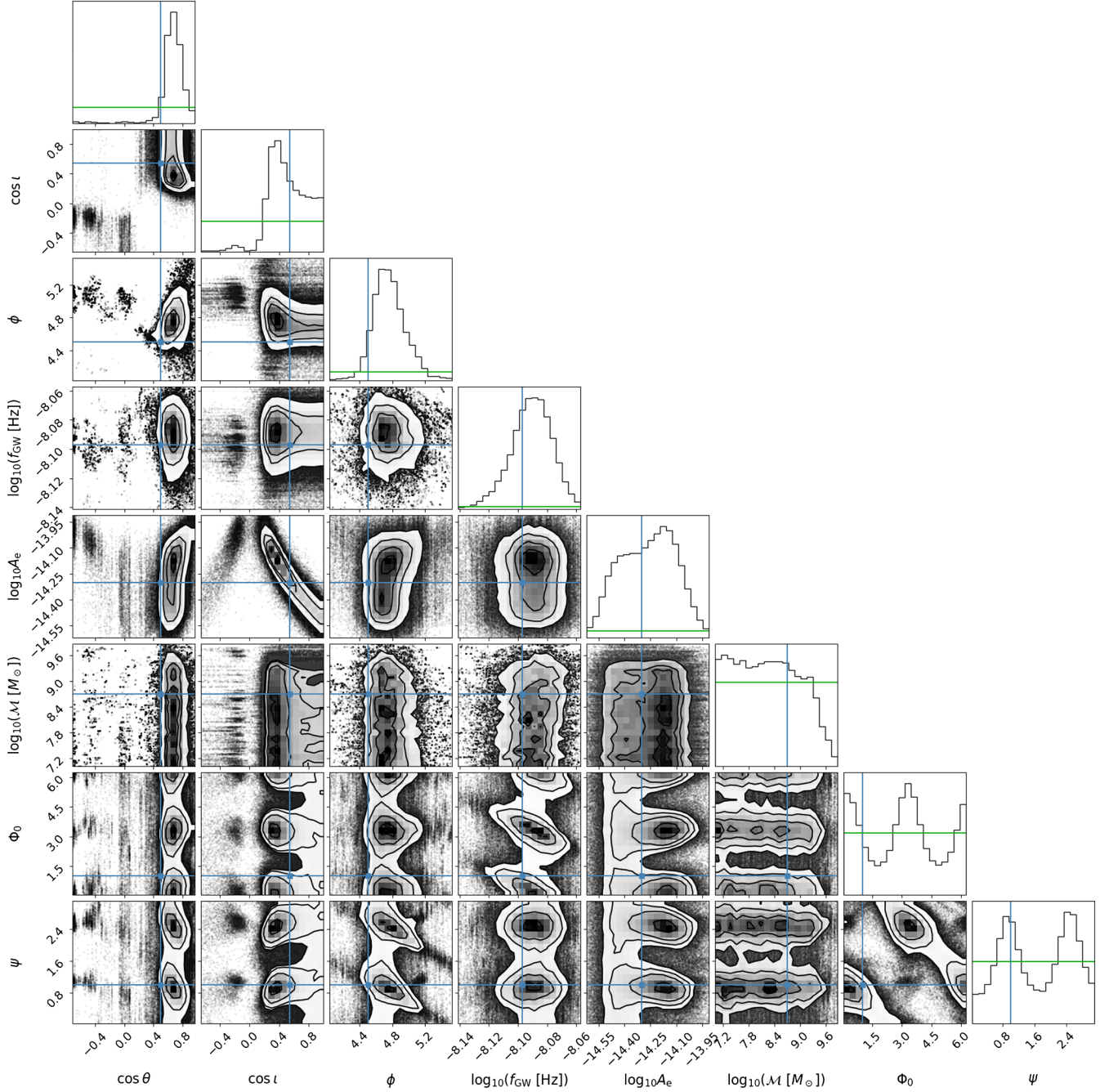


FIG. 2. Corner plot of parameters common to all pulsars for the slowly evolving signal (SNR = 10.4, Dataset 1 in Table II). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the $1\text{-}\sigma/2\text{-}\sigma/3\text{-}\sigma$ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

pulsars in Fig. 4. Unlike in the slowly-evolving case, the observed frequency evolution results in an informative chirp mass posterior. This moderate-SNR source results in a chirp mass measurement with a $1\text{-}\sigma$ error of $\sim 10\%$. Given the high correlations between $\mathcal{M}$ and $\cos\theta$, the chirp mass measurement precision can be further improved if the sky location of the source can be fixed. This can be done e.g. if the host galaxy can be identified through electromagnetic observations (see e.g., [37–39]). The intricate multimodal

structure we have seen for the slowly evolving signal in Fig. 2 is much less prominent for this quickly evolving signal. Many of the one-dimensional marginal distributions we see in Fig. 4 are close to being Gaussian, and only Φ_0 and Ψ show a multimodal structure. This reduction in the complexity of the posterior distributions is due to the fact that the frequency evolution breaks some of the degeneracies present in the signal model when the evolution is negligible.

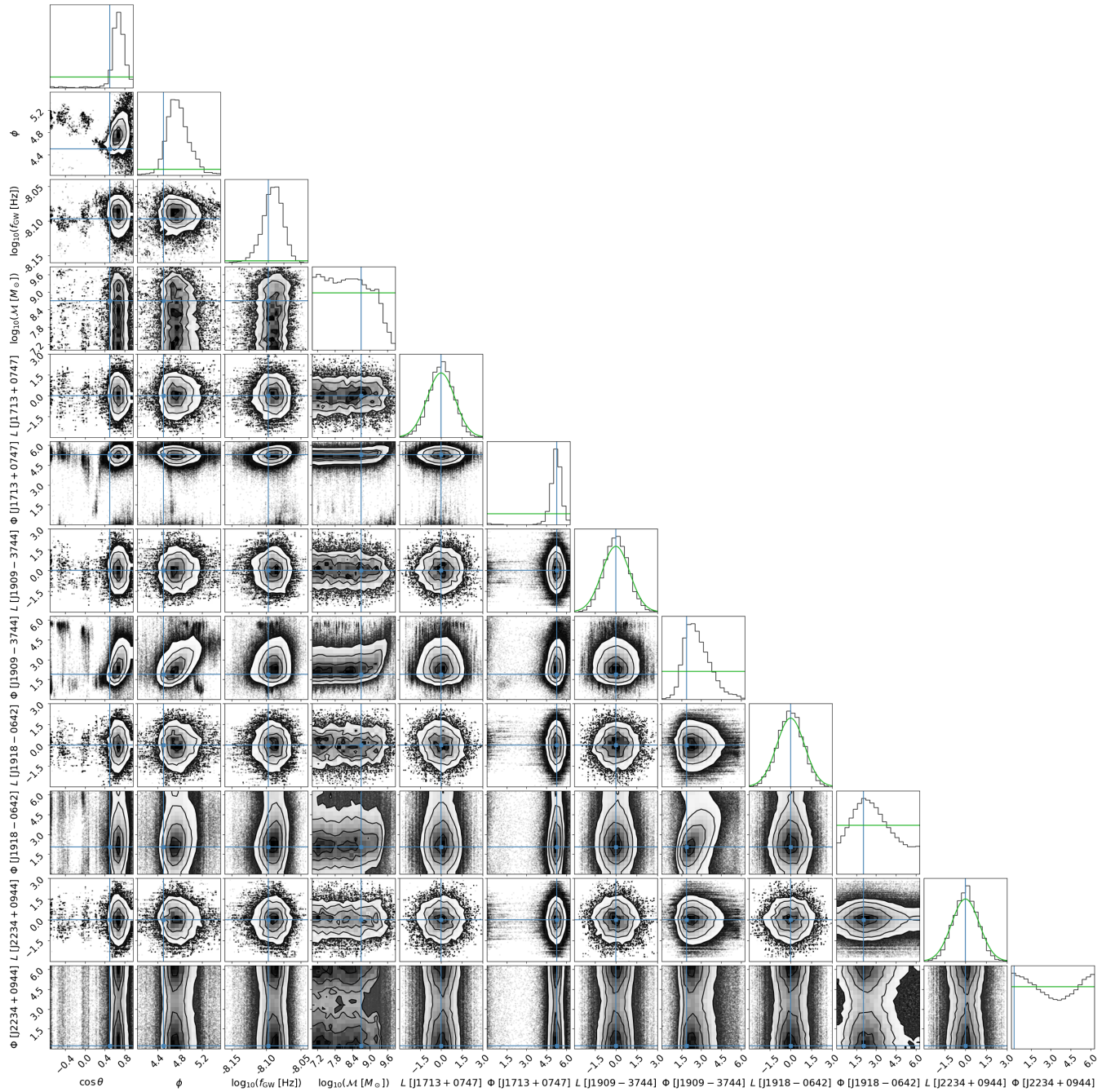


FIG. 3. Corner plot of four common shape parameters and selected pulsar distances and phases for the slowly evolving signal (SNR = 10.4, 1 in Table II). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the 1- σ /2- σ /3- σ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

We show the posterior distributions of phases and distances for four selected pulsars, and their correlations with the four common shape parameters in Fig. 5. We show these for the same four pulsars as in Fig. 3. For this signal, they have the following SNRs: 8.3 for PSR J1713 + 0747, 5.42 for PSR J1909 – 3744, 2.17 for PSR J1918 – 0642, and 0.41 for PSR J2234 + 0944. In this example, the GW phase at PSR J2234 + 0944 perfectly recovers its prior. Similarly to Fig. 3, we also see correlations between the

pulsar phases and sky location or chirp mass. In addition, we see that pulsar phases can also be correlated with the GW frequency and pulsar distances. These have similar explanations as discussed above.

3. Marginally detectable signal

We also analyzed a marginally detectable signal with a total SNR of 4.3, with 50%/90%/99% of the total SNR²

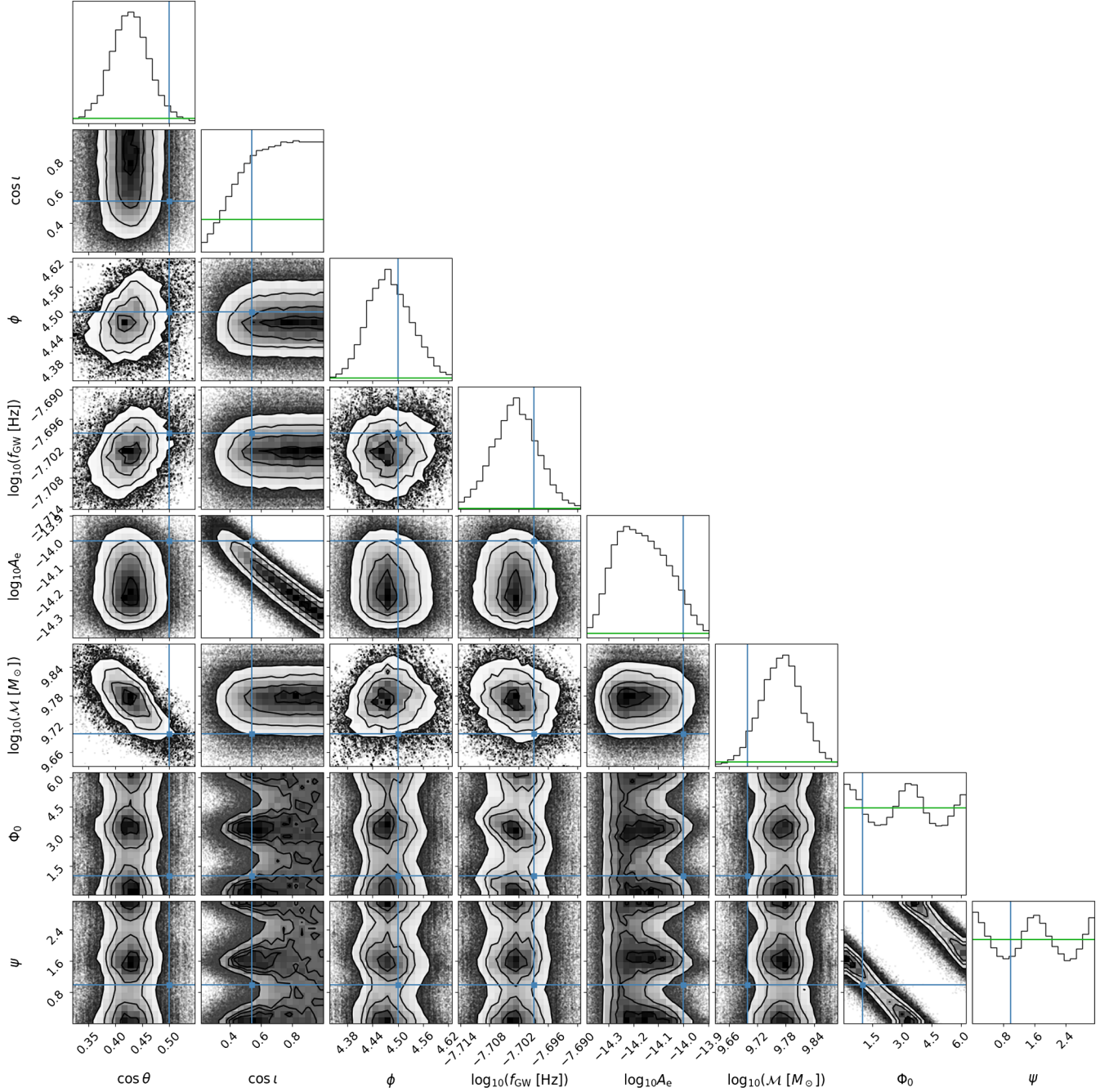


FIG. 4. Corner plot of parameters common to all pulsars for the fast evolving signal (SNR = 13, Dataset 2 in Table II). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the 1- σ /2- σ /3- σ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

coming from 3/12/22 pulsars (see Dataset 3 in Table II). The sky location and frequency ($f_{\text{GW}} = 8$ nHz) of the signal were chosen so that it roughly corresponds to the most sensitive part of the parameter space (see Sec. IV B) and thus represents a typical signal we might expect to detect first. We have also chosen a moderate chirp mass ($\mathcal{M} = 1 \times 10^9 M_{\odot}$), which is heuristically what we are most likely to detect, since systems with higher chirp mass

have higher GW amplitudes, but are also more rare. This combination of f_{GW} and $\mathcal{M}$ results in a signal with a relatively low amount of frequency evolution.

Figure 6 shows the one-dimensional and two-dimensional marginal posterior distributions of the common parameters for this signal. For most of the parameters, the posterior distributions are not significantly different from their respective priors. The main exceptions are f_{GW}

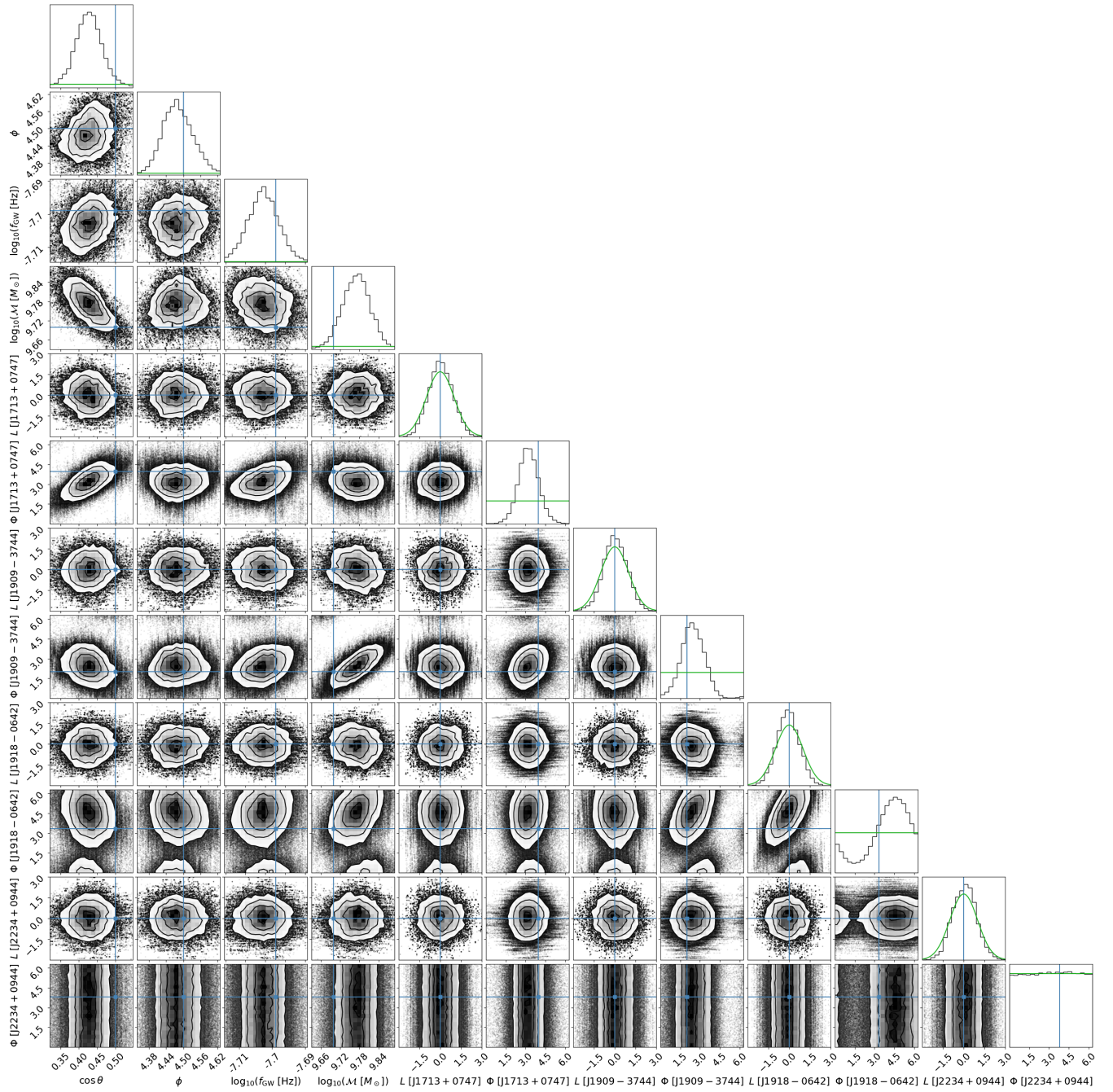


FIG. 5. Corner plot of four common shape parameters and selected pulsar distances and phases for the fast evolving signal (SNR = 13, Dataset 2 in Table II). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the $1\text{-}\sigma/2\text{-}\sigma/3\text{-}\sigma$ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

and A_e , both of which have a posterior peaked at the true location of the parameter. The posterior for f_{GW} also has non-negligible support over the entire prior range. A_e values significantly higher than the true parameter are ruled out, however, the posterior has support extending to the lower prior boundary, indicating the low significance of the signal. Note that the posteriors for the sky location and inclination parameters also show some deviation from the

priors, but they are not particularly peaked around the true parameter values.

These analyses illustrate the expected progression of a GW detection from an individual SMBHB as we gather SNR over time. We expect to first see a peak emerging in the posteriors of f_{GW} and A_e , while basically recovering the priors of other parameters. These other parameters will start to have more informative posteriors as we gather more SNR

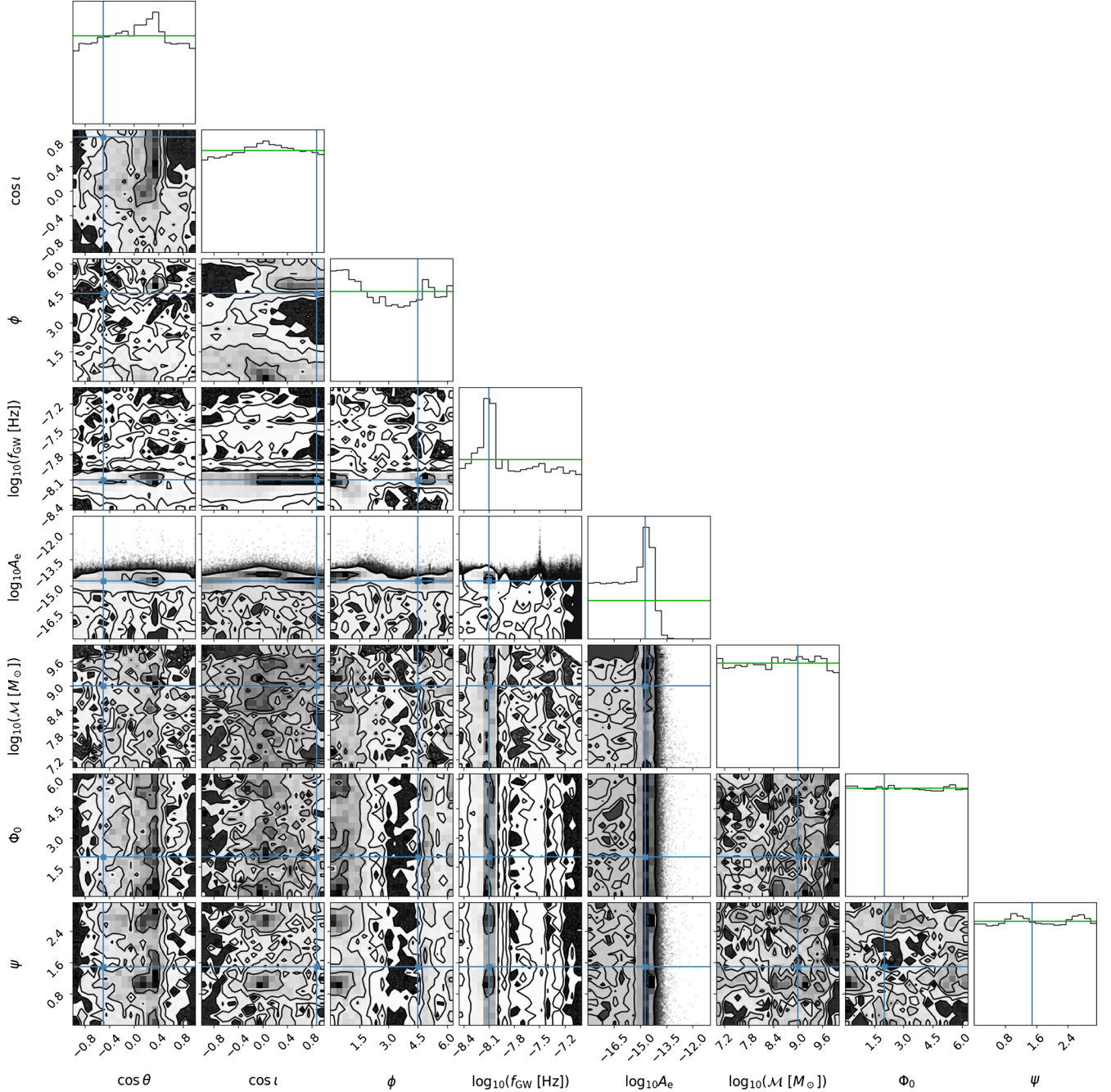


FIG. 6. Corner plot of parameters common to all pulsars for marginally detectable signal ($\text{SNR} = 4.3$, Dataset 3 in Table II). We show the posterior with black, the prior with green and the true values of parameters with blue.

(as we have seen on Figs. 2 and 4). The expected sequence of parameter constraint improvement will be important to keep in mind as we transition from placing upper limits to claiming detections. In Sec. IV B, we also investigate how the upper limit can be affected by a marginally detectable signal.

B. Upper limit analysis

If no significant GW candidates are found, we can place upper limits on the amplitude of GWs from individual

SMBHBs. Such an analysis poses slightly different challenges, because one needs to effectively explore the whole prior range in many parameters. Exploration of the full parameter space is required to ensure we gather a sufficient number of independent samples to get an accurate estimate of the upper limit as a function of different parameters. To test the performance of our pipeline in such a scenario, we analyzed a dataset similar to those discussed above, but with no GW signal added.

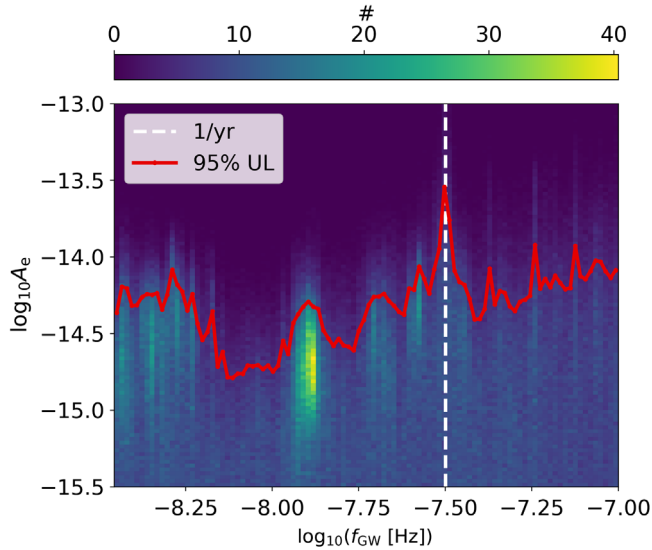


FIG. 7. GW amplitude upper limit as a function of GW frequency (red line). We also show the number of independent samples on the $\log_{10} f_{\text{GW}} - \log_{10} A_e$ plane. The vertical dashed line indicates the frequency of 1/yr, where we expect a significantly reduced sensitivity.

Figure 7 shows the upper limit on $\log_{10} A_e$ as a function of $\log_{10} f_{\text{GW}}$ for this simulated dataset. We also show the number of independent samples in each bin on the $\log_{10} f_{\text{GW}} - \log_{10} A_e$ plane. The upper limit is calculated by binning the samples in $\log_{10} f_{\text{GW}}$ and finding the 95th percentile of the amplitude in each bin. As expected, the dataset is most sensitive around 8 nHz ($\log_{10} f_{\text{GW}} \simeq -8.1$). At lower frequencies, we lose sensitivity due to the red noise present in the pulsars. At higher frequencies, we are progressively less sensitive due to the fact that we measure the integral of the GW signal. We also have particularly low sensitivity at $f_{\text{GW}} = 1/\text{yr}$ due to the degeneracy with Earth's orbital period introduced when we convert the observed TOAs to the SSB. Note that there is a peak in the posterior at $(\log_{10} f_{\text{GW}}, \log_{10} A_e) \simeq (-7.9, -14.7)$. The peak corresponds to a noise fluctuation being fitted with the GW model in this particular realization.

Even a pure noise dataset can exhibit features akin to marginal GWs. To investigate this feature, we analyzed a dataset with the same properties but a different random seed used for the noise realizations. Figure 8 shows the upper limit as a function of frequency for this dataset. We can see that while that peak does not show up in the alternate realization, other similar features appear at different parts of the parameter space. Candidate detections would need to be vetted with a full suite of cross-checks beyond the scope of this paper before a detection could officially be claimed. One possible approach is to reanalyze the dataset many times, while setting the sky location of the pulsars to random positions on the sky. This sky scrambling would be the same as currently used for the GWB [40,41], and since it destroys the coherence of the signal, it could be used the

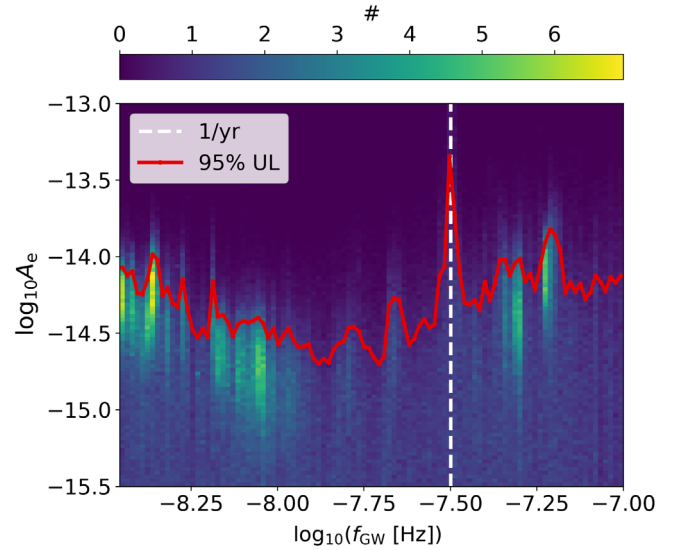


FIG. 8. Same as Fig. 7, but with a different noise realization. Note that bumps in the upper limit due to noise fluctuations appear at different random frequencies.

build a null distribution. Comparing the nonscrambled result with the null distribution gives a false alarm probability of the candidate. The dramatic speedup provided by our method significantly reduces the expense of conducting such reanalyses, which will help improve the robustness of future candidate vetting.

We also investigated how the upper limit changes in the presence of a marginal GW signal. Figure 9 shows the amplitude upper limit as a function of $\log_{10} f_{\text{GW}}$ for Dataset 3 from Table II. As a comparison, we also show the upper

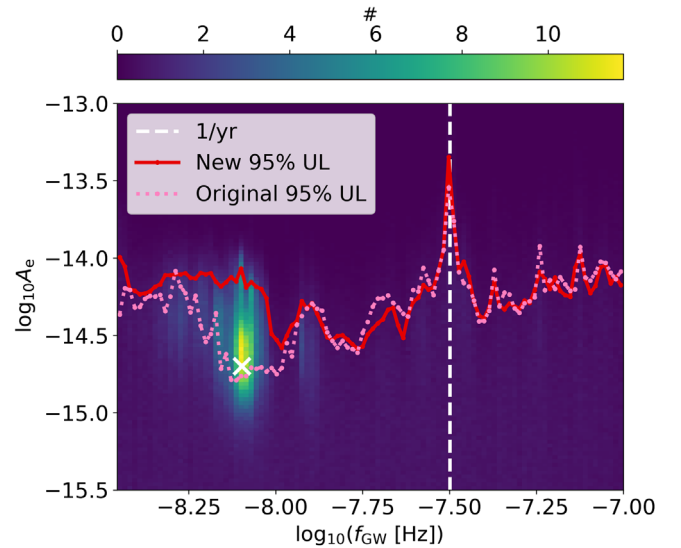


FIG. 9. Same as Fig. 7 but in the presence of a marginally detectable signal at $\log_{10} f_{\text{GW}} \simeq -8.1$. The red solid line shows the upper limit we get from analyzing this dataset, which is elevated around the signal compared to the original upper limit shown by the pink dotted line. The white cross indicates the true parameters of the signal.

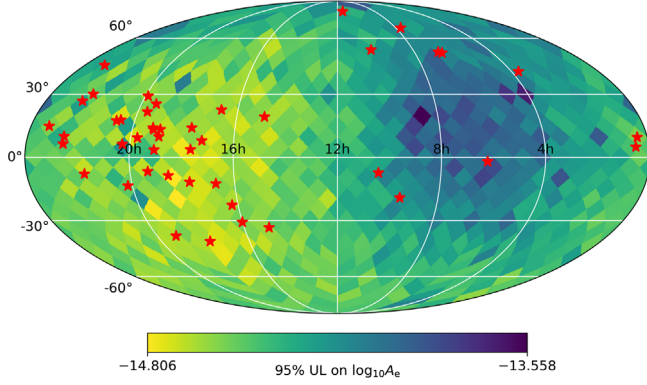


FIG. 10. Upper limit on $\log_{10} A_e$ as a function of sky location. Red stars indicate locations of the pulsars in the array.

limit we get from the dataset with no signal. We can see that the upper limit around the true frequency of the signal is elevated by almost an order of magnitude, even though the true amplitude is right at the level of the original upper limit. Note also that even this low-significance signal is much more prominent than the peak we have seen on Fig. 7 due to the noise fluctuation.

Figure 10 shows the frequency-marginalized upper limit we can place on $\log_{10} A_e$ as a function of sky location. There is more than an order of magnitude difference between the upper limit we can place towards the most and least sensitive sky location. The variable sky sensitivity is due to the fact that the pulsars in the NANOGrav 12.5-year dataset have a highly anisotropic distribution on the sky. That is expected for any pulsar timing array, because there are more pulsars towards the galactic center than in the antipodal direction.

V. RESULTS WITH REAL DATA

To validate our new methods with real data, we analyzed the NANOGrav 11-year dataset. Figure 11 shows the GW amplitude upper limit we get with `QuickCW` as a function of the GW frequency in red. We also show the official NANOGrav result in green [12]. Overall, there is good agreement between the two over the whole frequency range. Note that the official results show the amplitude upper limit at a set of fixed f_{GW} values, while our results show the upper limit at frequency bins with non-negligible width. Thus we do not expect perfect agreement between the two results, especially where the upper limit is quickly changing with frequency. To illustrate this, we also plot our results using narrow frequency bins around the fixed frequencies used in the official NANOGrav analysis (pink markers). We can see that at several frequencies, these are in better agreement with the official results (e.g., around $f_{\text{GW}} = 1/\text{yr}$). However, even these narrow frequency bin results show some discrepancy at the lowest frequency bin. We think this is due to a bug that was recently found in the so-called empirical distribution proposals, which were used

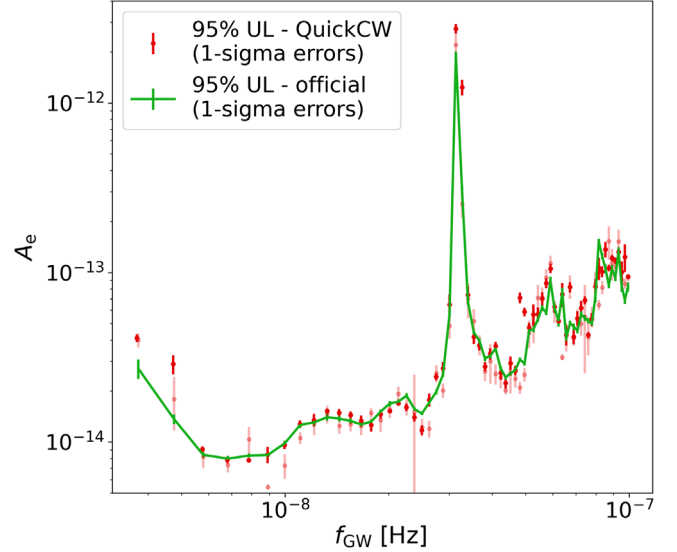


FIG. 11. GW amplitude upper limit as a function of GW frequency for the NANOGrav 11-year dataset analyzed by `QuickCW` (red), and as reported in Ref. [12] (green). We also show our results using narrow frequency bins centered at the fixed frequencies used in the official NANOGrav analysis (pink).

in the official analysis [12,42]. The bug resulted in a small overestimation of the red noise in some pulsars, which in turn meant that the upper limit on A_e was underestimated at low frequencies, where there is a strong correlation between the red noise and the individual binary models.

Allowing the GW frequency to explore the entire prior range has several advantages over running at a set of fixed frequencies. Especially at frequencies where the upper limit changes quickly, a fixed frequency analysis cannot fully explore parameter space and artificially underestimates the uncertainty in the upper limits. Varying the frequency also streamlines the analysis, since we only need a single run instead of dozens. A potential issue is that we do not get the same number of samples at all frequencies. Excessive focus on a particular set of frequencies is suboptimal, since if we require a fixed level of convergence in each frequency bin, the overall runtime is determined by the frequency bin with the least number of samples. To optimize the sampling, we could apply a pseudoprior on f_{GW} , which down-weights regions with many samples based on a pilot run. The pseudoprior would then ensure a more uniform number of samples at all frequencies while preserving good mixing. The undesired bias from the pseudoprior can be canceled in postprocessing by reweighting the posterior samples. This technique is similar to umbrella sampling [43], and the pseudoprior used when calculating Bayes factors with the product space method [44].

VI. CONCLUSION

In this paper we presented a new formulation of the likelihood function [see Eq. (22)], which results in a

significant speedup of a Bayesian search for individual SMBHBs in PTA data. Our formulation does not apply any new approximations, and recovers the canonical likelihood within numerical errors. This is achieved by separating the parameter space to shape parameters and projection parameters. Precalculating inner products for a given set of shape parameters allows cheap exploration of the projection parameters, thus speeding up the entire analysis. We demonstrated the performance using a new analysis pipeline employing the new likelihood in a Metropolis-within-Gibbs sampler with multiple-try MCMC.

These methods will drastically reduce the computational cost of searches for individual sources in upcoming PTA datasets, and improve the tractability of achieving a well-converged analysis as the size of datasets increases. Since all the latest PTA datasets show evidence for the presence of a common red noise process [2–5], we plan to incorporate that in our model as well. This would appear as an additional red noise term in the T and B matrices defined below Eq. (20), and would introduce two additional parameters to sample over, which describe the amplitude and spectral slope of the common red noise. This addition would not significantly change how the fast likelihood methods presented in this paper work. If the upcoming PTA datasets will show evidence for this common process being correlated between pulsars, we would ultimately want to include those correlations in our model as well. In Appendix A we outline how such correlations could be incorporated into the fast-likelihood framework we presented in this paper. The implementation of that addition will be presented in future work.

These methods will also make more sophisticated analyses computationally feasible. In particular, we plan to extend these methods to work with BayesHopper [22], a pipeline proposed to search for multiple individual sources simultaneously. We also plan to work on implementing similar methods for a search for eccentric SMBHBs and a search for non-Einsteinian polarization modes from SMBHBs [45]. It might also be worthwhile to extend these methods to sine-Gaussian wavelets in an effort to speedup BayesHopperBurst [46], a pipeline to search for generic GW bursts in PTA data.

ACKNOWLEDGMENTS

The authors thank the organizers of the 2021 NANOGrav Fall Meeting, which provided inspiration for this project. We thank Ken Olum and Stephen Taylor for feedback on the manuscript, and Sarah Vigeland for helping reproduce the NANOGrav 11-year results. We used the corner [47] software package to produce the corner plots presented in this paper. Some of the results in this paper have been derived using the healpy [48] and HEALPix [49] package. We appreciate the support of the NSF Physics Frontiers Center Award No. PFC-1430284 and NASA LISA foundation Science Grant No. 80NSSC19K0320.

APPENDIX A: FAST LIKELIHOOD IN THE PRESENCE OF CORRELATED RED NOISE

A red noise process can be modeled using a pseudo-Fourier basis $\mathbf{F}$, made up of a collection of sines and cosines at a discrete set of frequencies with amplitudes $\mathbf{a}$ [50]. The log likelihood then becomes

$$\log L = -\frac{1}{2}(\delta t - s - \mathbf{F} \cdot \mathbf{a} | \delta t - s - \mathbf{F} \cdot \mathbf{a}). \quad (\text{A1})$$

The usual approach is to marginalize over the $\mathbf{a}$'s analytically. If we assume the signal is not correlated between pulsars, Eq. (A1) can still be evaluated pulsar-by-pulsar [51]. However, the per-pulsar factorization breaks down if we intend to model a GWB with Hellings-Downs correlations between the pulsars [52]. This would significantly increase the cost of the CW analysis since the M^{kl} matrix would become dense.

Another approach is to use a method similar to what we are proposing for individual SMBHBs and to the methods presented in Ref. [53]. For simplicity let us ignore the individual sources for now and focus on the GWB. Expanding the log likelihood we have

$$\log L = -\frac{1}{2}[(\delta t | \delta t) - 2\mathbf{a} \cdot \mathbf{P} + \mathbf{a} \cdot \mathbf{Q} \cdot \mathbf{a}], \quad (\text{A2})$$

where $\mathbf{P} = (\delta t | \mathbf{F})$ and $\mathbf{Q} = (\mathbf{F} | \mathbf{F})$. Note that the vector $\mathbf{P}$ is much smaller than the vector $\mathbf{F}$. The vector $\mathbf{F}$ is a vector of vectors, made up of the time samples for the sines and cosines at each frequency and for each pulsar. After the inner products have been done, the $\mathbf{P}$ is a collection of numbers, one each for the sine and cosine at each frequency in each pulsar. The matrix $\mathbf{Q}$ has rows and columns that follow the pattern of the vector $\mathbf{P}$. To simplify the discussion, consider a single frequency f , and label the $\mathbf{a}$ such that $a_{(2i-1)}$ are the cosine terms in pulsar i and $a_{(2i)}$ are the sine terms. Our hyperprior for the $\mathbf{a}$ is such that each a_k is drawn from a Gaussian distribution with variance $S(f)$ and correlations given by

$$\begin{aligned} E[a_{(2i-1)}a_{(2j-1)}] &= H_{ij}S(f), \\ E[a_{(2i-1)}a_{(2j)}] &= 0, \\ E[a_{(2i)}a_{(2j)}] &= H_{ij}S(f), \end{aligned} \quad (\text{A3})$$

where H_{ij} is the Hellings-Downs correlation between pulsars i , and j . More schematically we can write $E[a_k a_l] = \varphi_{kl}$ where the entries of φ_{kl} are given by (A3). The posterior probability distribution for $\mathbf{a}$ can then be written as

$$\begin{aligned} p(\mathbf{a} | \delta t, S) &= \frac{e^{-\frac{1}{2}(\delta t | \delta t)}}{\sqrt{\det(2\pi C) \det(2\pi \varphi)}} \\ &\times e^{a_k P^k - \frac{1}{2}a_k a_l \varphi^{kl} - \frac{1}{2}a_k a_l (\varphi^{-1})^{kl}}. \end{aligned} \quad (\text{A4})$$

The next step is to marginalize the $p(\mathbf{a}|\delta t, S)$ over $\mathbf{a}$ to produce the marginal likelihood $p(\delta t|S)$. This marginalization can be done analytically since the expression for $p(\mathbf{a}|\delta t, S)$ is a multivariate Gaussian. The end product is the new likelihood for the power at frequency f . Notice that all the inner products were done pulsar by pulsar. We have a factorized likelihood. Unlike in the usual analysis where the marginalization is done before computing the inner products, by performing the marginalization after the per-pulsar inner products have been done we avoid any cross terms. The analysis still “knows” about the correlations, without actually having to cross-correlate the data between pulsars.

Note that the inner products $\mathbf{P}$ and $\mathbf{Q}$ have to be recomputed every time the noise model is updated. The red noise can be treated like the GWB, but just with a diagonal correlation matrix, so $\mathbf{P}$ and $\mathbf{Q}$ only have to be recalculated for white noise updates. The white noise inner products could be computed for some discrete set of values for each pulsar, then a joint marginalization across the white noise in all the pulsars could be done using a lookup table.

APPENDIX B: IMPLEMENTATION IN PYTHON USING Numba

The new formulation of the likelihood and the associated sampler is implemented in the `QuickCW` package. `QuickCW` relies on the `ENTERPRISE` software [54], which uses `PYTHON`, so it was a natural choice to implement `QuickCW` in `PYTHON` as well. To overcome the inherent speed limitations of `PYTHON`, we use the `Numba` software package [55,56], which is a just-in-time compiler capable of generating fast machine code from `PYTHON` syntax. This is particularly important, as in the new formulation of the likelihood, the computational cost after the inner products have been precomputed consists of a small number of simple multiplicative and additive operations. As a result, run times are dominated by `PYTHON`-specific overheads if we do not use `Numba`.

Table III shows representative runtimes of the old and new likelihood for different scenarios. We carried out these tests on an AMD Ryzen Threadripper 3970X 32-core processor. We tested on three different simulated datasets:

one made to resemble the NANOGrav 12.5-year dataset (same as used for all other results in this paper); one with double the number of pulsars and TOAs as the NANOGrav 12.5-year dataset; and one with five times as many pulsars and TOAs as the NANOGrav 12.5-year dataset. The latter two were used to show how performance will change in the future as the PTA datasets grow in size.

For each of these datasets we timed the execution of the old likelihood for randomly drawn GW parameters and randomly drawn red noise parameters. We can see in Table III that these take a different amount of time. The variation is due to the fact that `ENTERPRISE` caches some of its internal function calls, so when the red noise is changed, some parts of the likelihood can be reused from before, which results in different execution times. We also tested the new likelihood formulation in a scenario when only projection parameters are changed. We can see that for the dataset resembling the NANOGrav 12.5-year, this is $\sim 20,000$ times faster than the old likelihood. We also timed the recalculation of the inner products, which is necessary when the shape parameters are updated. We can see that recalculating the filters for a new set of shape parameters is still about a factor of three faster than the old likelihood. The speedup is due to a combination of algorithmic optimization and the speedup we get with `Numba`. If GW parameters are updated, the N^k and M^{kl} needs to be recomputed using the updated S^k , but one can use C^{-1} from memory. However, when the red noise parameters are updated, S^k remains the same, but an updated C^{-1} needs to be used. As we can see in Table III, this results in the red noise updates being more expensive.

The results for larger datasets show that runtimes for the old likelihood roughly scale linearly with the number of pulsars and TOAs. Evaluating the new likelihood at given shape parameters scales slower than linear, resulting in less than twice as long runtimes for the 5 times larger dataset. On the other hand, shape parameter updates scale faster than linear, resulting in about a factor of 8 slower evaluation for the 5 times larger dataset. Note, however, that we optimized the pipeline for a dataset like the NANOGrav 12.5-year, so the performance for larger datasets might be suboptimal.

TABLE III. Runtimes of old likelihood and different components of the new likelihood for various representative datasets.

	NANOGrav 12.5-year	$2 \times$ NANOGrav 12.5-year	$5 \times$ NANOGrav 12.5-year
Number of pulsars	45	90	225
Number of TOAs	410,064	820,128	2,050,320
Old likelihood GW update	300 ms	610 ms	1500 ms
Old likelihood red noise update	120 ms	210 ms	630 ms
New likelihood ^a	0.016 ms	0.020 ms	0.029 ms
GW shape parameter update ^b	90 ms	220 ms	770 ms
Pulsar red noise update	180 ms	400 ms	1300 ms

^aWith fixed shape parameters.

^bCommon parameters or pulsars distances.

- [1] Stephen R. Taylor, [arXiv:2105.13270](https://arxiv.org/abs/2105.13270).
- [2] Zaven Arzoumanian, Paul T. Baker, Harsha Blumer *et al.*, *Astrophys. J. Lett.* **905**, L34 (2020).
- [3] S. Chen, R. N. Caballero, Y. J. Guo *et al.*, *Mon. Not. R. Astron. Soc.* **508**, 4970 (2021).
- [4] Boris Goncharov, R. M. Shannon, D. J. Reardon *et al.*, *Astrophys. J. Lett.* **917**, L19 (2021).
- [5] J. Antoniadis, Z. Arzoumanian, S. Babak *et al.*, *Mon. Not. R. Astron. Soc.* **510**, 4873 (2022).
- [6] Joseph D. Romano, Jeffrey S. Hazboun, Xavier Siemens, and Anne M. Archibald, *Phys. Rev. D* **103**, 063027 (2021).
- [7] Nihan S. Pol, Stephen R. Taylor, Luke Zoltan Kelley *et al.*, *Astrophys. J. Lett.* **911**, L34 (2021).
- [8] Pablo A. Rosado, Alberto Sesana, and Jonathan Gair, *Mon. Not. R. Astron. Soc.* **451**, 2417 (2015).
- [9] Luke Zoltan Kelley, Laura Blecha, Lars Hernquist, Alberto Sesana, and Stephen R. Taylor, *Mon. Not. R. Astron. Soc.* **477**, 964 (2018).
- [10] Zaven Arzoumanian, Paul T. Baker, Harsha Blumer *et al.*, *Phys. Rev. Lett.* **127**, 251302 (2021).
- [11] Maria Charisi, Stephen R. Taylor, Jessie Runnoe, Tamara Bogdanovic, and Jonathan R. Trump, *Mon. Not. R. Astron. Soc.* **510**, 5929 (2022).
- [12] K. Aggarwal, Z. Arzoumanian, P. T. Baker *et al.*, *Astrophys. J.* **880**, 116 (2019).
- [13] S. Babak, A. Petiteau, A. Sesana *et al.*, *Mon. Not. R. Astron. Soc.* **455**, 1665 (2016).
- [14] X. J. Zhu, G. Hobbs, L. Wen *et al.*, *Mon. Not. R. Astron. Soc.* **444**, 3709 (2014).
- [15] Vincent Corbin and Neil J. Cornish, [arXiv:1008.1782](https://arxiv.org/abs/1008.1782).
- [16] 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).
- [17] J. A. Ellis, *Classical Quantum Gravity* **30**, 224004 (2013).
- [18] Stephen Taylor, Justin Ellis, and Jonathan Gair, *Phys. Rev. D* **90**, 104028 (2014).
- [19] B. P. Abbott, R. Abbott, T. D. Abbott *et al.*, *Classical Quantum Gravity* **37**, 055002 (2020).
- [20] S. R. Taylor, E. A. Huerta, J. R. Gair, and S. T. McWilliams, *Astrophys. J.* **817**, 70 (2016).
- [21] Abhimanyu Susobhanan, Achamveedu Gopakumar, George Hobbs, and Stephen R. Taylor, *Phys. Rev. D* **101**, 043022 (2020).
- [22] Bence Bécsy and Neil J. Cornish, *Classical Quantum Gravity* **37**, 135011 (2020).
- [23] Publicly available at: <https://github.com/bencebecsy/QuickCW>.
- [24] J. A. Ellis, X. Siemens, and J. D. E. Creighton, *Astrophys. J.* **756**, 175 (2012).
- [25] Zaven Arzoumanian, Adam Brazier, Sarah Burke-Spolaor *et al.*, *Astrophys. J. Suppl. Ser.* **235**, 37 (2018).
- [26] Piotr Jaranowski, Andrzej Królak, and Bernard F. Schutz, *Phys. Rev. D* **58**, 063001 (1998).
- [27] Curt Cutler and Bernard F. Schutz, *Phys. Rev. D* **72**, 063006 (2005).
- [28] Neil J. Cornish and Edward K. Porter, *Classical Quantum Gravity* **24**, 5729 (2007).
- [29] Stanislav Babak and Alberto Sesana, *Phys. Rev. D* **85**, 044034 (2012).
- [30] Note: the following notation follows the sign convention of the ENTERPRISE software package, which is slightly different than the conventions usually used in the literature.
- [31] Yan Bai, An Adaptive Directional Metropolis-within-Gibbs algorithm, 2009, https://www.researchgate.net/publication/228389131_An_Adaptive_Directional_Metropolis-within-Gibbs_algorithm.
- [32] Jun S. Liu, Faming Liang, and Wing Hung Wong, *J. Am. Stat. Assoc.* **95**, 121 (2000).
- [33] Luca Martino, *Digit. Signal Process.* **75**, 134 (2018).
- [34] Radu Craiu and Christiane Lemieux, *Stat. Comput.* **17**, 109 (2007).
- [35] Roberto Casarin, Radu V. Craiu, and Fabrizio Leisen, *Stat. Comput.* **23**, 185 (2013).
- [36] Md F. Alam, Zaven Arzoumanian, Paul T. Baker *et al.*, *Astrophys. J. Suppl. Ser.* **252**, 4 (2021).
- [37] Janna M. Goldstein, Alberto Sesana, A. Miguel Holgado, and John Veitch, *Mon. Not. R. Astron. Soc.* **485**, 248 (2019).
- [38] Zaven Arzoumanian, Paul T. Baker, Adam Brazier *et al.*, *Astrophys. J.* **900**, 102 (2020).
- [39] Tingting Liu and Sarah J. Vigeland, *Astrophys. J.* **921**, 178 (2021).
- [40] Neil J. Cornish and Laura Sampson, *Phys. Rev. D* **93**, 104047 (2016).
- [41] S. R. Taylor, L. Lentati, S. Babak, P. Brem, J. R. Gair, A. Sesana, and A. Vecchio, *Phys. Rev. D* **95**, 042002 (2017).
- [42] Note that this issue was recently resolved here: https://github.com/nanograv/enterprise_extensions/commit/3a17943dddf-005867cdbaa960d4e72b747eb373.
- [43] Charles Matthews, Jonathan Weare, Andrey Kravtsov, and Elise Jennings, *Mon. Not. R. Astron. Soc.* **480**, 4069 (2018).
- [44] Bradley P. Carlin and Siddhartha Chib, *J. R. Stat. Soc.* **57**, 473 (1995).
- [45] Logan O’Beirne, Neil J. Cornish, Sarah J. Vigeland, and Stephen R. Taylor, *Phys. Rev. D* **99**, 124039 (2019).
- [46] Bence Bécsy and Neil J. Cornish, *Classical Quantum Gravity* **38**, 095012 (2021).
- [47] Daniel Foreman-Mackey, *J. Open Source Software* **1**, 24 (2016).
- [48] Andrea Zonca, Leo Singer, Daniel Lenz, Martin Reinecke, Cyrille Rosset, Eric Hivon, and Krzysztof Gorski, *J. Open Source Software* **4**, 1298 (2019).
- [49] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, *Astrophys. J.* **622**, 759 (2005).
- [50] Lindley Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren, *Phys. Rev. D* **87**, 104021 (2013).
- [51] Stephen R. Taylor, Joseph Simon, Levi Schult, Nihan Pol, and William G. Lamb, *Phys. Rev. D* **105**, 084049 (2022).
- [52] Neil J. Cornish and Joseph D. Romano, *Phys. Rev. D* **87**, 122003 (2013).
- [53] Xavier Siemens, William G. Lamb, Stephen R. Taylor, Joseph D. Romano, Sarah J. Vigeland, Ken Olum, Aaron Johnson, Michele Vallisneri, Jeffrey S. Hazboun, and Joseph Simon, NANOGrav Fall Meeting, 2021.

-
- [54] Justin A. Ellis, Michele Vallisneri, Stephen R. Taylor, and Paul T. Baker, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust Pulsar Inference Suite (Zenodo, 2020), <https://doi.org/10.5281/zenodo.4059815>.
- [55] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15, New York, NY, USA, 2015. Association for Computing Machinery.
- [56] Siu Kwan Lam, stuartarchibald, Antoine Pitrou *et al.*, numba/numba: Version 0.55.0 (Zenodo, 2022), <https://doi.org/10.5281/zenodo.5847553>.