arXiv:1606.09180v1 [astro-ph.IM] 29 Jun 2016

All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar timing arrays

S. R. Taylor,[1,*] L. Lentati,[2] S. Babak,[3] P. Brem,[3] J. R. Gair,[4] A. Sesana,[5] and A. Vecchio[5]

[1] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91106, USA*
[2] *Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, UK*
[3] *Max-Planck-Institut für Gravitationsphysik, Albert Einstein Institut, Am Mühlenberg 1, 14476 Golm, Germany*
[4] *School of Mathematics, University of Edinburgh, King's Buildings, Edinburgh EH9 3JZ, UK*
[5] *School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*

(Dated: June 30, 2016)

We present two methods for determining the significance of a stochastic gravitational-wave background affecting a pulsar-timing array, where detection is based on recovering evidence for correlations between different pulsars, i.e. spatial correlations. Nulling these spatial correlations is crucial to understanding the response of our detection statistic under the null hypothesis so that we can properly assess the significance of plausible signals. The usual approach of creating many noise-only simulations is, albeit useful, undesirable since in that case detection significance is predicated on our (incomplete) understanding of all noise processes. Alternatively, destroying any possible correlations in our real datasets and using those (containing all actual noise features) is a much superior approach. In our first method, we perform random phase shifts in the signal-model basis functions. This phase shifting is performed separately for each sampling frequency of each pulsar in the timing array, having the effect of eliminating signal phase coherence between pulsars, while keeping the statistical properties of the pulsar timing residuals intact. We also explore a method to null correlations between pulsars by using a "scrambled" overlap-reduction function in the signal model for the array. This scrambled overlap-reduction function should be effectively orthogonal to what we expect of a real background signal. We demonstrate the efficacy of these methods in a set of simulated datasets that contain a stochastic gravitational wave background, using Bayesian model selection to compare models that do, or do not, account for the correlation between pulsars induced by this signal. Finally, we introduce an overarching formalism under which these two techniques can be seen as natural companions to each other. These methods are immediately applicable to all current pulsar-timing array datasets, and should become standard tools for future analyses.

I. INTRODUCTION

The existence of gravitational waves (GWs) was recently confirmed with the detection of a binary black-hole merger by LIGO, ushering in the era of GW astronomy [1]. This detection relied on (amongst other factors) precision engineering, extensive theoretical development, and detailed detector noise characterization. The latter is incredibly important because it dictates to what degree we are confident that the detector output contains a signal rather than a spurious noise feature. Methods for this (such as "time sliding") are well-developed in the ground-based and space-based interferometry literature, but were heretofore lacking for pulsar-timing arrays [PTA, 14]. We explore such methods here.

In the pulsar timing experiment, an ensemble of millisecond pulsars makes use of the expected form for the spatial correlation of the GW signal between pulsars in the array to discriminate between the GW signal of interest and other sources of noise in the data. These latter noise processes are intrinsic to each pulsar, and include intrinsic spin-noise due to rotational irregularities [e.g. 34], or delays in the pulse arrival time due to propagation through the interstellar medium [e.g. 22]. In the specific case of an isotropic stochastic gravitational-wave background (GWB), this spatial correlation is quadrupolar, and known as the *Hellings and Downs* curve [17]. This correlation signature is only a function of the angular separation between pairs of pulsars in the array, although there are more general correlation signatures for anisotropic backgrounds [15, 29, 39], and GWBs composed of non-GR polarisations [7, 16]

Upper limits on an isotropic stochastic gravitational wave background from the three main pulsar-timing arrays (PPTA [35]; EPTA [27]; NANOGrav [3]) are now reaching the sensitivities required to constrain current models of backgrounds generated by a population of supermassive black hole binaries [e.g., 24, 28, 30, 32, 33]. Recent projections suggest that there is significant probability that a stochastic GWB will be detected within the

* Stephen.R.Taylor@jpl.nasa.gov

next decade [31, 37, 41].

Several detection statistics exist for a GWB signal in pulsar timing data. Frequentist methods such as the "optimal statistic" [2, 6] measure how likely it is (in terms of number of standard deviations from zero) that a cross-correlated signal is present in our data rather than a common uncorrelated signal. This therefore implicitly assumes that all cross correlated power comes from one signal – in our case a GWB – rather than from several other sources [42] such as errors in the time standard [18], or in the Solar System ephemeris [8].

Bayesian methods instead make use of the fully-marginalized likelihood (or "evidence") to determine the probability of one model over another. This allows for a statistically robust comparison of models that includes contributions to the correlated signal from clock or Solar System ephemeris errors, and can also be used to perform model comparison for more general scenarios, when the correlation between pulsars has been modeled either using a smooth functional form [40], or pairwise for each pulsar pair [26]. The shortcoming of this approach is the explicit dependence on the appropriateness of the models being used for the evidence comparison. In pulsar timing, every data point contains contributions from a potential GWB, timing-noise intrinsic to the pulsar, and any additional systematics such as unmodelled system offsets or calibration errors. In order to obtain a convincing detection of a GWB, it will therefore be necessary to show that the observed signal truly is a result of quadrupolar spatial correlations between pulsars, and not simply a mismodelling of the intrinsic noise, or systematic effects present in the timing data.

In this paper we present two possible approaches to determining the significance of correlated signals present in pulsar timing data. The first approach exploits correlations in phase, where random phase shifts are introduced between pulsars to destroy signal phase coherence, but which preserve the statistical properties of the individual pulsar datasets. The second approach exploits the spatial correlations of many expected correlated signals. By "scrambling" the positions of pulsars on the sky (and therefore their angular separations) we will incorrectly sample these expected correlation functions, which nulls any cross-correlated signal in question. We note that, although in the following we frame the problem and analysis in a Bayesian context, these phase-shifting and sky-scrambling techniques can be straightforwardly applied to real datasets with frequentist detection statistics, such as the aforementioned optimal-statistic. Previous work in Cornish and Sampson [9] has investigated the requirments of a robust detection of GWs in PTAs, but focused on issues regarding the position scrambling approach.

In Section II we introduce the pulsar-timing likelihood to be used in our Bayesian analysis. In Section III we discuss the phase and spatial correlations of pulsars due to the influence of a common GWB, and introduce our two methods of "phase shifting" and "sky scrambling". We also demonstrate that both approaches are equivalent in terms of their effect on the correlated signal. These methods are applied in Sec. IV to (*i*) an idealized simulation, (*ii*) a more realistic simulation, and (*iii*) a simulation for which our noise model is incomplete, where we show the superiority of these methods over repeated noise-only simulations. We summarize our findings in Sec. V.

II. A PULSAR TIMING LIKELIHOOD

For any pulsar we can write the times of arrival (TOAs) for the pulses as a sum of both a deterministic and a stochastic component:

$$\mathbf{t}_{\rm tot} = \mathbf{t}_{\rm det} + \mathbf{t}_{\rm sto}, \tag{1}$$

where $\mathbf{t}_{\rm tot}$ is a vector of length $N_{\rm TOA}$ for a single pulsar, with $\mathbf{t}_{\rm det}$ and $\mathbf{t}_{\rm sto}$ the deterministic and stochastic contributions to the total respectively. Any contributions to the latter will be modelled as random Gaussian processes. In estimating the timing model parameters for the pulsar (either through a standard weighted least-squares fit, as performed in packages such as TEMPO2 [10, 19], or through a Bayesian analysis such as performed by the pulsar timing package TempoNest [25]) an initial estimate for the m timing model parameters included in the model can be obtained, which we denote $\boldsymbol{\beta_0}$. This in turn allows us to generate an initial set of timing residuals, which we denote $\delta\mathbf{t}$.

We now assume that the difference between this initial solution $\boldsymbol{\beta_0}$, and the final solution $\boldsymbol{\beta_f}$ obtained from a joint analysis that includes a GWB term will be small. Therefore a linear approximation of the timing model can be used such that any deviations from the initial guess of the timing model parameters are encapsulated using the vector $\boldsymbol{\epsilon}$ of length m, such that $\epsilon_i = \beta_{{\rm f}i} - \beta_{0i}$. These small timing model deviations influence the timing residuals via $\mathbf{M}\boldsymbol{\epsilon}$, where $\mathbf{M}$ is the $N_{\rm TOA} \times m$ timing model "design matrix" describing the dependence of the residuals on the timing model parameters.

Furthermore, we include the influence of all low-frequency processes on the timing residuals (such as intrinsic spin-noise, a common red-noise process, and a GWB) via the term $\mathbf{Fa}$. The vector $\mathbf{a}$ of length $2N_{\rm freqs}$ describes the Fourier coefficients of any low-frequency process at a limited number of harmonics of the base sampling frequency $1/T$ (where T is the observation timespan of a single pulsar, or the maximum coverage of the entire pulsar timing array), and $\mathbf{F}$ is the $N_{\rm TOA} \times 2N_{\rm freqs}$ "Fourier design matrix" consisting of alternating columns of sines and cosines.

We can also explicitly include the influence of white-noise terms on the timing residuals, such as from TOA measurement uncertainties (which may be modified by additional system-dependent scaling parameters such as EFACs and EQUADs), or correlated measurement uncertainties in simultaneous multi-frequency observations

(ECORR). However we implicitly marginalize over these effects in the following such that their influence is confined to the $N_{\rm TOA} \times N_{\rm TOA}$ white noise covariance matrix, $\mathbf{N}$, for each pulsar.

The model-dependent timing residuals, $\mathbf{r}$, for each pulsar can thus be written in terms of the input residuals, $\delta\mathbf{t}$ as

$$\mathbf{r} = \delta\mathbf{t} - \mathbf{M}\boldsymbol{\epsilon} - \mathbf{F}\mathbf{a}, \tag{2}$$

with a likelihood given by

$$p(\delta\mathbf{t}|\boldsymbol{\epsilon}, \mathbf{a}, \boldsymbol{\eta}) = \frac{\exp\left(-\frac{1}{2}\mathbf{r}^T\mathbf{N}^{-1}\mathbf{r}\right)}{\sqrt{\det(2\pi\mathbf{N})}}, \tag{3}$$

where $\boldsymbol{\eta}$ encapsulates any parameters not already represented by $\boldsymbol{\epsilon}$ or $\mathbf{a}$.

We group all low-frequency and reduced rank signals into a common description, such that

$$\mathbf{r} = \delta\mathbf{t} - \mathbf{T}\mathbf{b}, \tag{4}$$

where

$$\mathbf{T} = \begin{bmatrix}\mathbf{M} & \mathbf{F}\end{bmatrix}, \ \mathbf{b} = \begin{bmatrix}\boldsymbol{\epsilon} \\ \mathbf{a}\end{bmatrix}. \tag{5}$$

This allows us to put a prior on the coefficients, $\mathbf{b}$, which we choose such that the underlying processes are Gaussian. Therefore,

$$p(\mathbf{b}|\boldsymbol{\phi}) = \frac{\exp\left(-\frac{1}{2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}\right)}{\sqrt{\det(2\pi\mathbf{B})}}, \tag{6}$$

with,

$$\mathbf{B} = \begin{bmatrix}\boldsymbol{\infty} & 0 \\ 0 & \boldsymbol{\varphi}\end{bmatrix}, \tag{7}$$

such that the timing model portion of $\mathbf{b}$ has an infinite variance to approximate a uniform unconstrained prior on small timing model parameter deviations, $\boldsymbol{\epsilon}$.

The low-frequency portion of $\mathbf{b}$ has a variance, φ, given by the spectrum of all low-frequency processes in the data. Since this may include a GWB we must naturally model spatial correlations in the data:

$$[\varphi]_{(ai),(bj)} = \Gamma_{ab}\rho_i\delta_{ij} + \kappa_{ai}\delta_{ab}\delta_{ij}, \tag{8}$$

where κ_{ai} is the intrinsic low-frequency ("spin-noise") spectrum of pulsar a at the $i^{\rm th}$ sampling frequency; ρ_i is the GWB spectrum at the $i^{\rm th}$ sampling frequency; and Γ_{ab} is the overlap reduction function (ORF) between pulsars a and b describing the reduction in correlated power due to spatial separation of the pulsars. For an isotropic sotchastic GWB this Γ_{ab} depends only on the separation between pulsars and is commonly known as the "Hellings and Downs curve". We note that both κ and ρ can either be modelled with a functional form (such as a power-law or a smooth turnover) or as a free spectrum with a parameter per frequency. In the following we consider all low-frequency processes to be well described by power-law spectra at all sampling frequencies ν_i, such that (taking κ_{ai} as an example)

$$\kappa_{ai} = \frac{A_a^2}{12\pi^2}\frac{1}{T}\left(\frac{\nu_i}{1{\rm yr}^{-1}}\right)^{-\gamma_a} {\rm yr}^2. \tag{9}$$

For a GWB the exponent has a value of $\gamma = 13/3$ for a circular GW-driven population of SMBHBs. All intrinsic red noise and GWB power-law spectral parameters are grouped into the parameter vector $\boldsymbol{\eta}$.

We can now write the joint probability density of the timing model and reduced rank signal parameters, $p(\mathbf{b}, \boldsymbol{\eta}|\delta\mathbf{t})$, as:

$$p(\mathbf{b}, \boldsymbol{\eta}|\delta\mathbf{t}) \propto p(\delta\mathbf{t}|\mathbf{b}) \times p(\mathbf{b}|\boldsymbol{\eta}) \times p(\boldsymbol{\eta}). \tag{10}$$

Taking the logarithm of Eq. (10) and extremizing gives the maximum likelihood vector of coefficients $\widehat{\mathbf{b}}$:

$$\widehat{\mathbf{b}} = \boldsymbol{\Sigma}^{-1}\mathbf{d}, \tag{11}$$

where $\boldsymbol{\Sigma} = (\mathbf{T}^T\mathbf{N}^{-1}\mathbf{T} + \boldsymbol{\varphi}^{-1})$ and $\mathbf{d} = \mathbf{T}^T\mathbf{N}^{-1}\delta\mathbf{t}$.

We can also analytically marginalize Eq. (10) over the coefficients $\mathbf{b}$, giving:

$$p(\boldsymbol{\eta}|\delta\mathbf{t}) \propto \frac{\exp\left(-\frac{1}{2}\delta\mathbf{t}^T\mathbf{C}^{-1}\delta\mathbf{t}\right)}{\sqrt{\det(2\pi\mathbf{C})}} \times p(\boldsymbol{\eta}), \tag{12}$$

where $\mathbf{C} = \mathbf{N} + \mathbf{T}\mathbf{B}\mathbf{T}^{\mathbf{T}}$. In practice, the Woodbury matrix identity [44] is used to reduce Eq. (12) to lower rank operations and thus accelerate computations.

III. DESTROYING SIGNAL COVARIANCE

Our aim in the following is to operate on real pulsar timing datasets in order to assess the statistical significance of any potential GWB signal in the data. We want to mitigate the cross-pulsar influence of any possible GWB already present, but keep the statistical properties of each individual pulsar dataset the same. The alternative to this procedure would be to simulate pulsar datasets with many different realizations of the noise processes we find in the real dataset, and thus assess the significance of findings in our real dataset against the distribution of detection statistics in noise-only datasets.

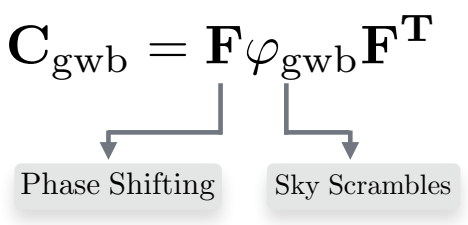


FIG. 1: The two detection significance techniques presented here for application to real pulsar timing datasets are naturally linked through the common result of destroying cross-pulsar GWB signal correlations.

This is undesirable since it requires us to have a complete description of the noise model of each pulsar, which is not likely achievable in practice and adds a level of uncertainty in any subsequent claims of GWB detection significance.

The reduced rank description of the GW signal covariance (as presented above) provides two natural avenues of inquiry for us to remove the GWB's correlated influence between pulsars. We recall that the time-domain covariance between pulsar TOAs which is induced by a GWB takes the following form:

$$\begin{aligned}\mathbf{C}_{\rm gwb} &= \langle \mathbf{F}\,\mathbf{a}_{\rm gwb}\mathbf{a}^*_{\rm gwb}\,\mathbf{F}^T\rangle \\ &= \mathbf{F}\,\langle \mathbf{a}_{\rm gwb}\mathbf{a}^*_{\rm gwb}\rangle\,\mathbf{F}^T \\ &= \mathbf{F}\boldsymbol{\varphi}_{\rm gwb}\mathbf{F}^T, \end{aligned} \tag{13}$$

where $\mathbf{F}$ is the Fourier design matrix of the signal, composed of length $N_{\rm TOA}$ columns of the signal basis functions (sines and cosines of harmonics of the base frequency, $1/T$), and $\boldsymbol{\varphi}_{\rm gwb}$ is the variance of the zero-mean signal coefficients $\mathbf{a}_{\rm gwb}$. This variance is proportional to the power spectral density of the GWB-induced time delays:

$$\begin{aligned}\boldsymbol{\varphi}_{\rm gwb} &= \langle \mathbf{a}_{\rm gwb}\mathbf{a}^*_{\rm gwb}\rangle \\ &= \Gamma_{ab} \times \frac{A^2_{\rm gwb}}{12\pi^2}\frac{1}{T}\left(\frac{\nu_i}{1{\rm yr}^{-1}}\right)^{-\gamma} {\rm yr}^2. \end{aligned} \tag{14}$$

In the following two techniques we destroy signal covariance by either operating on the phase coherence through $\mathbf{F}$, or on the induced spatial correlations through Γ_{ab}. But in Eq. (13) we see that these are naturally linked through the common result of mitigating cross-pulsar signal correlations in the data.

We now explore the two techniques displayed in Fig. 1 in detail.

A. Phase shifting

Phase shifting attacks the phase coherence of the GWB signal between different pulsars in our array. In fact, phase shifting will remove *any* spatial correlations in the data, regardless of whether they are due to a GWB or other influences (e.g. clock errors, Solar System ephemeris errors). There are two approaches one can take in phase shifting — we can either construct phase shifted datasets (*data-driven*), or we can search for the GWB with a phase shifted model (*model-driven*).

Data-driven — In this approach we must first reconstruct the signal in each individual pulsar. We do this by analyzing each pulsar in turn, determining the maximum likelihood parameters for the intrinsic pulsar noise obtained without including a GWB in the model, then solving Eq. (11) to obtain the maximum likelihood signal coefficients. By taking the subset of the vector $\widehat{\mathbf{b}}$ that corresponds only to the low-frequency process coefficients, $\widehat{\mathbf{a}}$, for those frequencies we wish to shift, we can then reconstruct the maximum likelihood signal realization $\widehat{\mathbf{s}}$ for those frequencies via:

$$\mathbf{s} = \mathbf{F}\widehat{\mathbf{a}}. \tag{15}$$

We also construct a shifted signal, which we will denote $\mathbf{s}'$, using the adjusted matrix $\mathbf{F}'$, defined as:

$$F'(\nu, t) = \sin\left(2\pi\nu t + \delta\right), \tag{16}$$

and equivalent cosine terms, with δ a random phase between 0 and 2π. This gives our shifted signal as:

$$\mathbf{s}' = \mathbf{F}'\widehat{\mathbf{a}}. \tag{17}$$

We can then construct a new, shifted dataset $\delta\mathbf{t}'$:

$$\delta\mathbf{t}' = \delta\mathbf{t} - \mathbf{s} + \mathbf{s}'. \tag{18}$$

An example of this shifting process is shown in Fig. 2. Crucially (as will be shown in Sec. IV B) this process retains the statistical properties of the original dataset, including any unmodeled stochastic or systematic effects. However by shifting the phases of the signal we are able to destroy the correlations that exist between pulsars.

Model-driven — This approach is rather simple, in that we employ the pulsar datasets as they are, and search on the data with new phase shifted "Fourier design matrices", $\mathbf{F}'$, in the model for low-frequency processes. This scrambles our model for the GWB by destroying phase coherence between pulsars, such that the data will prefer models where the low frequency processes are explained by either separate intrinsic spin-noise or a common uncorrelated process. Because of the simplicity with which the *model-driven* phase shifting can be

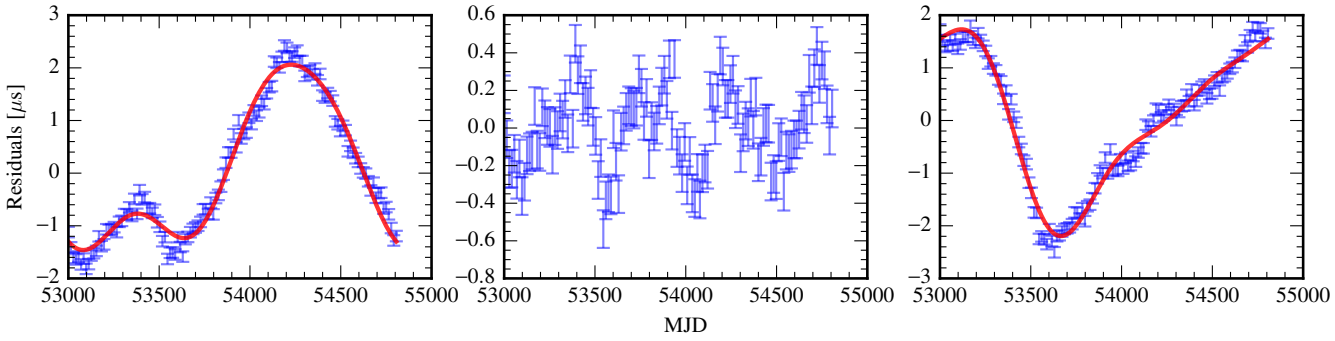


FIG. 2: Example of the phase shifting process for one pulsar in IPTA MDC open 1 (blue points, left plot), described in Sec. IV A. In this example we include only the lowest 3 frequencies in the set to be shifted, and obtain the maximum likelihood signal realization for this set (red line, left plot). We subtract this signal from the dataset to obtain a set of residuals that contain everything not modelled by this set of frequencies (middle plot). In this case there is still clear structure left over. We then perform the phase shift described in Section III A to obtain the new signal (red line, right plot), which we add back into the residuals to obtain a new, shifted dataset (blue points, right plot). The statistical properties of the dataset remain unchanged, while phase coherence with other pulsars has been destroyed for the shifted frequencies.

implemented, we use it in all of the following, but note that both approaches produce consistent results, and the *data-driven* approach is important in illustrating that the statistical properties of each pulsar dataset remain unaffected.

B. Sky scrambles

"Sky scrambles" attack the *spatial* correlations induced by the GWB signal between different pulsars in the array. This is based on the fact that the expected spatial correlation induced by an isotropic stochastic GWB is distinctive, being representative of the quadrupolar antenna response of pulsars to transverse-tensor mode GWs, and unlike any other possible conceived spatial correlations. As mentioned before, this is commonly referred to as the "Hellings and Downs correlation". A common uncorrelated low-frequency signal will have zero spatial correlations; stochastic clock-standard drifts will have the same spatial correlations between all pulsars (monopole correlation); and inaccuracies in the solar system ephemerides may lead to a dipole spatial correlation between pulsars. Unlike phase shifting, sky scrambling requires us to make specific assumptions about the expected spatial correlation signature of the signal. However, as in phase shifting, there are *model-driven* and *data-driven* approaches.

Model-driven — When constructing the spatial correlations of the GWB signal model in our search pipelines, we artificially move pulsar positions from their true values (their true values are retained for fitting astrometric terms in the timing model). This means that the angular separations between pulsars will also be scrambled. Thus when we impose the "Hellings and Downs" correlation between pulsars for searches, this spatial correlation will be at odds with the spatial correlations of any true signal in the data. Our goal is to make the overlap of the true spatial correlations as orthogonal as possible to our scrambled correlation model. As such we effectively null the influence of GWB-induced spatial correlations in the data.

An intuitive picture arises by considering the behavior of the signal-to-noise ratio from the stochastic background "optimal statistic" [2, 6, 37] in the frequency domain:

$$\langle\rho\rangle^2 \propto \sum_{a,b\neq a}\sum_i \left[\frac{\Gamma'_{ab}(\nu_i)S'_i \times \Gamma_{ab}(\nu_i)S_i}{P_a(\nu_i)P_b(\nu_i)}\right], \qquad (19)$$

where a and b index pulsars, i indexes sampling frequencies of the pulsar time series, primes denote template (or model) quantities, and lack of primes indicate true signal or noise quantities. In principle we can apply different template ORFs at each frequency, however for the purposes of this study we assume no frequency evolution of the GWB angular-power distribution, and thus no evolution of the true or template ORF. Likewise, the power spectral density of the GWB-induced time delays, S_i, takes the usual power-law functional form throughout.

The influence of a sky scramble in Eq. (19) would be to minimize the overlap of the template with the signal, and thus diminish the detection significance of spatial correlations in the data. This equation can also be used as a generator of sky scrambles, where we insert typical pulsar noise properties along with signal assumptions to find the scrambled pulsar positions which lead to a scrambled template ORF which makes $\langle\rho\rangle$ as small as

possible. The scrambles found in this way can then be used in constructing the spatial correlation model in a Bayesian analysis of real datasets. Repeating for many sky scramble analyses of the real dataset allows a distribution of Bayes factors to be obtained under the null hypothesis where spatial correlations have been destroyed, thus putting the significance of the actual signal-model Bayes factor in context.

In practice, a more straightforward generator of sky scrambles is through minimization of the normalized overlap of the template ORF with the true "Hellings and Downs" ORF. This "match statistic" has been explored in [9], and we reiterate its form here:

$$M = \frac{\sum_{a,b} \Gamma_{ab}\Gamma'_{ab}}{(\sum_{a,b} \Gamma_{ab}\Gamma_{ab} \times \sum_{a,b} \Gamma'_{ab}\Gamma'_{ab})^{1/2}},$$
$$\overline{M} = \frac{\sum_{a,b\neq a} \Gamma_{ab}\Gamma'_{ab}}{(\sum_{a,b\neq a} \Gamma_{ab}\Gamma_{ab} \times \sum_{a,b\neq a} \Gamma'_{ab}\Gamma'_{ab})^{1/2}}, \tag{20}$$

where in M the sum is over all unique pulsar pairings, while in $\overline{M}$ the sum excludes pulsar self-pairings since these merely add positive terms regardless of whether the pulsar positions are scrambled or not. The benefits of these match statistics is that they rely purely on the geometric properties of the array through the sky locations of the pulsars. We use the minimization of $\overline{M}$ to generate sky scrambles for the analyses in the rest of this paper, employing a particle swarm optimization (PSO) algorithm [23, 36] to find scrambled positions for which $\overline{M}$ is below a given threshold with respect to the true ORF *and* all other previously discovered sky-scrambles. In the first simulation with 36 pulsars (discussed in greater detail in Sec. IV) we found that the PSO algorithm was able to discover at least several hundred sky-scrambles with a threshold of $\lesssim 0.2$, while with only 10 pulsars in the second simulation we had to raise the threshold to $\lesssim 0.5$ in order to find an adequate number of scrambles. Nevertheless, in Sec. IV we show that the distribution of log-Bayes factors under the null hypothesis which were found with these scrambles match the phase-shifting and noise-only distributions very well.

There is a potential concern that, for a given number of pulsars, there are only a finite number of unique sky scrambles which produce ORFs that are orthogonal to the true ORF and all other scrambled ORFs. In the course of performing many sky-scramble analyses, there may be repetition of these scrambled ORFs which one might worry would lead to biases in assessing detection significance. The problem is akin to trying to produce the posterior distributions of sampled parameters given an MCMC chain with a large autocorrelation length. In practice we do not believe this to be problematic since (much like in our MCMC analogy) provided we sample a sufficiently large number of sky scrambles we will overcome any local clumpiness due to repetition of non-unique scrambles. The histogram of Bayes factors under the null hypothesis should then be unbiased, as will our assessment of the significance of the true dataset's Bayes factor.

Data-driven — This approach proceeds in much the same way as *data-driven* phase shifting. We perform an initial search on the *entire* pulsar array dataset for a GWB signal, from which we can extract the maximum likelihood signal coefficients $\widehat{\mathbf{a}}_{\rm gwb}$ in each pulsar. At each frequency, the variance of these signal coefficients is equal to the power spectral density of the GWB-induced time delays, scaled by the ORF between the pair of pulsars in question: $\langle a_a a_b^* \rangle_i = \Gamma_{ab} S_i$. In the following we denote the vector of all pulsars' GWB signal coefficients at a particular frequency, ν_i, by $\widehat{\mathbf{a}}_i$. The expected covariance matrix of this vector of coefficients is then the "Hellings and Downs" spatial correlation matrix, scaled by S_i. Explicitly, the spatial correlation matrix, $\mathbf{\Gamma}$ is:

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \vdots \\ \Gamma_{21} & \Gamma_{22} & \vdots \\ \cdots & \cdots & \ddots \end{bmatrix}. \tag{21}$$

We Cholesky factorize this spatial correlation matrix, followed by operating on $\widehat{\boldsymbol{a}}_i$ with the inverse Cholesky factor to decorrelate the signal between different pulsars. Hence,

$$\mathbf{\Gamma} = \mathbf{L}\mathbf{L}^T,$$
$$\widehat{\mathbf{a}}'_i = \mathbf{L}^{-1}\widehat{\mathbf{a}}_i, \tag{22}$$

where $\widehat{\mathbf{a}}'_i$ is the vector of new pulsar signal coefficients at frequency i, which are uncorrelated between pulsars but retain the same spectral properties. We repeat this process at all sampling frequencies in our rank-reduced approximation of the GWB signal, giving new vectors of signal coefficients for each pulsar. As in the *data-driven* phase shifting approach, we now form new pulsar datasets such that:

$$\delta\mathbf{t}' = \delta\mathbf{t} - \mathbf{F}\widehat{\mathbf{a}} + \mathbf{F}\widehat{\mathbf{a}}'. \tag{23}$$

We now have new pulsar datasets which have their spectral properties intact, but which are decorrelated with respect to each other according to our model of what the GWB would induce. The analog with Eq. (18) is now easy to see: when phase shifting we modify $\mathbf{F}$ while in sky scrambling we modify $\widehat{\mathbf{a}}$. As in the case of phase shifting, the *model-driven* sky-scrambling approach is a more straightforward practical implementation, so we use it in all of the following.

C. Unified formalism

Let us examine the combined influence of phase shifting and sky scrambling on the timing-residual correlation

between two pulsars. For simplicity we consider only the correlation due to the GWB. We also initially consider only one sampling frequency in the reduced rank description of the signal, but generalize later. The covariance between timing residuals at t_k in pulsar a and at t_l in pulsar b is:

$$\mathbf{C}_{(ak),(bl)} = \mathbf{F}_{(ak)} \boldsymbol{\varphi}_{ab} \mathbf{F}^T_{(bl)}, \tag{24}$$

where, with only one sampled frequency at ν_i, $\mathbf{F}_{(ak)}$ takes the form:

$$\mathbf{F}_{(ak)} = \left[\sin(2\pi\nu_i t_k) \;\; \cos(2\pi\nu_i t_k)\right], \tag{25}$$

and $\mathbf{F}_{(bl)}$ is likewise. The spectrum, φ_{ab}, at each frequency is as in Eq. (14), but is explicitly represented here as a $2N_{\rm freqs} \times 2N_{\rm freqs}$ matrix (since each frequency has a sine and cosine basis function):

$$\boldsymbol{\varphi}_{ab} = \begin{bmatrix} \varphi_{ab} & 0 \\ 0 & \varphi_{ab} \end{bmatrix}. \tag{26}$$

The result of phase shifting and sky scrambling is to convert $\mathbf{F}_{(ak)}$ and φ_{ab} into the following:

$$\begin{aligned} \mathbf{F}'_{(ak)} &= \left[\sin(2\pi\nu_i t_k + \delta_{ai}) \;\; \cos(2\pi\nu_i t_k + \delta_{ai})\right], \\ \varphi'_{ab} &= \frac{\Gamma'_{ab}}{\Gamma_{ab}} \times \varphi_{ab}. \end{aligned} \tag{27}$$

If we explicitly evaluate Eq. (24) with the phase shifted and scrambled quantities (and generalize to multiple sampling frequencies) we get the following for our scrambled model of the induced correlations:

$$\begin{aligned} \mathbf{C}'_{(ak),(bl)} = \sum_i^{N_{\rm freqs}} \frac{\Gamma'_{ab}}{\Gamma_{ab}} \varphi_{ab} \big[&\cos(2\pi\nu_i(t_k - t_l)) \cos(\delta_{ai} - \delta_{bi}) \\ &- \sin(2\pi\nu_i(t_k - t_l)) \sin(\delta_{ai} - \delta_{bi})\big]. \end{aligned} \tag{28}$$

One can easily see that without phase shifting (or with a common phase shift for all pulsars at each frequency) and without sky scrambling, the correlation is:

$$\mathbf{C}'_{(ak),(bl)} = \sum_i^{N_{\rm freqs}} \varphi_{ab} \cos(2\pi\nu_i(t_k - t_l)), \tag{29}$$

which is the just the discrete Wiener-Khinchin conversion between the power spectral density of a process and the time-domain correlation.

Also, the autocovariance of each pulsar can be examined by setting $a = b$ in Eq. (28). In this case $\Gamma_{aa} = \Gamma'_{aa} = 1$, and phase shifts cancel at each sampling frequency, such that the statistical properties of each individual pulsar dataset remain intact.

IV. APPLICATION TO SIMULATIONS

In the following we test our two techniques against several different types of PTA datasets. All Bayesian analysis and evidence recovery is performed using the PTA analysis suite NX01 [38] with the MultiNest sampler [4, 11–13], where 1000 live points are employed in Secs. IV A and IV C (for a 2-D parameter space), and 5000 live points are employed Sec. IV B (for a 22-D parameter space).

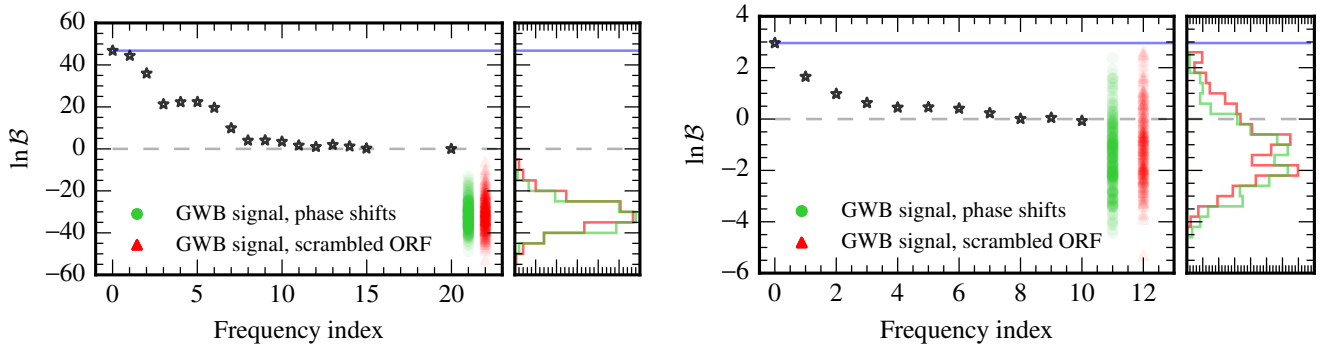
A. Simulation 1 (IPTA MDC open 1)

We first apply the techniques described in Sec. III A and Sec. III B to the first open dataset of the publicly available, International Pulsar Timing Array [43] mock data challenge.[1] This dataset contains 5 years of observations for a set of 36 pulsars, each with a 14 day cadence, and each with uncorrelated TOA measurement uncertainties of 10^{-7} seconds. These datasets do not include any intrinsic red noise processes, but include an additional power-law GWB term with spectral index, $\gamma_{\rm gwb} = 13/3$, and amplitude $A_{\rm gwb} = 5 \times 10^{-14}$. While this does not represent a realistic dataset by any metric, it is a simple initial test case upon which to explore the effectiveness of our approaches for eliminating the correlation between pulsars due to the GWB. In the following analyses, our noise model includes only TOA uncertainties given by the observations (i.e. no search parameters) while the signal model is a power-law GWB (two search parameters).

To apply phase shifts in our GWB model, we first need to assess how many sampling frequencies contain information about correlations between the pulsars. We can investigate this by analyzing the dataset with a model which neglects spatial correlations in a successive number of sampling frequencies, beginning with the base frequency $1/T$ and then increasing. As we see in Fig. 3a, the evidence for a GWB without spatial correlations in the lowest ~ 20 frequencies is indistinguishable from a common uncorrelated red process. As such, so long as we apply random phase shifts to at least the first 20 sampling frequencies in our model, the phase coherence between pulsars will be destroyed. Sky scrambles are generated using the match-statistic minimization approach described in Sec. III B.

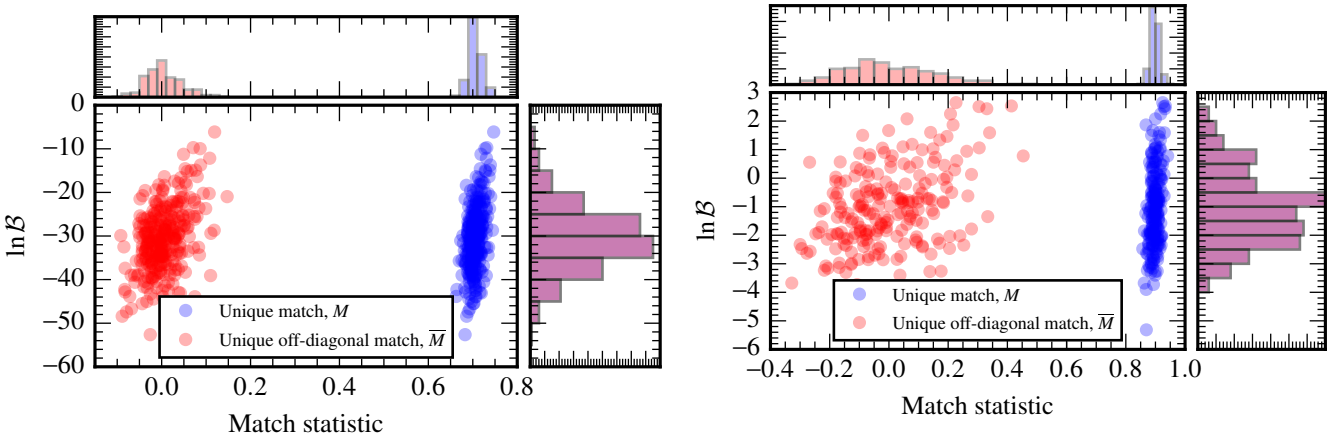
The result of carrying out 300 phase-shifting and 300 sky-scrambling analyses of the dataset are also shown in Fig. 3a as green and red log-Bayes factor histograms, respectively. The Bayes factor is for a "Hellings and Downs" GWB model versus a common uncorrelated red process. Therefore the actual values correspond to the

[1] http://www.ipta4gw.org/?page_id=214



(a) Simulation 1 (b) Simulation 2

FIG. 3: Logrithm of the Bayes factor, $\mathcal{B}$, for a correlated GWB model versus a common uncorrelated red process. Black stars correspond to operating on the unmodified dataset without modeling spatial correlations in the lowest N coefficients, e.g. the black star at a frequency index of 1 signifies that the correlation has not been included in the lowest frequency. The green circles and histogram correspond to operating on the unmodified dataset with 300 phase-shift instances, while the red triangles and histogram are for 300 sky-scramble instances. The case of equal model evidences is indicated with a dashed grey line at zero, and the true Bayes factor of spatial correlations in the dataset is indicated with a solid blue line.

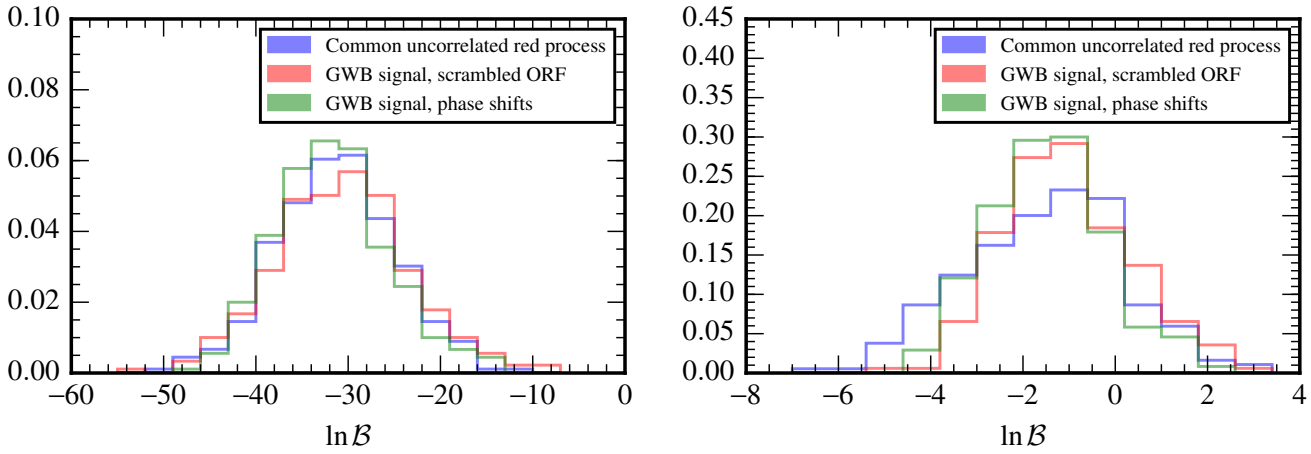


(a) Simulation 1 (b) Simulation 2

FIG. 4: Distribution of scrambled-sky Bayes factors and match-statistics upon application to Simulation 1 (IPTA MDC Open 1) and Simulation 2 (more realistic dataset).

difference between the ln $\mathcal{Z}$ of a GWB signal in the shifting/scrambling analyses and the ln $\mathcal{Z}$ of a common uncorrelated red process (this latter log-evidence value only needs to be computed once). The fact that these histograms are centred around ~ -30 shows that the data now strongly favor a common uncorrelated red process under the modified models, and both the phase-shifting and sky-scrambling techniques produce consistent histograms. The result of these techniques is to produce a distribution of the log-Bayes factor under the null hypothesis i.e. where there are no correlations between pulsars. The true log-Bayes factor of $\sim +47$ is seen to be highly significant in the context of these null hypothesis distributions. To make quantitative statements about the p-value of correlations one would need to perform many more analyses than our 300 to fill out the tails of the distributions, however it is clear that spurious noise correlations are not capable of producing the very high Bayes factor given by the data.

We realize that to speak of p-values resulting from Bayesian detection statistics is to mix two separate inference philosophies, however we are merely trying to contextualize the significance of our computed Bayes factors by seeing how often spurious correlations could be forged by noise alone. Standard rule-of-thumb guides for the significance of Bayes factors [20, 21] are unsatisfactory because they do not account for the specifics of a given problem.



(a) Simulation 1

(b) Simulation 2

FIG. 5: Comparison of the log-Bayes factor histograms obtained via phase-shifting and sky-scrambling the original dataset, and the histogram obtained by simulating many noise-only datasets having the same statistical properties as the original dataset but without spatial correlations.

In Figure 4a we show the distribution of log-Bayes factors for the 300 sky scrambles along with the corresponding distribution of match statistics M and $\overline{M}$, which were introduced in Eq. (20). We see that $\overline{M}$ is tightly distributed around zero by construction, while the diagonal self-pairing terms of the ORFs merely add a positive offset to M which renders it less useful than $\overline{M}$ as a geometric generator of sky scrambles.

Finally, we must check how phase-shifting and sky-scrambling compare to noise-only simulations as a means of determining the distribution of the Bayes factor under the null hypothesis. In Fig. 5a we compare our histograms of log-Bayes factors from phase-shifting and sky-scrambling with a histogram from analyzing 300 independent noise-only datasets, where each pulsar has the same statistical properties as in the original dataset (i.e. 100 ns of white noise, and a power-law red noise process with $A = 5 \times 10^{-14}$ and $\gamma = 13/3$) but without spatial correlations. Two analyses are needed per simulated dataset: a GWB signal analysis, and a common uncorrelated red process analysis, where the log-Bayes factor is given by the difference in the log-evidence value of each model. One can see that our techniques match the performance of noise-only simulations very well.

B. Simulation 2 (a more realistic simulation)

We now apply the techniques to a more realistic 10 pulsar dataset. Table I lists the timespan, rms of the white noise, and the properties of the red noise for each pulsar in the simulation. These values are chosen to be similar to those given in Caballero *et al.* [5] in order to provide as realistic a simulation as possible. In addition we add a correlated power-law GWB signal with spectral index, $\gamma_{\rm gwb} = 13/3$, and amplitude $A_{\rm gwb} = 5 \times 10^{-15}$. While this is significantly in excess of current upper limits, we choose this amplitude such that the change in log evidence between models that do or do not include spatial correlations is $\sim +3$. Therefore the phase-shifting and sky-scrambling operations can produce a measurable change in the evidence.

TABLE I: Details of simulation 2.

Pulsar	$T_{\rm obs}$ [years]	σ_w [μs]	$\log_{10} A_{\rm red}$	$\gamma_{\rm red}$
J0613−0200	16.054	1.58	-13.90	3.18
J0751+1807	17.606	2.60	-14.14	2.58
J1012+5307	16.831	1.47	-13.09	1.65
J1640+2224	16.735	1.99	-13.24	0.03
J1643−1224	17.300	1.65	-18.56	4.04
J1713+0747	17.657	0.26	-14.90	4.85
J1744−1134	17.250	0.65	-13.60	2.00
J1857+0943	17.310	1.51	-16.00	1.35
J1909−3744	9.379	0.12	-13.99	2.06
J2145−0750	17.161	1.19	-13.87	4.02

As in the previous section, we first investigate how many frequencies are informative of spatial correlations in the data, where we see in Fig. 3b that the majority of information is contained in the lowest $\sim 2-3$ frequencies. The evidence is reduced to that of a common uncorrelated red process after neglecting spatial correlations in the lowest $\sim$ 10 frequencies. We therefore apply phase shifts to at least these lowest 10 frequencies in the GWB signal model. As before, sky scrambles are generated by minimizing the unique off-diagonal match statistic, $\overline{M}$.

Figure 3b also shows the histograms of log-Bayes factors produced from several hundred phase-shifting and sky-scrambling experiments, where the techniques are

shown to match very well. Although not as significant as the signal in IPTA MDC open 1, we see that the Bayes factor for spatial correlations in this more realistic simulation is still highly convincing and unlikely to have been formed via spurious noise correlations. As discussed previously, in a real analysis we would desire a quantitative assessment of the significance; this would require many more phase shifting or sky scrambling experiments than are examined here in order to produce smooth distributions which are well sampled in the tails. In Fig. 4b we see that, despite the larger $\overline{M}$ of the generated sky scrambles, the scrambling process produces an adequate distribution of the Bayes factor under the null hypothesis, and in fact (as we have seen) is well matched to the phase-shifting distribution.

As before, we confirm that phase-shifting and sky-scrambling produces Bayes factor distributions under the null hypothesis which compare well with the distribution produced from analyzing many noise-only simulations. The results for this are shown in Fig. 5b, where all distributions are shown to be in good agreement.

Finally, we use this more realistic dataset to demonstrate that phase-shifting and sky-scrambling do not alter the statistical properties of each individual pulsar dataset. Figure 6 shows the posterior distributions of the intrinsic red noise parameters for a subset of the pulsars in simulation 2. Red lines show the mean parameter estimates over 50 realizations of simultaneous phase-shifting and sky-scrambling on the original dataset, where the model includes separate red noise per pulsar and a GWB. Blue lines show parameter estimates from an analysis of the unmodified dataset, where the model includes separate red noise per pulsar and a common uncorrelated red process. As desired, the shifting and scrambling processes have not significantly affected the parameter estimates for individual pulsars, indicating that the underlying statistics of the dataset remain consistent whether we analyze with a shifted/scrambled GWB model or a common uncorrelated red process.

C. Simulation 3 (a mismodelled simulation)

So far we have demonstrated that phase-shifting and sky-scrambling perform just as well as analyzing many noise-only simulations. However, if this equality of performance were all that we could claim then the techniques would certainly be useful, but not strictly necessary. In this section we show that phase-shifting and sky-scrambling on the real dataset can in fact be superior to analyzing many simulations when our understanding of noise processes in the dataset is incomplete.

We generate an extreme example of a dataset with a large unmodelled noise component by injecting loud independent glitches (negative ramps in the residual time-series due to spontaneous increases in the pulsar rotational frequency) into each pulsar of the IPTA MDC open 1 dataset (simulation 1, described in Sec. IV A). This glitch term is injected as:

$$s_{\rm glitch}(t) = -\mathcal{A} \times (t - t_e) H(t - t_e) \times {\rm spd}, \qquad (30)$$

where t is the MJD of a given pulsar observation, $H(\cdot)$ is the Heaviside step function, $\mathcal{A}$ is the glitch amplitude, t_e is the MJD of the glitch epoch, and spd = 86400 is the number of seconds per day. The glitch epochs are randomly drawn from a uniform distribution between the earliest (MJD 53000) and latest (MJD 54806) observation times of the PTA, while the $\log_{10}\mathcal{A}$ values are randomly drawn from a uniform distribution in the range $[-18, -17]$. The glitch parameters injected into each pulsar are shown in Table II.

We analyze this simulation as we would do for a real dataset, with the following checklist:

1. Analyze the full PTA dataset for a GWB signal.

2. Analyze the full PTA dataset for a common uncorrelated red process.

3. Generate many simulated datasets with the maximum-likelihood pulsar noise properties of the true dataset, and with a common uncorrelated red process having the maximum-likelihood parameters of the recovered GWB signal in the true dataset. Analyze each dataset for a GWB signal and for a common uncorrelated red process.

4. Analyze the true dataset for the number of frequencies which are informative of spatial correlations. Perform many phase-shift analyses on the true dataset.

5. Generate scrambled sky positions from either an optimal statistic analysis or the match statistics. Perform many sky-scramble analyses on the true dataset.

6. Item (3) gives the distribution of the Bayes factor under the null hypothesis from simulations. Items (4) + (2) are used to give the phase-shifting estimation of the null hypothesis Bayes factor distribution. Items (5) + (2) are used to give the sky-scrambling estimation of the null hypothesis Bayes factor distribution.

When performing GWB analyses on the true dataset, our model consists of white-noise given by the reported TOA measurement uncertainties, and a GWB signal with "Hellings and Downs" spatial correlations. We purposefully do not try to model the glitch in each pulsar so that we can assess how phase-shifting and sky-scrambling perform when we have an incomplete noise model. The parameter estimation of the GWB spectrum will be dominated by the self-pairings of the pulsars (since the off-diagonal elements of the ORF matrix are at most half of the diagonal terms). Therefore the

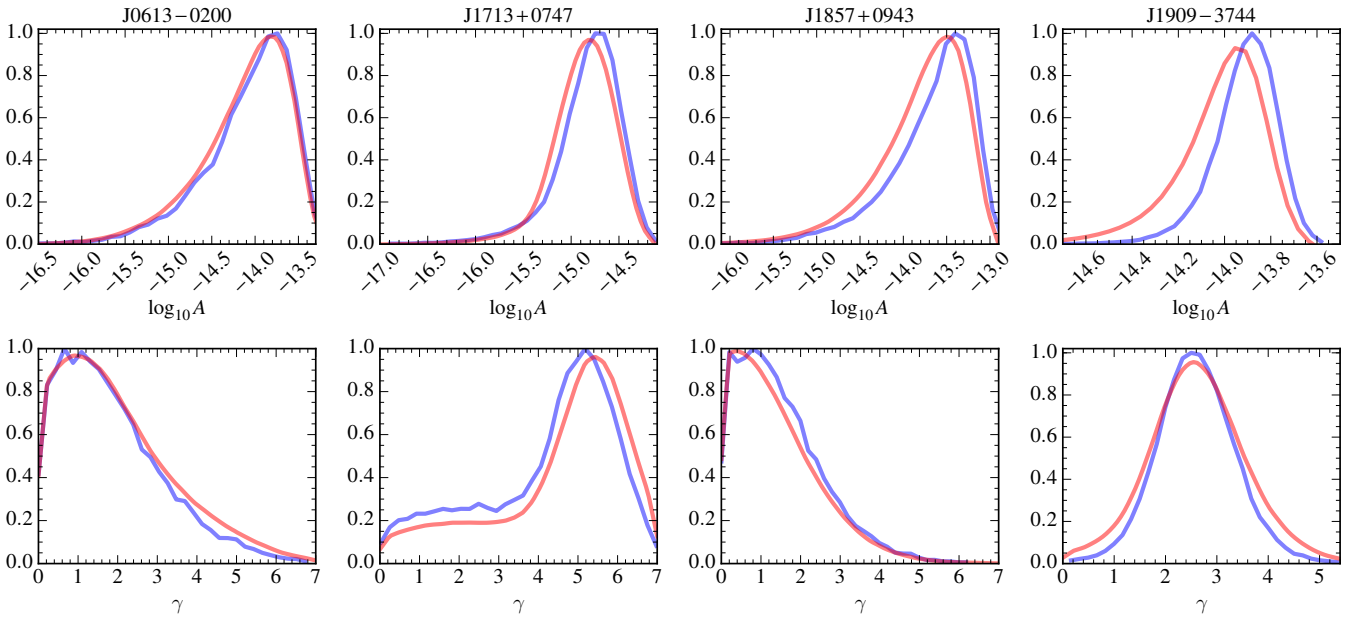


FIG. 6: One dimensional marginalized posterior distributions for the log-amplitude (top) and spectral index (bottom) of the intrinsic red noise for four of the pulsars in simulation 2. Blue lines represent the analysis for the unmodified dataset containing the uncorrelated common red noise term, red lines are the mean of the posterior distributions over 50 realizations of the combined phase and position shifted datasets. We find that the parameter estimates before and after performing the shifting process are consistent with one another.

TABLE II: Parameters of glitch term added to each pulsar in IPTA MDC open 1 to generate simulation 3. The logarithm of the amplitude is drawn such that $\log_{10}\mathcal{A} \in U[-18,-17]$, while the glitch epoch is drawn such that $t_e \in U[\text{MJD } 53000, \text{MJD } 54806]$.

Pulsar	$\log_{10}\mathcal{A}$	t_e [MJD]	Pulsar	$\log_{10}\mathcal{A}$	t_e [MJD]	Pulsar	$\log_{10}\mathcal{A}$	t_e [MJD]
J0030+0451	-17.00	53204	J1455−3330	-17.78	54376	J1853+1303	-17.27	54235
J0218+4232	-17.42	53601	J1600−3053	-17.78	53240	J1857+0943	-17.26	54188
J0437−4715	-17.05	53276	J1603−7202	-17.66	53012	J1909−3744	-17.92	53190
J0613−0200	-17.76	53501	J1640+2224	-17.62	53326	J1910+1256	-17.69	53803
J0621+1002	-17.96	54148	J1643−1224	-17.66	53179	J1918−0642	-17.77	54359
J0711−6830	-17.09	53816	J1713+0747	-17.80	54132	J1939+2134	-17.08	54252
J0751+1807	-17.05	53637	J1730−2304	-17.40	53560	J1955+2908	-17.38	54680
J0900−3144	-17.25	53798	J1732−5049	-17.26	53682	J2019+2425	-17.61	53864
J1012+5307	-17.85	53620	J1738+0333	-17.19	53119	J2124−3358	-17.34	53259
J1022+1001	-17.27	53257	J1741+1351	-17.83	54620	J2129−5721	-17.78	53122
J1024−0719	-17.85	54022	J1744−1134	-17.96	53526	J2145−0750	-17.09	53626
J1045−4509	-17.61	53099	J1751−2857	-17.93	54800	J2317+1439	-17.59	53167

glitches, which are rather loud negative ramps, are interpreted as an additional low-frequency component of the GWB such that the recovered signal will have a higher amplitude than the true signal. This is indeed the case, where the maximum-likelihood GWB signal parameters are found to be $A_{\rm gwb} = 9.60 \times 10^{-14}$ and $\gamma_{\rm gwb} = 4.24$, and the posterior distributions are inconsistent with the true GWB signal parameters of $A_{\rm gwb} = 5 \times 10^{-14}$ and $\gamma_{\rm gwb} = 4.33$. The log-Bayes factor for spatial correlations in this dataset is $\sim +4$.

We create 300 PTA dataset simulations which contain white-noise at the level of the reported TOA uncertainties, and a red process in each pulsar with $A = 9.60 \times 10^{-14}$ and $\gamma = 4.24$ which is uncorrelated across pulsars. Since this amplitude is larger than what is actually in the true dataset, the significance of the GWB signal should be biased high. The histogram of log-Bayes factors from these noise simulations is contrasted with the phase-shifting histogram and sky-scrambling histogram in Figure 7. In Fig. 7a we see that phase-shifting gives a more conservative estimate of the significance of the GWB spatial correlations in the dataset than noise simulations, while in Fig. 7b we see that sky-scrambling implies a significance somewhere between the

aforementioned two.

Phase-shifting and sky-scrambling preserve all properties of the original dataset (including e.g. systematic offsets, clock-standard drifts, calibration errors), rather than producing simulations with properties which match the mismodelled noise processes in the true dataset. Phase shifting is the more reliable and general approach, since it makes no assumptions about the nature of spatial correlations in the data, and simply eliminates phase coherence between pulsars. Therefore *any* spatial correlations in the data will be eliminated by phase shifting. On the other hand sky scrambling requires us to make a specific assumption about the nature of spatial correlations in the data. For example, we assume that the spatial correlations in the data are only due to an isotropic GWB, which is a necessary assumption in order to generate the scrambled positions.

The differences in the approaches can be seen most clearly by looking at the *data-driven* procedures. When phase shifting we analyze each pulsar individually, applying random phase shifts at all sampling frequencies of the low-frequency processes to produce new pulsar datasets with the same statistical properties (see Eq. (16)). At no point do we make any specific assumption about possible spatial correlations in the data. When sky scrambling we perform a full PTA search for a GWB, where we assume a "Hellings and Downs" correlated model. This model may soak up more than just any real GWB which is present, including unmodeled noise features. Thus the recovered signal coefficients, which we attempt to decorrelate between pulsars (see Eq. (22)), may contain a component which was never correlated in the first place. As such, sky scrambling may exaggerate the significance of spatial correlations when we have an incomplete noise model, in qualitatively the same way as noise simulations. Phase shifting should provide an estimate of the true Bayes factor significance which is similar to or more conservative than noise simulations and sky scrambling.

V. CONCLUSIONS

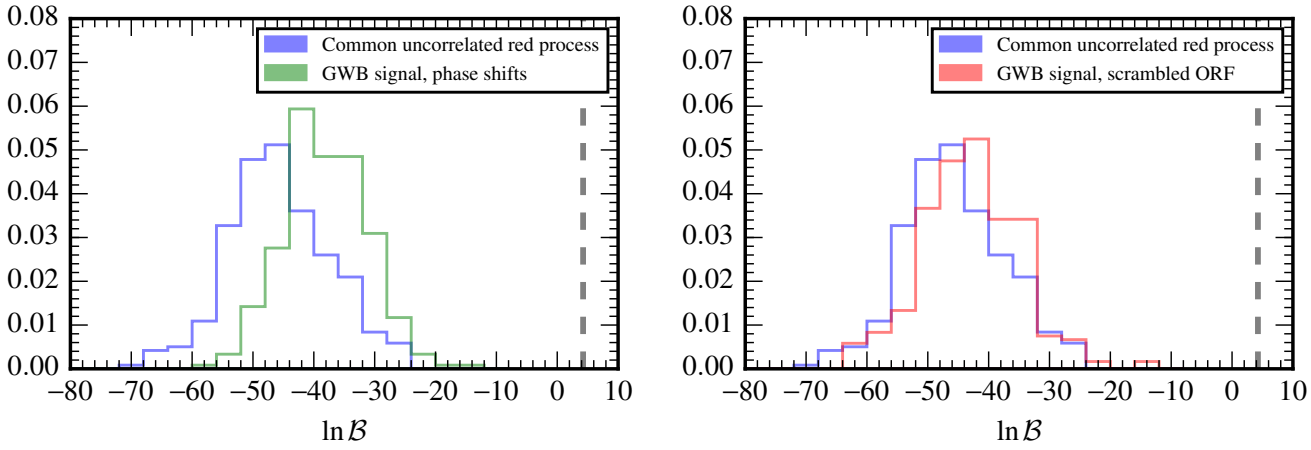
We have studied two methods of determining the significance of spatially correlated signals in a pulsar timing array within a Bayesian context. Assessing how often spurious noise correlations can lead to Bayes factors as large as what we find upon analyzing our given PTA dataset is crucial for placing this Bayes factor in context. If noise alone can often produce Bayes factors as large as what we find in the PTA dataset then the value of this detection statistic is clearly not very significant. Although standard rule-of-thumb guides exist for assessing Bayes factor significance [20, 21], using them in any production-level analysis is deeply unsatisfactory because they are not designed with the specifics of a given problem in mind. Rather, we must resort to numerical experiments to produce distributions of Bayes factors for spatial correlations under the null hypothesis i.e. where there are no spatial correlations in the dataset.

Our first technique involved adding random phase shifts to all basis functions modeling the low-frequency processes in a given pulsar dataset. This is performed separately for each pulsar, such that the statistical properties of each pulsar dataset remain intact, but all phase coherence between pulsars due to a potential GWB, Solar-system ephemeris errors, or clock-standard drifts are eliminated. The second technique involves scrambling the pulsar positions used to model the expected overlap reduction function (ORF) of the spatially-correlated signal, in such a way that the modeled ORF is effectively orthogonal to the true signal's ORF, and thus correlations between pulsars are destroyed. Both of these techniques operate on the actual PTA dataset rather than analyzing noise-only simulations, which has the benefit of incorporating all idiosyncrasies of the true dataset into measures of the detection statistic significance instead of being biased by our (possibly incomplete) noise model assumptions.

We tested our techniques against several different types of PTA datasets, including an idealized dataset (with a large number of evenly sampled pulsars, high timing precision, and no intrinsic red noise), a more realistic dataset (realistic cadence, timing precision, and red noise levels), and a dataset for which our noise model is incomplete (i.e. includes a process which we do not explicitly model). By performing several hundred phase-shifting and sky-scrambling analyses, we constructed a distribution of the Bayes factor for spatial correlations under the null hypothesis, which in practice would allow us to quote the p-value of the true Bayes factor. Quoting p-values when our detection statistic is Bayesian may seem like an ill-conceived mixture of two distinct inference philosophies, but it is merely trying to answer the question of what our recovered Bayes factor actually means, and how often noise alone could produce it.

For both the idealized and realistic datasets, we found that the distribution of Bayes factors produced by phase-shifting and sky-scrambling compared well with that of analyzing many noise-only simulations. However, we took a further step in showing that our two techniques are actually superior to noise-only simulations in the case where our noise model is an incomplete description of processes in the data. Noise-only simulations will include only the processes of our incomplete noise model, and so will provide a distribution of Bayes factors which exaggerate the significance of the true Bayes factor. Phase shifting operates on the actual PTA dataset and provides a more conservative estimate of the detection significance, while sky scrambling is found to give a significance somewhere between noise-only and phase-shifting results (but closer to noise-only simulations).

These techniques can be readily deployed on all existing and future PTA datasets, and should become standard tools for contextualizing the Bayes factors for spatial correlations that we report in future PTA analyses. Understanding the significance of our quoted detection



(a) (b)

FIG. 7: The histogram of log-Bayes factors for spatial correlations from 300 noise-only simulations is compared to 300 (*a*) phase-shifting, and (*b*) sky-scrambling analyses of the true simulation 3 dataset. The log-Bayes factor found from analyzing the true dataset is shown as a vertical dashed grey line in both panels. See the text for more details and discussion.

statistics is of vital importance as pulsar-timing moves closer to the first detection of nanohertz gravitational waves.

ACKNOWLEDGMENTS

We wish to thank Justin Ellis, Michele Vallisneri, Rutger van Haasteren, Joseph Lazio, Neil Cornish, Laura Sampson, as well as the entire NANOGrav and EPTA detection working groups for many useful suggestions and fruitful discussions. SRT was supported by appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, administered by Oak Ridge Associated Universities and the Universities Space Research Association through a contract with NASA. AS is supported by a University Research Fellowship of the Royal Society. The authors acknowledge the support of colleagues in the EPTA. This work was supported in part by National Science Foundation Grant No. PHYS-1066293 and by the hospitality of the Aspen Center for Physics. A majority of the computational work was performed on the Nemo cluster at UWM supported by NSF grant No. 0923409. The research was partially carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. © 2016. All rights reserved.

[1] Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., and et al.,, Physical Review Letters **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[2] Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., and Siemens, X., Phys. Rev. D **79**, 084030 (2009), arXiv:0809.0701 [gr-qc].

[3] Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., Chamberlin, S. J., Chatterjee, S., Christy, B., Cordes, J. M., Cornish, N. J., Crowter, K., Demorest, P. B., Deng, X., Dolch, T., Ellis, J. A., Ferdman, R. D., Fonseca, E., Garver-Daniels, N., Gonzalez, M. E., Jenet, F., Jones, G., Jones, M. L., Kaspi, V. M., Koop, M., Lam, M. T., Lazio, T. J. W., Levin, L., Lommen, A. N., Lorimer, D. R., Luo, J., Lynch, R. S., Madison, D. R., McLaughlin, M. A., McWilliams, S. T., Mingarelli, C. M. F., Nice, D. J., Palliyaguru, N., Pennucci, T. T., Ransom, S. M., Sampson, L., Sanidas, S. A., Sesana, A., Siemens, X., Simon, J., Stairs, I. H., Stinebring, D. R., Stovall, K., Swiggum, J., Taylor, S. R., Vallisneri, M., van Haasteren, R., Wang, Y., Zhu, W. W., and The NANOGrav Collaboration,, ApJ **821**, 13 (2016), arXiv:1508.03024.

[4] Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., Donley, J., and Kocevski, D., A&A **564**, A125 (2014), arXiv:1402.0004 [astro-ph.HE].

[5] Caballero, R. N., Lee, K. J., Lentati, L., Desvignes, G., Champion, D. J., Verbiest, J. P. W., Janssen, G. H., Stappers, B. W., Kramer, M., Lazarus, P., Possenti, A., Tiburzi, C., Perrodin, D., Osłowski, S., Babak, S., Bassa, C. G., Brem, P., Burgay, M., Cognard, I., Gair, J. R., Graikou, E., Guillemot, L., Hessels, J. W. T., Karuppusamy, R., Lassus, A., Liu, K., McKee, J., Mingarelli, C. M. F., Petiteau, A., Purver, M. B., Rosado, P. A., Sanidas, S., Sesana, A., Shaifullah, G., Smits, R., Taylor, S. R., Theureau, G., van Haasteren, R., and Vecchio,

A., MNRAS **457**, 4421 (2016), arXiv:1510.09194 [astro-ph.IM].
[6] Chamberlin, S. J., Creighton, J. D. E., Siemens, X., Demorest, P., Ellis, J., Price, L. R., and Romano, J. D., Phys. Rev. D **91**, 044048 (2015), arXiv:1410.8256 [astro-ph.IM].
[7] Chamberlin, S. J. and Siemens, X., Phys. Rev. D **85**, 082001 (2012), arXiv:1111.5661 [astro-ph.HE].
[8] Champion, D. J., Hobbs, G. B., Manchester, R. N., Edwards, R. T., Backer, D. C., Bailes, M., Bhat, N. D. R., Burke-Spolaor, S., Coles, W., Demorest, P. B., Ferdman, R. D., Folkner, W. M., Hotan, A. W., Kramer, M., Lommen, A. N., Nice, D. J., Purver, M. B., Sarkissian, J. M., Stairs, I. H., van Straten, W., Verbiest, J. P. W., and Yardley, D. R. B., ApJ **720**, L201 (2010), arXiv:1008.3607 [astro-ph.EP].
[9] Cornish, N. J. and Sampson, L., Phys. Rev. D **93**, 104047 (2016), arXiv:1512.06829 [gr-qc].
[10] Edwards, R. T., Hobbs, G. B., and Manchester, R. N., MNRAS **372**, 1549 (2006), astro-ph/0607664.
[11] Feroz, F. and Hobson, M. P., MNRAS **384**, 449 (2008), arXiv:0704.3704.
[12] Feroz, F., Hobson, M. P., and Bridges, M., MNRAS **398**, 1601 (2009), arXiv:0809.3437.
[13] Feroz, F., Hobson, M. P., Cameron, E., and Pettitt, A. N., ArXiv e-prints (2013), arXiv:1306.2144 [astro-ph.IM].
[14] Foster, R. S. and Backer, D. C., ApJ **361**, 300 (1990).
[15] Gair, J., Romano, J. D., Taylor, S., and Mingarelli, C. M. F., Phys. Rev. D **90**, 082001 (2014), arXiv:1406.4664 [gr-qc].
[16] Gair, J. R., Romano, J. D., and Taylor, S. R., Phys. Rev. D **92**, 102003 (2015), arXiv:1506.08668 [gr-qc].
[17] Hellings, R. W. and Downs, G. S., ApJ **265**, L39 (1983).
[18] Hobbs, G., Coles, W., Manchester, R. N., Keith, M. J., Shannon, R. M., Chen, D., Bailes, M., Bhat, N. D. R., Burke-Spolaor, S., Champion, D., Chaudhary, A., Hotan, A., Khoo, J., Kocz, J., Levin, Y., Oslowski, S., Preisig, B., Ravi, V., Reynolds, J. E., Sarkissian, J., van Straten, W., Verbiest, J. P. W., Yardley, D., and You, X. P., MNRAS **427**, 2780 (2012), arXiv:1208.3560 [astro-ph.IM].
[19] Hobbs, G. B., Edwards, R. T., and Manchester, R. N., MNRAS **369**, 655 (2006), astro-ph/0603381.
[20] Jeffreys, H., *Theory of Probability*, 3rd ed. (Oxford, Oxford, England, 1961).
[21] Kass, R. E. and Raftery, A. E., Journal of the American Statistical Association **90**, 773 (1995).
[22] Keith, M. J., Coles, W., Shannon, R. M., Hobbs, G. B., Manchester, R. N., Bailes, M., Bhat, N. D. R., Burke-Spolaor, S., Champion, D. J., Chaudhary, A., Hotan, A. W., Khoo, J., Kocz, J., Osłowski, S., Ravi, V., Reynolds, J. E., Sarkissian, J., van Straten, W., and Yardley, D. R. B., MNRAS **429**, 2161 (2013), arXiv:1211.5887.
[23] Kennedy, J. and Eberhart, R., in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, Vol. 4 (1995) pp. 1942–1948 vol.4.
[24] Kulier, A., Ostriker, J. P., Natarajan, P., Lackner, C. N., and Cen, R., ApJ **799**, 178 (2015), arXiv:1307.3684.
[25] Lentati, L., Alexander, P., Hobson, M. P., Feroz, F., van Haasteren, R., Lee, K. J., and Shannon, R. M., MNRAS **437**, 3004 (2014), arXiv:1310.2120 [astro-ph.IM].
[26] Lentati, L., Alexander, P., Hobson, M. P., Taylor, S., Gair, J., Balan, S. T., and van Haasteren, R., Phys. Rev. D **87**, 104021 (2013), arXiv:1210.3578 [astro-ph.IM].
[27] Lentati, L., Taylor, S. R., Mingarelli, C. M. F., Sesana, A., Sanidas, S. A., Vecchio, A., Caballero, R. N., Lee, K. J., van Haasteren, R., Babak, S., Bassa, C. G., Brem, P., Burgay, M., Champion, D. J., Cognard, I., Desvignes, G., Gair, J. R., Guillemot, L., Hessels, J. W. T., Janssen, G. H., Karuppusamy, R., Kramer, M., Lassus, A., Lazarus, P., Liu, K., Osłowski, S., Perrodin, D., Petiteau, A., Possenti, A., Purver, M. B., Rosado, P. A., Smits, R., Stappers, B., Theureau, G., Tiburzi, C., and Verbiest, J. P. W., MNRAS **453**, 2576 (2015), arXiv:1504.03692.
[28] McWilliams, S. T., Ostriker, J. P., and Pretorius, F., ApJ **789**, 156 (2014).
[29] Mingarelli, C. M. F., Sidery, T., Mandel, I., and Vecchio, A., Physical Review D **88**, 062005 (2013), arXiv:1306.5394 [astro-ph.HE].
[30] Ravi, V., Wyithe, J. S. B., Shannon, R. M., and Hobbs, G., MNRAS **447**, 2772 (2015), arXiv:1406.5297.
[31] Rosado, P. A., Sesana, A., and Gair, J., MNRAS **451**, 2417 (2015), arXiv:1503.04803 [astro-ph.HE].
[32] Sesana, A., MNRAS **433**, L1 (2013), arXiv:1211.5375.
[33] Sesana, A., Vecchio, A., and Colacino, C. N., MNRAS **390**, 192 (2008), arXiv:0804.4476.
[34] Shannon, R. M. and Cordes, J. M., ApJ **725**, 1607 (2010), arXiv:1010.4794 [astro-ph.SR].
[35] Shannon, R. M., Ravi, V., Lentati, L. T., Lasky, P. D., Hobbs, G., Kerr, M., Manchester, R. N., Coles, W. A., Levin, Y., Bailes, M., Bhat, N. D. R., Burke-Spolaor, S., Dai, S., Keith, M. J., Oslowski, S., Reardon, D. J., van Straten, W., Toomey, L., Wang, J.-B., Wen, L., Wyithe, J. S. B., and Zhu, X.-J., Science **349**, 1522 (2015), arXiv:1509.07320.
[36] Shi, Y. and Eberhart, R., in *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on* (1998) pp. 69–73.
[37] Siemens, X., Ellis, J., Jenet, F., and Romano, J. D., Classical and Quantum Gravity **30**, 224015 (2013), arXiv:1305.3196 [astro-ph.IM].
[38] Taylor, S. R., "NX01: Updated release," (2016).
[39] Taylor, S. R. and Gair, J. R., Phys. Rev. D **88**, 084001 (2013), arXiv:1306.5395 [gr-qc].
[40] Taylor, S. R., Gair, J. R., and Lentati, L., Phys. Rev. D **87**, 044035 (2013), arXiv:1210.6014 [astro-ph.IM].
[41] Taylor, S. R., Vallisneri, M., Ellis, J. A., Mingarelli, C. M. F., Lazio, T. J. W., and van Haasteren, R., ApJ **819**, L6 (2016), arXiv:1511.05564 [astro-ph.IM].
[42] Tiburzi, C., Hobbs, G., Kerr, M., Coles, W. A., Dai, S., Manchester, R. N., Possenti, A., Shannon, R. M., and You, X. P., MNRAS **455**, 4339 (2016), arXiv:1510.02363 [astro-ph.IM].
[43] Verbiest, J. P. W., Lentati, L., Hobbs, G., van Haasteren, R., Demorest, P. B., Janssen, G. H., Wang, J.-B., Desvignes, G., Caballero, R. N., Keith, M. J., Champion, D. J., Arzoumanian, Z., Babak, S., Bassa, C. G., Bhat, N. D. R., Brazier, A., Brem, P., Burgay, M., Burke-Spolaor, S., Chamberlin, S. J., Chatterjee, S., Christy, B., Cognard, I., Cordes, J. M., Dai, S., Dolch, T., Ellis, J. A., Ferdman, R. D., Fonseca, E., Gair, J. R., Garver-Daniels, N. E., Gentile, P., Gonzalez, M. E., Graikou, E., Guillemot, L., Hessels, J. W. T., Jones, G., Karup-

pusamy, R., Kerr, M., Kramer, M., Lam, M. T., Lasky, P. D., Lassus, A., Lazarus, P., Lazio, T. J. W., Lee, K. J., Levin, L., Liu, K., Lynch, R. S., Lyne, A. G., Mckee, J., McLaughlin, M. A., McWilliams, S. T., Madison, D. R., Manchester, R. N., Mingarelli, C. M. F., Nice, D. J., Osłowski, S., Palliyaguru, N. T., Pennucci, T. T., Perera, B. B. P., Perrodin, D., Possenti, A., Petiteau, A., Ransom, S. M., Reardon, D., Rosado, P. A., Sanidas, S. A., Sesana, A., Shaifullah, G., Shannon, R. M., Siemens, X., Simon, J., Smits, R., Spiewak, R., Stairs, I. H., Stappers, B. W., Stinebring, D. R., Stovall, K., Swiggum, J. K., Taylor, S. R., Theureau, G., Tiburzi, C., Toomey, L., Vallisneri, M., van Straten, W., Vecchio, A., Wang, Y., Wen, L., You, X. P., Zhu, W. W., and Zhu, X.-J., MNRAS **458**, 1267 (2016), arXiv:1602.03640 [astro-ph.IM].

[44] Woodbury, M. A., Memorandum report **42**, 106 (1950).