EXPLORING THE LOW-FREQUENCY GRAVITATIONAL-WAVE UNIVERSE

WITH PULSAR TIMING ARRAYS

by

Bence Bécsy

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Physics

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2022

DEDICATION

I dedicate this thesis to my parents Ildikó and Ferenc, and to my wife Villő. I could not have done any of this without their love and support.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Neil Cornish, who I learned a tremendous amount from. I am also greatful for all the feedback I received over the years from the members of the MSU Gravity Group and the NANOGrav Collaboration.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

Pulsar timing arrays monitor millisecond pulsars to detect gravitational waves with nanohertz frequencies. They provide valuable information about various astophysical processes inaccessible to electromagnetic observations. In particular, they could shed light on unsolved problems related to the formation and evolution of supermassive black holes. We present several new methods which will help us fully realize the detection potential of pulsar timing arrays. We explore how the large collection of supermassive black hole binaries in the Universe can appear as a stochastic gravitational wave background, and how it might also result in a few individually detectable binaries. We describe a new method to efficiently search for such individual binaries, and also how we can detect multiple binaries in the presence of the confusion noise from the stochastic background. Finally, we introduce a new approach to search for generic gravitational-wave bursts, which enables us to hunt for unexpected new types of sources on the nanohertz gravitational-wave sky.

CHAPTER ONE

INTRODUCTION

Einstein's theory of general relativity is the modern description of the gravitational force, which passed every observational test we ever conducted. One of its predictions is the existence of wavelike perturbations of spacetime traveling at the speed of light called gravitational waves (GWs). Their first direct observation was achieved by the Laser Interferometer Gravitational-Wave Observatory (LIGO), which detected GWs from an inspiralling binary black hole system in 2015 [1]. Since then, LIGO has observed GWs from dozens of compact binaries [2], which are now used to study a diverse array of phenomena in astrophysics.

Pulsar timing arrays (PTAs) offer a different GW detection method. They monitor a collection of millisecond pulsars looking for the signature of nHz GWs in the arrival times of radio pulses. There are three major PTA efforts around the globe: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav, [3]), the European Pulsar Timing Array (EPTA, [4]), and the Parkes Pulsar Timing Array (PPTA, [5]). These collaborations share their data and analyze combined datasets as part of the International Pulsar Timing Array (IPTA, [6]). Both individual PTAs and the IPTA recently reported evidence for a common red noise process seen in their datasets [7, 8, 9, 10], which could potentially be the first sign of a stochastic GW background (GWB). The only way to definitively distinguish a GWB from other noise sources common to all pulsars is to measure the characteristic correlations between the signal in different pulsars. The expected correlation as a function of the angular separation between two pulsars is described by the Hellings–Downs curve [11], which predicts positive correlation at small and large angular separations and anticorrelation at moderate angular separation. Currently, there is no evidence for such correlations, but if the observed noise process is indeed due to the GWB, the correlations are expected to emerge within the next few years [12]. With the imminent detection of low-frequency GWs, PTAs will undergo a similar transition as LIGO: we will soon be able to use them not only to detect GWs, but also to answer exciting questions in astrophysics. This thesis presents new methods to understand and search for various GW signals in the nHz frequency band. These will help us realize the full potential of the emerging field of PTA astrophysics.

The most promising sources in the nHz frequency range monitored by PTAs are supermassive black hole binaries (SMBHBs). The vast majority of these result in the emergence of a GWB, while the few brightest binaries could be detected individually. In Chapter 2 we explore realistic simulated datasets using synthetic catalogs of SMBHBs based on the Illustris cosmological hydrodynamic simulation [13]. We show how such a realistic GWB is different from a simple isotropic Gaussian background with a power-law spectrum. We also investigate how the confusion noise from the background affects the prospects of individually detecting SMBHBs.

While the GWB provides information about the overall population of SMBHBs, detecting GWs from an individual SMBHB would allow us to study that particular system in detail, and could also allow for multi-messenger studies. The intrinsic GW waveform of individual SMBHBs is relatively simple, because most of them will not evolve appreciably during the observational timespan of PTAs. However, the way they appear in the data can be quite complex, which makes it challenging to effectively sample the large parameter space involved. In Chapter 3 we present an improved method to search for individual binaries, which represents a substantial speedup compared to previously used methods. This allows for more efficient parameter exploration, and potentially for carrying out analyses that were previously computationally prohibitive.

As PTAs get more sensitive, they will be able to individually detect GWs from multiple SMBHBs. However, we do not know a priori how many SMBHBs should be modeled individually, and how many can be described collectively as a GWB. In Chapter 4 we present a new approach that uses a trans-dimensional Markov chain Monte Carlo (MCMC) algorithm to find the number of individual sources most preferred by the data. This methodology enables us to search for an unspecified number of individual binaries in the presence of a GWB, and assess the influence of these two models on each other. This will be crucial in understanding how the presence of individual sources can bias the parameter estimation of the GWB, and vice versa.

Detecting SMBHBs can be extremely valuable in understanding their formation and evolution. However, detecting an unexpected source would have an even larger impact on our understanding of astrophysics, as it was demonstrated many times throughout the history of astronomy (e.g. the cosmic microwave background, gamma-ray bursts, soft gamma repeaters, or fast radio bursts). So it is worthwhile to search for previously unimagined sources that may emit GW bursts with unknown morphology in the nHz band. There are several searches targeting such unexpected sources at higher frequencies using LIGO data. In Chapter 5 we present new methods to search for such generic burst signals in PTA data. The signal is modeled with a collection of sine-Gaussian wavelets, where the number of wavelets is searched over using trans-dimensional MCMC techniques. We similarly model noise transients that may occur in individual pulsars. Whether the data prefers the coherent signal model or the incoherent noise transient model will determine if we detected a GW burst. In addition, these methods can also be used to search for previously undiscovered noise transients in the data.

CHAPTER TWO

EXPLORING REALISTIC NANOHERTZ GRAVITATIONAL-WAVE BACKGROUNDS

2.1 Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Bence Bécsy

Contributions: Developed methods and code needed for the study. Performed the analysis. Wrote the first draft of the manuscript.

Author: Neil J. Cornish

Contributions: Conceived the study design. Provided guidance on the analysis. Edited the manuscript.

Author: Luke Z. Kelley

Contributions: Supplied expertise and code to produce synthetic datasets of binaries. Provided feedback on the analysis and the manuscript.

2.2 Manuscript Information

Bence Bécsy, Neil J. Cornish, Luke Z. Kelley

Astrophysical Journal

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
_X_ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

2.3 Introduction

Pulsar timing arrays (PTAs) probe gravitational waves (GWs) with frequencies between a few and a few hundred nanohertz (nHz) by continuously monitoring millisecond pulsars (for a review see e.g. Refs. [14, 15]). The NANOGrav collaboration recently found evidence for a low-frequency stochastic process common to all pulsars in their 12.5-year dataset [16]. The three other PTAs also all found evidence for such a common red noise process (International Pulsar Timing Array (IPTA) [10] , EPTA [8], PPTA [9]). It is possible that these results are the first signs of a stochastic gravitational-wave background (GWB), but at this point none of the above analyses found strong evidence for Hellings-Downs (HD) spatial correlations that are the tell-tale sign of a gravitational background. However, this is not surprising given that the power in cross-correlations is about an order of magnitude less then the total power of a GWB. This means that we can expect to detect a common process before we detect the HD correlations between pulsars (see e.g. Ref. [12]). Upcoming PTA datasets will decide whether this red noise process is the GWB or not.

Once a firm detection of the GWB is established, the next task will be to identify its origin. The theoretically most favored source of a GWB at nHz frequencies is an ensemble of inspiralling supermassive black hole binaries (SMBHBs, see e.g. Refs. [17, 18, 19, 20]). However, there are several other proposed processes that can form such a low-frequency GWB, like cosmological phase transitions [21], cosmic strings [22, 23], primordial black holes [24], inflation [25], etc. These models all have different predictions for the spectral shape of the common signal, which presents the possibility of distinguishing these models. However, definitively identifying the source of the GWB based solely on its spectrum will be challenging given the large uncertainty in both measurement [12] and model predictions. Thus any additional predictions of these models beyond their spectra could be useful in determining the source of the GWB.

In this paper we explore various properties of realistic GWBs from SMBHBs, some of which might help distinguish that scenario from others. As we will see, most of these properties rely on the fact that an SMBHB-based GWB is built from a finite number of sources. The general question of when the signal from a finite collection of individual sources becomes effectively stochastic was investigated in Ref. [26], while the effects of the finite population on the detection prospects of the GWB have been studied in Ref. [27]. Here we are instead focusing on how these effects can help us discern the origin of the GWB. Our simulations are using synthetic SMBHB catalogs based on the Illustis cosmological simulation [28]. We introduce and validate a method that can rapidly produce simulated PTA datasets from millions of SMBHBs by modeling the large majority of the sources as a GWB and directly simulating the few brightest binaries in each frequency bin (see Section 2.4). We use this simulation method to investigate how the central limit theorem breaks down for such an SMBHB population (see Section 2.5.1), how the finite number of sources can lead to anisotropic GWBs (see Section 2.5.2), and how the variance of the HD correlations is affected by the properties of the SMBHB population (see Section 2.5.3). We also explore the detectability of individual binaries in the presence of a GWB by calculating the distribution of signal-to-noise ratios (SNRs) of the brightest binaries in multiple realizations (see Section 2.6). We summarize and offer concluding remarks in Section 2.7.

2.4 Simulation methods

2.4.1 SMBH population model

Our simulations are based on synthetic catalogs of the SMBHB population representative of the entire observable universe. These are derived from the Illustris cosmological hydrodynamic simulations (see e.g. Ref. [29]) by modeling small-scale astrophysical processes in a post-processing step. This treatment accounts for environmental effects relevant to binary hardening like dynamical friction, stellar scattering, drag from a circumbinary disk, and GW emission (for more details see Refs. [28, 30, 20]). The collection of mergers observed in the Illustris simulation volume can be resampled multiple times to create new realizations of a simulated catalog of SMBHBs for the entire past light-cone of the observer.

Figure 2.1 shows the distribution of the source-frame chirp mass ($\mathcal{M}$), the luminosity distance ($d_{\rm L}$), and the observer-frame GW frequency ($f_{\rm obs}$) of such a simulated dataset of SMBHBs. Here we applied a lower frequency cutoff at (15 year)$^{-1}$, corresponding to the observational timespan we consider in this paper and the observational timespan of the upcoming NANOGrav 15-year dataset. This particular realization has about 207 million binaries. Note that the number density is dominated by low-frequency, low-mass, faraway systems.

2.4.2 Isotropic GWB plus bright binaries

The timing residuals in a PTA dataset at observing epochs t_i can be written as:

$$r(t_i) = r_{\rm n}(t_i) + r_{\rm GW}(t_i), \tag{2.1}$$

where $r_{\rm n}$ describes residuals due to every noise source, while $r_{\rm GW}$ is the contribution of GWs. In case of a population of N SMBHBs producing GWs in the observable universe, one can write:

$$r_{\rm GW}(t_i) = \sum_{j=1}^{N} s_j(t_i; \boldsymbol{\theta}_j), \tag{2.2}$$

where s_j is the timing residual response to the GW signal from the jth binary, which is described by the parameters $\boldsymbol{\theta}_j = \{\mathcal{M}, d_{\rm L}, f_{\rm obs}, \theta, \phi, \iota, \psi, \Phi_0\}$, where θ and ϕ parametrize the location of the source on the sky, ι is the inclination of the binary's orbit, ψ is the polarization angle, and Φ_0 is the initial phase of the GW signal (for an explicit expression of $s_j(t_i; \boldsymbol{\theta}_j)$ see e.g. Eq. (10) in Ref. [31]). While Eq. (2.2) is a valid description of the total GW signal, it is impractical both for data analysis and for simulating signals. In terms of data analysis, this

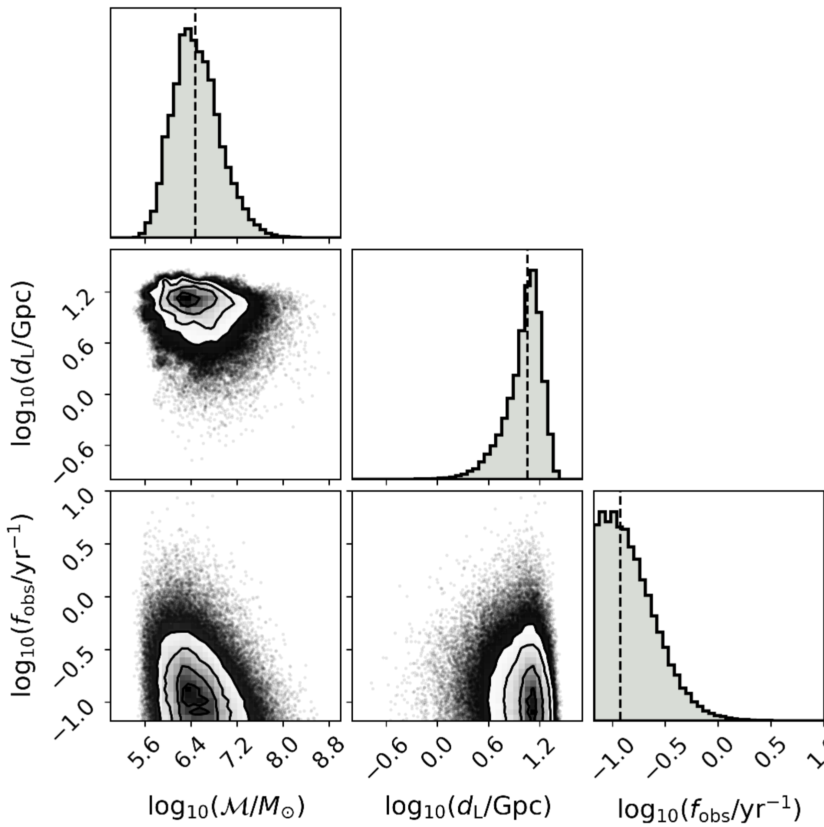


Figure 2.1: Distribution of binary parameters in the simulated SMBHB catalog we use. The distribution was artificially cut off below observer-frame GW frequencies of (15 year)$^{-1}$, which would not be observable with the 15 year observing timespan considered here. Vertical dashed lines indicate the median value of each parameter.

model has $8N$ parameters, which are practically impossible to explore in the realistic case, where $N \sim \mathcal{O}(10^6 - 10^8)$. In terms of simulating PTA datasets, individually calculating the response of millions of binaries becomes a computationally expensive task.

We will show that both of these problems can be averted by expressing the total contribution to the residuals as:

$$r_{\rm GW}(t_i) = r_{\rm GWB}(t_i, h_c^{\rm GWB}) + \sum_{k=1}^{N_{\rm freq}} \sum_{j=1}^{M} s_j^{(k)}(t_i; \boldsymbol{\theta}_j), \tag{2.3}$$

where $r_{\rm GWB}$ is an isotropic Gaussian stochastic background with the characteristic strain spectrum $h_c^{\rm GWB}$, $N_{\rm freq}$ is the number of frequency bins considered, and $s_j^{(k)}$ is the binary with the jth largest characteristic strain in the kth frequency bin. The second term in Eq. (2.3) loops over all $N_{\rm freq}$ frequency bins and directly adds the contribution of the M brightest sources in each bin. We set $h_c^{\rm GWB}$ in each frequency bin based on the remaining SMBHBs as:

$$h_c^{\rm GWB}(f_k) = \sqrt{\sum_{j=M+1}^{N_k} \left[(h_c)_j^{(k)} \right]^2}, \tag{2.4}$$

where N_k is the total number of binaries in the kth frequency bin, f_k is the central frequency of that bin, and

$$(h_c)_j^{(k)} = h_j^{(k)} \sqrt{f_j^{(k)} T_{\rm obs}} \tag{2.5}$$

is the characteristic strain of the jth brightest source in the kth frequency bin, $T_{\rm obs}$ is the total observation time, $f_j^{(k)}$ is the observer-frame GW frequency of that source, and $h_j^{(k)}$ is the sky and polarization averaged GW amplitude [18]:

$$h_j^{(k)} = \frac{8}{\sqrt{10}} \frac{(G\mathcal{M}_{\rm obs})^{5/3} (\pi f_{\rm obs})^{2/3}}{c^4 d_{\rm L}}, \tag{2.6}$$

where $\mathcal{M}_{\rm obs} = (1+z)\mathcal{M}$ is the observer-frame chirp mass. Note that Eq. (2.4) determines $h_c^{\rm GWB}$ by summing up the square of the characteristic strain for all binaries except the ones directly modeled in Eq. (2.3). This is justified by the fact that when summing up sine-waves of a given frequency with random phase offsets, the expectation value of the total amplitude squared is given by the sum of the squares of individual amplitudes. This approximately still holds in this scenario, where the frequencies of sources in the same frequency bin is approximately the same.

This method of describing a realistic background by an idealized isotropic Gaussian GWB and a sum of the few brightest sources can be used for both data analysis and simulation. The `BayesHopper` [1] algorithm uses this approach to search for multiple individual SMBHBs in the presence of a GWB [32]. In this paper we use this approach to efficiently create realistic simulated datasets. Note that for data analysis purposes, the ideal number of SMBHBs to model individually is dictated by Bayesian parsimony. However, for simulating signals we want to make sure that all the relevant details are captured, so we add more binaries directly than strictly necessary, and than what can be picked up by data analysis. To validate the use of this description for rapidly simulating realistic GWBs, we simulated 800 realizations of a GWB by directly simulating all binaries (as in Eq. (2.2)) and also by simulating a GWB and 1000 outlier SMBHBs in each frequency bin (as in Eq. (2.3)). Figure 2.2 shows the median power spectral density (PSD) for both methods. The 90% credible intervals are shown as shaded bands. Both methods produce spectra with a significant variance over realizations, but both their median values and 90% credible intervals show good agreement. Note that the two peaks correspond to frequencies of 1/year and 2/year, where the timing model fit introduces extra power. The agreement between the PSDs produced by the two methods show that they result in statistically equivalent frequency content. In Sections 2.5.1 and 2.5.2 we show that after removing the brightest binaries, the remaining signal satisfies the assumption of Gaussianity and statistical isotropy as well.

2.5 Properties of a realistic GW background

In this section we explore how various characteristics of the background are affected by the fact that there is a finite number of sources contributing to it. We investigate the spectrum of the GWB in Section 2.5.1, we test for statistical isotropy in Section 2.5.2, and quantify the variance of the HD correlations in Section 2.5.3.

[1] Publicly available at: `https://github.com/bencebecsy/BayesHopper`

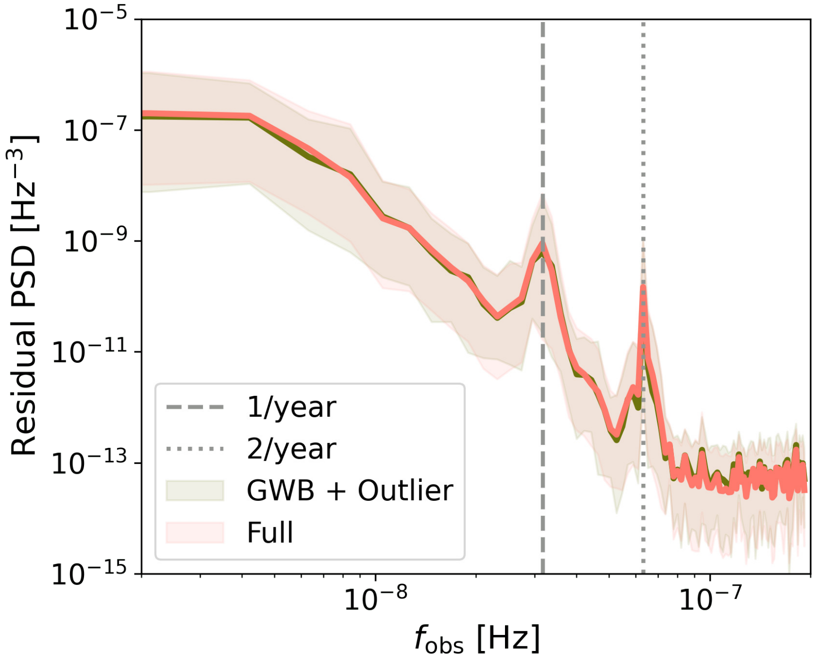


Figure 2.2: Comparison of individually simulating all binaries (red) and simulating a stochastic background and one thousand SMBHBs in each frequency bin (green). We show the median and 90% credible intervals over 800 realizations as solid lines and shaded regions, respectively.

2.5.1 GW spectrum

In the canonical description of a GWB arising from an infinite number of circular SMBHBs evolving purely through GW emission, the characteristic strain spectrum can be described as a power-law with a spectral index of -2/3, i.e. $h_c^{\mathrm{GWB}} \sim f_{\mathrm{obs}}^{-2/3}$ [33]. The spectrum arising from a population of SMBHBs differs from this simple spectrum (see purple dots and dashed line in Fig. 2.3). The difference is most striking at high frequencies, where the population-based spectrum has a lower amplitude than the power-law model. This is due to the fact that the power-law model counts non-physical contributions from fractional sources [34]. The population-based spectrum also exhibits significant fluctuations at high frequencies. Fig. 2.3 also shows the spectrum when we exclude the brightest 1/10/100/1000 binaries in each frequency bin [2]. We can see that as we exclude more binaries, the scatter in the spectrum decreases. This suggests that by setting $M \gtrsim 100$, the resulting spectrum is not dominated by a few bright binaries anymore. Note that as we remove more binaries, the spectra in Fig. 2.3 terminate at lower frequencies, because at higher frequencies we removed all the binaries.

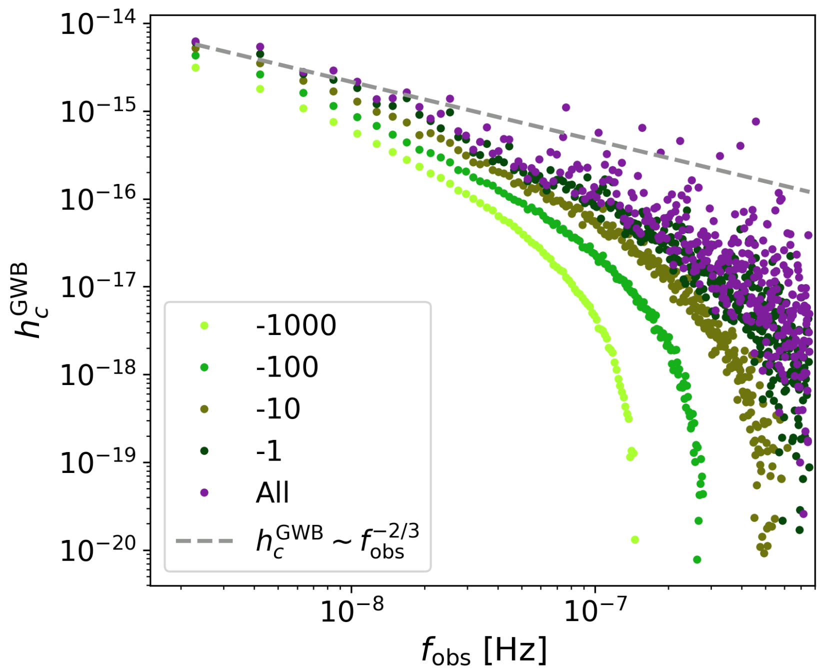


Figure 2.3: Characteristic strain spectrum for all binaries (purple) and for various numbers of the brightest binaries removed (different shades of green). We also show a power-law spectrum for reference (dashed gray line).

To quantify how much $h_c^{\rm GWB}$ is dominated by a few sources at a given frequency, we look at the minimum number of binaries contributing at least 90% of $h_c^{\rm GWB}$ at a given frequency (N_{90}). Figure 2.4 shows the median N_{90} value over 20 realizations. We show this for the total population and also for populations where we removed the top $M = 1/10/100/1000$ binaries. The total number of binaries in each frequency bin is also indicated. While the bulk of the

[2]This corresponds to setting $M = 1/10/100/1000$ in Eq. (2.4)

GWB signal is built up from $\mathcal{O}(10^3)$ binaries at the lowest frequency bins, it is dominated by less than 10 binaries above about 100 nHz. This is expected from our qualitative assessment of the spectra in Fig. 2.3, where we have seen an increased scatter at high frequencies. We can also see that as we remove the brightest binaries from the population, the effective number of binaries contributing to the GWB increases. Note that as we remove more and more binaries, we can reach a limit where we are removing a significant fraction of the total number of binaries, and thus removing more binaries actually decreases N_{90}. See e.g. "-1000" and "-100" lines around 40 nHz in Fig. 2.4). We also show the frequencies where the total number of binaries reach 100 and 1000 (vertical dashed lines). Note that the "-1000" and "-100" lines approach these horizontal lines, indicating the point where we remove all the binaries in the given bin.

Figure 2.4 illustrates why modeling a realistic background as a GWB and a few individual sources is well motivated. Modeling all the binaries as a GWB is not justified, since the background level for the full population is dominated by a few sources at high frequencies, so the central limit theorem breaks down. However, if we only consider the population with hundreds of the brightest binaries removed, the background level is determined by a large ($\gtrsim 100 - 1000$) number of binaries, so treating it as a Gaussian background is a good approximation. In this case the non-Gaussian nature of the background is modeled by adding the brightest binaries as individual sources to the full signal (see Eq. (2.3)).

2.5.2 Statistical isotropy

In the simplest description, it is also usually assumed that the GWB exhibits statistical isotropy. However, this is not expected to be true for realistic backgrounds. SMBHBs are hosted by galaxies, which cluster into galaxy clusters, so nearby galaxy clusters can result in an overabundance of nHz GW sources in a particular sky location. In addition, SMBHBs can produce anisotropy even when they are uniformly distributed in volume, if a small number

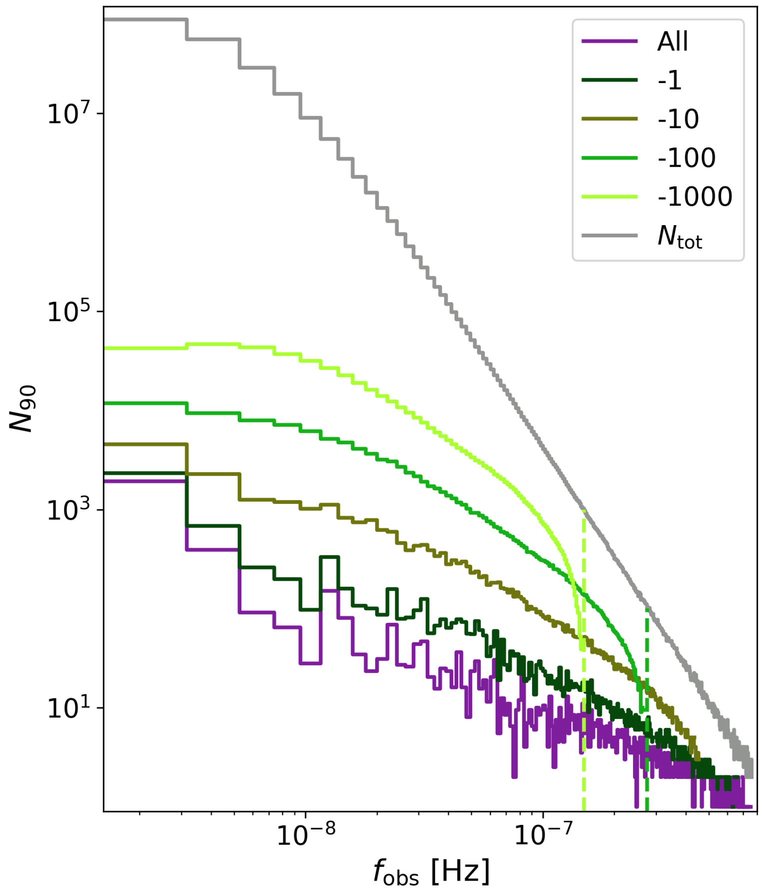


Figure 2.4: Number of binaries contributing 90% of the total strain in each frequency bin (N_{90}). We show this for the whole dataset (purple) and for datasets with the top 1/10/100/1000 binaries removed (different shades of green). We also show the total number of binaries in each bin ($N_{\rm tot}$, gray histogram). Vertical dashed lines indicate frequencies where the total number of binaries reach 1000 and 100. Note that the lines for the datasets without the top 1000/100 asymptote to the corresponding vertical lines, indicating frequency bins where we remove a large fraction of the binaries.

of them dominate the nHz GW sky. This is in stark contrast with a GWB of primordial cosmological origin, which would be perfectly statistically isotropic, so the origin of the GWB can be firmly established by detecting anisotropy [35]. Indeed, there are numerous methods to search for signatures of anisotropy in the GWB [36, 35, 37, 38, 39, 40], and by some estimates PTAs might be able to detect anisotropy just a few years after detecting the

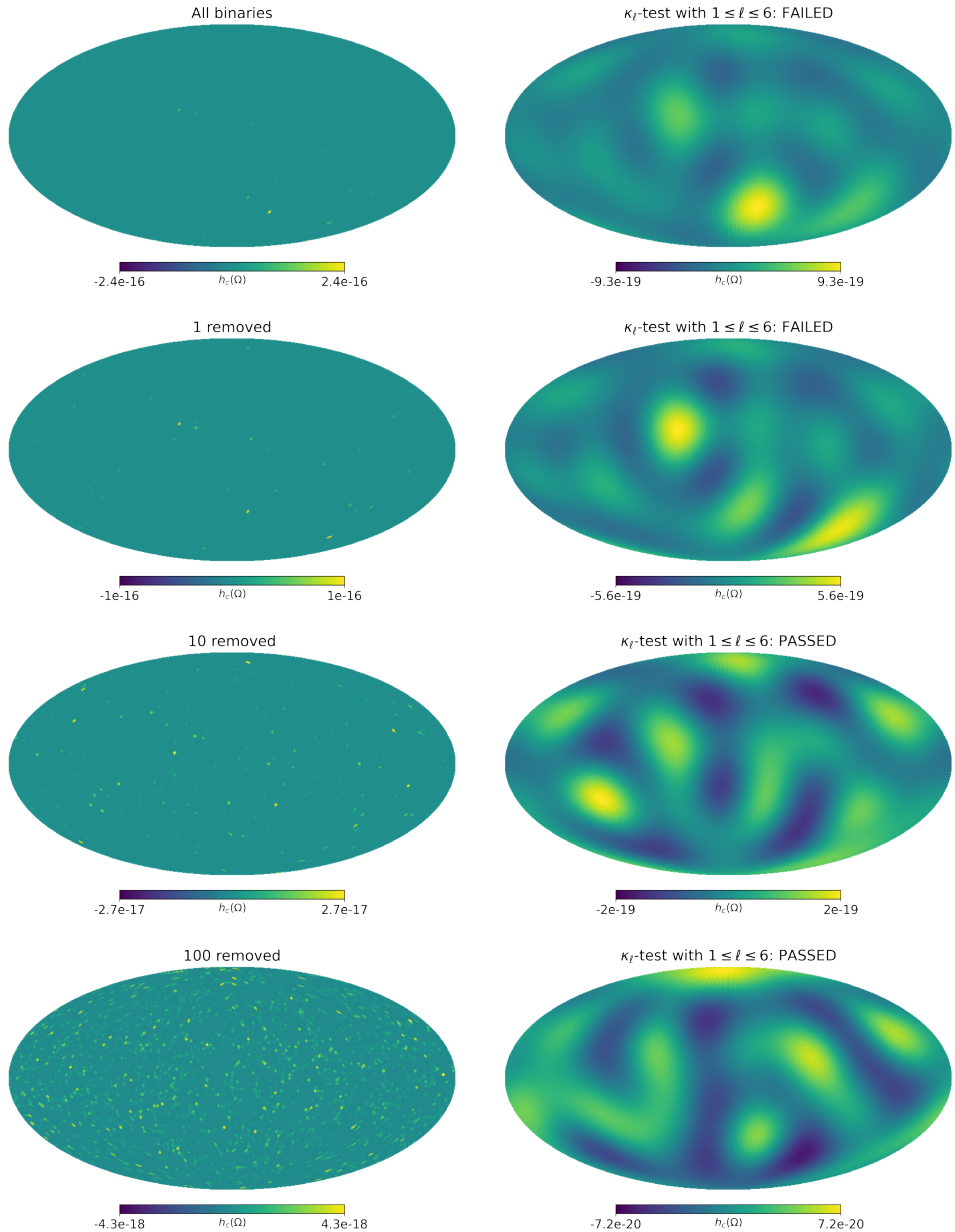


Figure 2.5: Angular distribution of the characteristic amplitude in the frequency bin with 64.4 nHz< $f_{\rm obs}$ <66.5 nHz. We show sky maps for all binaries (first row), and with 1/10/100 brightest binaries removed (2nd/3rd/4th rows). The left column shows the raw maps, while the right column shows the maps reconstructed with the same range of angular scales as was used for the anisotropy test ($1 \leq \ell \leq 6$). Only the maps with 10 and 100 binaries removed passed the isotropy test in this example. The monopole term was subtracted from each map. Note the changing scale of the colormaps, which we forced to be symmetric around 0.

isotropic component of the GWB [41].

In our modeling we assume a uniform in volume distribution of sources. This means that the following results on anisotropy can be taken as a lower limit, since the effect of local inhomogeneities would further increase the level of anisotropy. To model the angular distribution of GW power, we pixelate the sky using the `HEALPix` framework [42]. Since this produces equal-area pixels, randomly assigning each SMBHB to a pixel is equivalent to distributing them uniformly on the sky. We carry out our analysis independently for each frequency bin. For each pixel, we sum up h_c^2 values of all the binaries assigned to that pixel, resulting in a discretized function of sky location $h_c^2(\Omega)$.

Figure 2.5 shows that distribution for a particular realization at the frequency bin between 64.4 nHz and 66.5 nHz. This frequency bin has a total number of 20,077 binaries. The different rows show maps with different numbers of the brightest binaries removed. For each of these we show both the raw map of $h_c(\Omega)$ (left) and a reconstructed map with angular scales restricted to those used for the anisotropy test described below (right). Note that the mean amplitude has been subtracted to better show the fluctuations. We allow the colorscales to be different for each sky map, but force it to be symmetric around 0. This particular example shows a highly anisotropic distribution due to a few brightest binaries dominating the sky, but as more and more of the binaries are excised, the remaining GWB gets more and more isotropic. By the time we remove the top 10 binaries, the GWB sky map restricted to large angular scales is visibly indistinguishable from an isotropic distribution. This is similar to the findings of Ref. [43], where they found that the map can be made more isotropic by removing the brightest source (see their Fig. 2).

To quantitatively assess the statistical isotropy of maps like those shown in Fig. 2.5, we follow the formalism introduced in Refs. [44, 45]. This method was used to test the statistical isotropy of the cosmic microwave background using data from the Wilkinson Microwave Anisotropy Probe [46]. We start by decomposing our maps in terms of Y_ℓ^m

spherical harmonics:

$$h_c(\Omega) = h_c(\theta, \varphi) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} a_\ell^m Y_\ell^m(\theta, \varphi). \tag{2.7}$$

The a_ℓ^m coefficients can be used to compute the bipolar power spectrum, κ_ℓ as defined in Eq. (1) of Ref. [46]. After subtracting the contribution expected from an isotropic map (Eq. (14) of Ref. [44]), κ_ℓ has an expectation value of 0 for all $\ell \geq 1$ under statistical isotropy. The variance can also be analytically computed (Eq. (17) of Ref. [44]), so one can test for statistical isotropy by checking if κ_ℓ are consistent with zero within their theoretical variance.

We calculate κ_ℓ for $1 \leq \ell \leq 6$, which corresponds to probing angular scales larger than $30°$. The cut-off at $\ell = 6$ is motivated by the angular resolution sensitivity forecasts from Ref. [41]. We apply a Gaussian low-pass filter on the a_ℓ^m values as defined in Eq. (5) of Ref. [46] with $l_s = 20$. This ensures that there is no spectral leakage affecting the analysis. A key difference between our analysis and the one presented in Ref. [46] is that we do not use the theoretical variance of κ_ℓ to decide if a map is isotropic or not. Instead, we compare the measured κ_ℓ values to κ_ℓ values calculated from N_{iso} realizations of a simulated isotropic map with the same angular power spectrum. We call a map statistically isotropic if the measured κ_ℓ values are within the maximum and minimum κ_ℓ from the truly isotropic maps for all ℓ. This procedure allows us to set the false positive and false negative probabilities of our test by choosing the value of N_{iso}. As discussed below, we perform multiple trials of this test, so we want to make sure that anisotropic maps rarely pass the test by chance. We choose $N_{\text{iso}} = 5$, which results in the false positive and false negative rates reported in Table 2.1. Note that increasing N_{iso} would further decrease the probability of anisotropic maps passing the test at the cost of fewer truly isotropic maps passing it.

We performed κ_ℓ-tests on all four maps shown on the right column of Figure 2.5. The maps with all binaries and with the single brightest binaries removed failed the test, and are

Table 2.1: False negative and false positive probabilities of the anisotropy test with $N_{\rm iso} = 5$.

Isotropy test result	Pass	Fail
Truly isotropic map	57%	43%
Truly anisotropic map[3]	12%	88%

thus not consistent with being statistically isotropic. The maps with 10 and 100 binaries removed passed the test. It is instructive to compare the maps in the left and right columns. Based on the raw maps we might say that even the map with 10 binaries removed is not isotropic. However, the maps on the right correspond to what the κ_ℓ-test actually sees. It is evident even in those large angular scale features that the first two maps are dominated by a few sources. Note in particular that these maps are highly asymmetric around their mean values resulting in the lack of dark blue colors. However, once we remove the 10 brightest binaries, the map is consistent with statistical isotropy up to $\ell = 6$. This example reinforces our previous qualitative assessment that removing the few brightest binaries makes the GWB more isotropic.

We can ask how many binaries one needs to remove to get a background consistent with statistical isotropy. To answer that we start with a κ_ℓ-test on a map with all binaries, and remove binaries one at a time until we reach a statistically isotropic map. We call the number of binaries removed to achieve this $N_{\rm rem}$. We repeat this procedure over all frequency bins and multiple realizations (see Fig. 2.6). We can see that there is a trend of needing to remove more binaries as we move to higher frequencies. Figure 2.6 also shows the number of binaries with an amplitude within a factor of three of the brightest binary in the population that achieved statistical isotropy, $N_{\rm top}(33.3\%)$. We can see that we typically reach isotropy when there is more than ~ 10 binaries with comparable amplitudes. This is consistent with the fact that at $\ell = 6$ there are 13 different spherical harmonics, so once there are $\gtrsim 10$ comparable sources randomly placed on the sky, we cannot distinguish that from an isotropic sky. Note that at the highest frequencies $N_{\rm rem} \simeq N_{\rm tot}$, where $N_{\rm tot}$ is the total number of binaries in each frequency bin. This is because in our procedure we set $N_{\rm rem} = N_{\rm tot}$ if we cannot achieve statistical isotropy with any value of $N_{\rm rem}$. This is a similar feature to the one in Fig. 2.4, where removing more binaries does not get us closer to a true stochastic background when $N_{\rm tot}$ is small to start with. We also show the expected value of $N_{\rm rem}$ assuming all maps are isotropic, $\mathrm{E}[N_{\rm rem}]_{\rm isotropy} = 0.89$, which is based on the false negative probability of 43% reported in Table 2.1. We can see that at the lowest frequencies, the median value of $N_{\rm rem}$ is consistent with this expectation value, suggesting that these maps tend to be isotropic even without removing any bright binaries. Note that the expected value of $N_{\rm rem}$ assuming every map is anisotropic is ~ 60 and thus not expected to affect the results.

2.5.3 Variance of the Hellings-Downs curve

It can be shown that an isotropic, unpolarized, Gaussian GWB produces residuals that are correlated between pulsar pairs, and the amount of correlation between pulsar i and pulsar j is described by the HD curve [11][4]:

$$\Gamma_{ij} = \frac{1}{2} - \frac{1}{4}x_{ij} + \frac{3}{2}x_{ij}\ln(x_{ij}) + \frac{1}{2}\delta_{ij}, \tag{2.8}$$

where the term with δ_{ij} encodes correlations due to the pulsar terms and it vanishes when looking at different pulsars, and x_{ij} is defined as:

$$x_{ij} = \frac{1 - \cos\gamma_{ij}}{2}, \tag{2.9}$$

where γ_{ij} is the angular separation between pulsars i and j. It has been shown in Ref. [47] that Eq. (2.8) not only holds for GWBs, but also for a quasi-constant frequency continuous-

[4]Note that we follow the normalization convention where $\Gamma_{ii} = 1$.

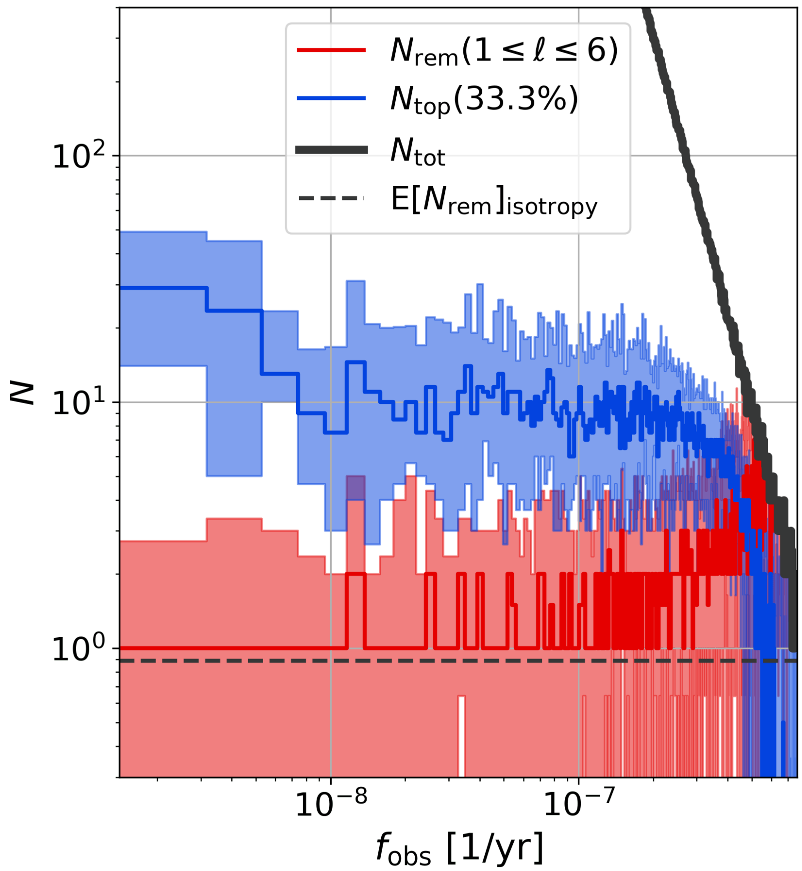


Figure 2.6: Minimum number of sources removed to achieve statistical isotropy based on the κ_ℓ-test with ℓ between 1 and 6, $N_{\mathrm{rem}}(1 \leq \ell \leq 6)$. We show the median and 68% credible interval of N_{rem} over 30 GWB realizations. We also show the number of binaries with an h_c within a factor of 3 of the brightest binary in the frequency bin, $N_{\mathrm{top}}(30\%)$. Note how statistical isotropy is reached when $N_{\mathrm{top}}(30\%) \sim 10$. The black line shows the total number of sources in each frequency bin. The dashed horizontal line shows the expected value of N_{rem} if all maps are isotropic and removals are only due to the false negative probability of our test (see Table 2.1).

wave (CW) signal from an individual SMBHB.

While Eq. (2.8) predicts the expectation value of the correlations as a function of γ, individual correlations can have a significant variance (see Ref. [48] and references therein). It has been suggested that the way the variance varies with angle might be used to distinguish between primordial and astrophysical sources of the GWB [48].

There are two sources of that variance that are present even in idealized noise-free observations, which we can call pulsar variance and cosmic variance, following the nomenclature of Ref. [48]. Pulsar variance arises from the fact that individual correlations can depend not only on γ, but also on three other angles describing the position of the two pulsars on the sky. Once we average over those three additional angles the mean correlations follow Eq. (2.8), but individual correlations can be significantly different. It is clear that having a large number of pulsar pairs, one can in principle average out the pulsar variance. Cosmic variance refers to statistical deviations from Eq. (2.8) in a given realization of the GWB. Unlike pulsar variance, we cannot average out cosmic variance, since we only have one universe to make observations of.

To quantify the variance of the HD correlations, we performed three different simulations, each with 100 realization:

(i) GWB: a true stochastic background with a characteristic strain amplitude of 1×10^{-15} measured at $f_{\rm obs} = 1$ year^{-1} and spectral index of -2/3;

(ii) CW: GW signal of an individual SMBHB with fixed amplitude, $\mathcal{M} = 10^9\ M_\odot$, $f_{\rm obs} = 10$ nHz, and nuisance parameters drawn randomly from uniform distributions;

(iii) Population: realistic GWB simulated as in Eq. (2.3).

For these simulations we use the sky location of pulsars in the NANOGrav 15-year dataset, and simulate evenly sampled observations over 15 years with a 30 day cadence and with no noise. From the resulting $r_k(t_i)$ residuals at times t_i in pulsar k we calculate the correlation between each pulsar pair as:

$$\mathcal{C}_{kl} = \frac{1}{N_{\rm obs}} \sum_{i=1}^{N_{\rm obs}} r_k(t_i) r_l(t_i), \tag{2.10}$$

where $N_{\rm obs}$ is the number of observations per pulsar. We also introduce:

$$\hat{\mathcal{C}}_{kl} = \mathcal{C}_{kl} \left(\frac{1}{N_{\rm PSR}} \sum_{k=1}^{N_{\rm PSR}} \mathcal{C}_{kk} \right)^{-1}, \tag{2.11}$$

which normalizes $\mathcal{C}_{kl}$ with the auto-correlation terms, so that it follows the same convention as Eq. (2.8).

Figure 2.7 shows the mean and 1-σ region of $\hat{\mathcal{C}}_{kl}$ as a function of γ. The means follow the expected HD curve for all three simulations. We also show the theoretical variance for a purely stochastic background from Ref. [48], and for a single CW with pulsar terms from Appendix A based on results from Ref. [48]. We can see that our results are in good agreement with these predictions. In particular, the GWB curve follows the prediction for an isotropic Gaussian background perfectly. The Population simulation also follows the shape of the same theoretical curve, but with a slightly lower variance at all angles. This constant shift is most likely due to its different spectral properties, which influence the normalization between the mean and the variance of the correlations (cf. Eqs. (C16) and (C23) of Ref. [48]). This also seems to suggest that our realistic GWB behaves more like an idealized GWB than a single CW in this regard. The CW simulations also show some agreement with the corresponding theoretical predictions. In particular, these correctly predict that a single source has less variance than a GWB at large angles, but more at small angles. The deviation can be explained by the anisotropic distribution of pulsars on the sky, since they agree very well for a simulation with isotropically distributed pulsars (see Appendix A).

The 1-σ bounds we see in Fig. 2.7 include both pulsar and cosmic variance. To combat the pulsar variance one can take the mean correlation in a set of angular separation bins. In

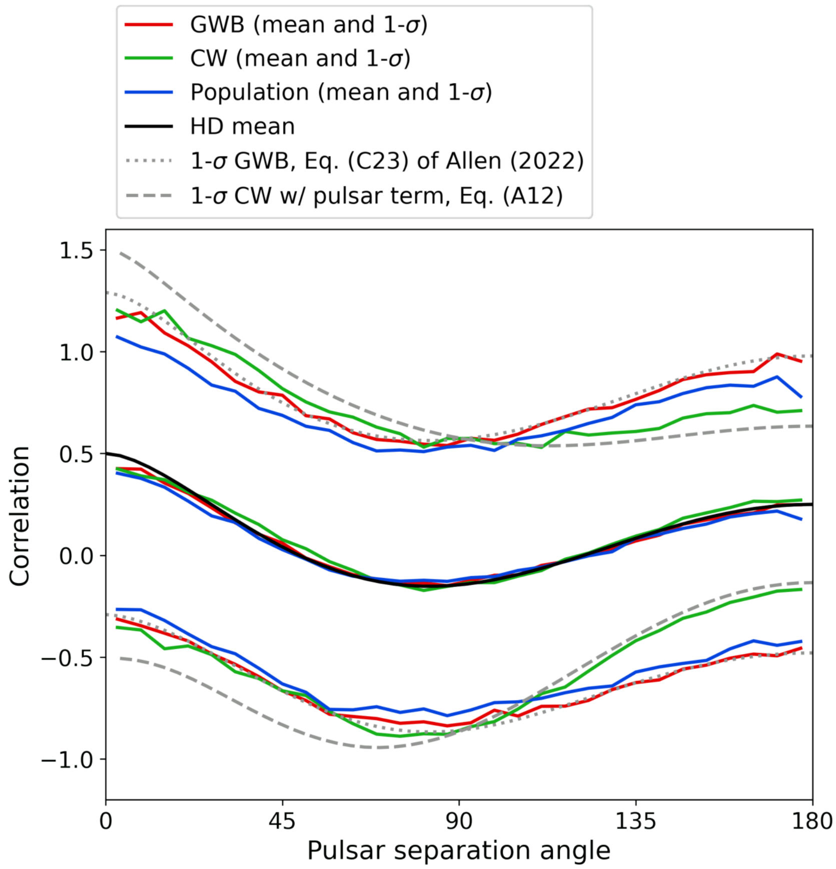


Figure 2.7: Variance of inter-pulsar correlations over 20 angular separation bins. We show results for an idealized GWB, a single CW source, and a realistic population-based GWB. We also show the theoretically expected mean and variance of the correlations under different assumptions. The small discrepancy between the theoretical and simulated CW curves is due to the anisotropic distribution of pulsars used for the simulation (see Appendix A).

the limit of large number of pulsars, such an averaging should factor out the pulsar variance, so the only source of variance is the cosmic variance (assuming noiseless observations). To test this scenario we analyzed the same simulations as above, but instead of calculating the variance of individual correlations, we first took the mean of the correlations in each realization, and then calculated the variance of those means over all realizations. The results are shown in Figure 2.8. We also show the same theoretically expected variances as in

Fig. 2.7. In addition, solid gray lines indicate the cosmic variance (Eq. (G10) in Ref. [48]). As expected, the variance of the correlations is drastically reduced by the binning and averaging procedure. However, the variance is still larger than the cosmic variance at some angular separations. This is due to the finite number of pulsar pairs available for averaging in our simulations. Also note that the results from our three simulations are all consistent with each other within their statistical uncertainty in this averaged representation.

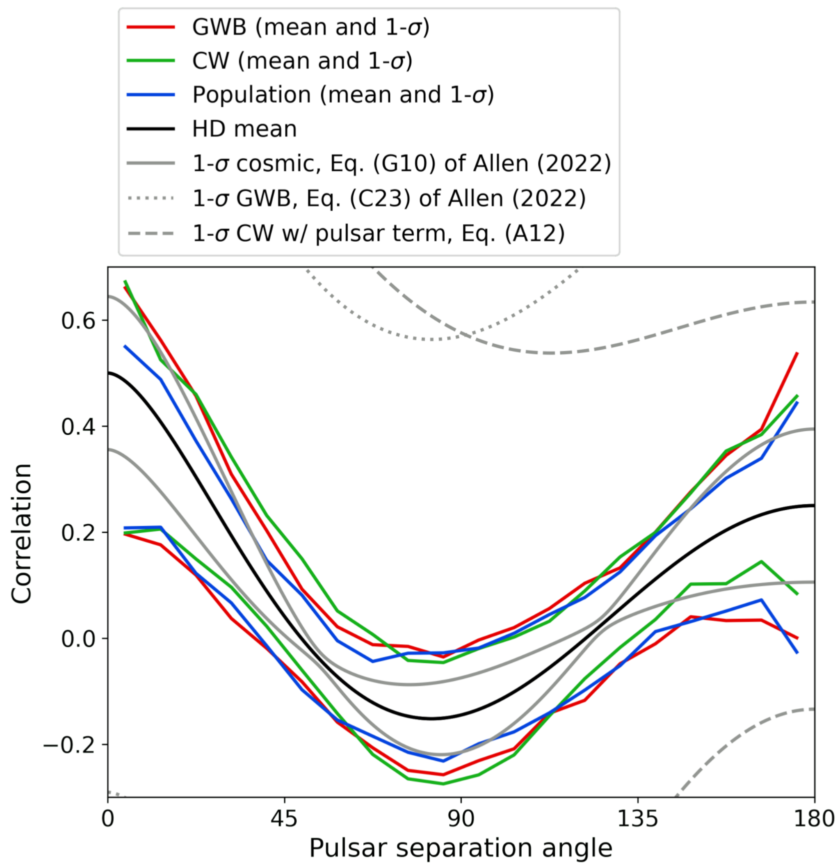


Figure 2.8: Variance of the mean correlations over 100 realizations. We show results for an idealized GWB, a single CW source, and a realistic population-based GWB. Besides the same theoretical curves as in Fig. 2.7, we also show the theoretical variance from cosmic variance only [48]. Note the different vertical scale compared to Fig. 2.7

Figures 2.7 and 2.8 only show the first and second moments of the distributions of correlations. However, one advantage of our direct simulation approach over analytical calculations is that we can easily produce the full distribution of correlations. To that end, we show the distribution of the deviation from the HD curve both with and without binning in Figure 2.9. We can see that the binned correlations show a consistent distribution for all three of our simulations. This is not surprising given that their means and variances are indistinguishable, as we have seen in Fig. 2.8. As expected, the individual correlations show a significantly wider distribution. We can also see that while the population-based and idealized GWB simulations show almost identical distributions, the histogram is significantly different for the simulations with an individual CW signal. As it was previously pointed out in Ref. [47], the overall distribution of correlation deviations has a heavier tail for a stochastic background then for a single CW source. Also note that the distribution of individual correlations shows significant non-Gaussian features.

2.6 Prospects for distinguishing individual binaries

Our method of modeling realistic backgrounds is also uniquely suitable to study detection prospects of individual binaries in the presence of a stochastic background. By setting $M = 1$ in Eqs. (2.3) and (2.4) we can separate the single brightest binary and the rest of the GWB in each bin. Thus we can take into account the confusion noise coming from the GWB itself, along with the white and red noise in each pulsar. Then we can calculate signal-to-noise ratios (SNRs) for those outlier sources as (see e.g. Eq. (219) of Ref. [49]):

$$\mathrm{SNR} = (s(t;\boldsymbol{\theta})|s(t;\boldsymbol{\theta})), \tag{2.12}$$

where $s(t;\boldsymbol{\theta})$ is the CW signal in question, and we define the inner product on residuals as:

$$(a|b) = a^T C^{-1} b, \tag{2.13}$$

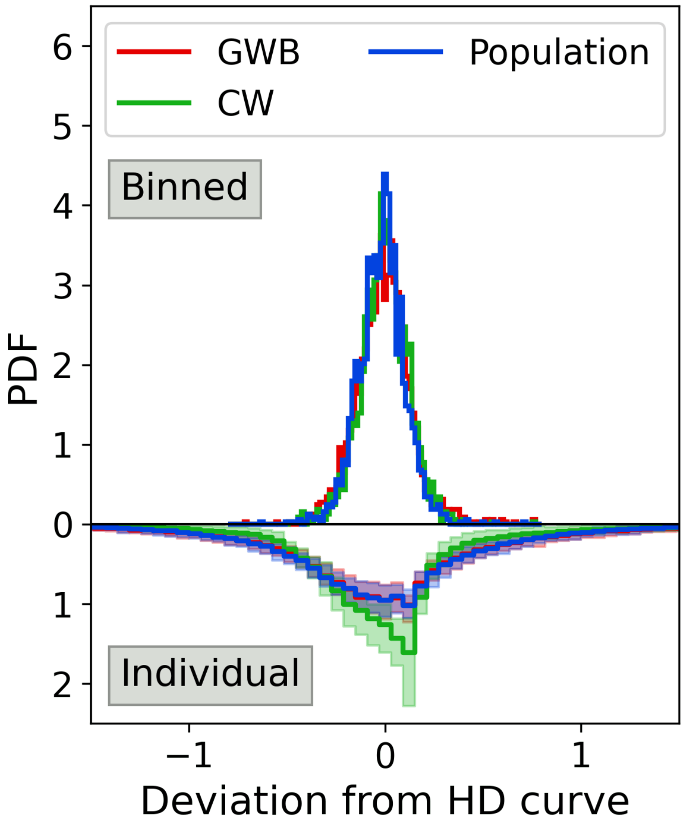


Figure 2.9: Histogram of deviations from the HD curve for all three simulations we performed. The bottom panel shows the mean and 90% credible interval for the distribution of all individual correlation values. The top panel shows the distribution of mean correlation values calculated in each angular bin and each realization.

where C is the noise covariance matrix that takes into account white and red noise in each pulsar and a stochastic background based on Eq. (2.4) with $M = 1$. We calculate inner products using the `ENTERPRISE` software package [50].

2.6.1 Variance of SNR over extrinsic parameters

The SNR can have a significant variance even when we fix the intrinsic parameters of the source. These are due to effects of extrinsic parameters: inclination angle, polarization angle, initial phase, and two angles describing the sky location of the source. To explore the effect of these on SNR, we focus on a particular binary with fixed amplitude/distance, chirp mass, and GW frequency, and calculate the distribution of SNRs over extrinsic parameters. Figure 2.10 shows this binary on the h_c–f_{obs} plane (orange dot). We also show the GWB spectrum for this particular realization (blue histogram), and the single brightest binary in each frequency bin (blue dots). Note that we chose our binary to be the one that produces the highest SNRs in this realization. We use a simulated dataset based on the properties of pulsars in the NANOGrav 12.5-year dataset [51], and we extend the observation time to 15 years, as it was done in Ref. [12].

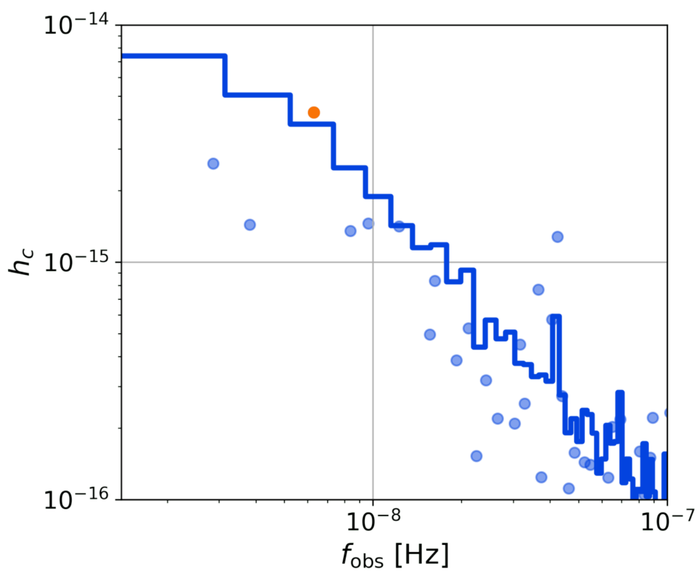


Figure 2.10: GWB spectrum in a particular realization (blue histogram) and the brightest binary in each bin (blue dots). The binary we focus our attention on is indicated by the orange dot.

Figure 2.11 shows the distribution of the median SNR over the sky for this particular source. We also show the location of the pulsars in our simulated array as red stars. The sizes of the stars are scaled by a rough estimate of the sensitivity of the pulsars defined as:

$$\text{Sensitivity} \propto \left(\sum_{i=1}^{N_{\rm obs}} \frac{1}{(\Delta t_i)^2} \right)^{1/2} , \tag{2.14}$$

where Δt_i is the nominal TOA error of the ith observation. Note that this simple expression reproduces the expected scaling both with the number of observations ($\sim \sqrt{N_{\rm obs}}$) and with the TOA errors ($\sim 1/\Delta t_i$). We can see in Fig. 2.11 that the sensitivity on the sky shows a dipolar structure, where one gets SNRs almost a factor of four higher in one direction than the antipodal direction. This is in agreement with the highly anisotropic upper limits found by the search for individual binaries in the NANOGrav 11-year dataset [31]. The reason for this is the highly anisotropic distribution of NANOGrav pulsars, which tend to be concentrated around the galactic center.

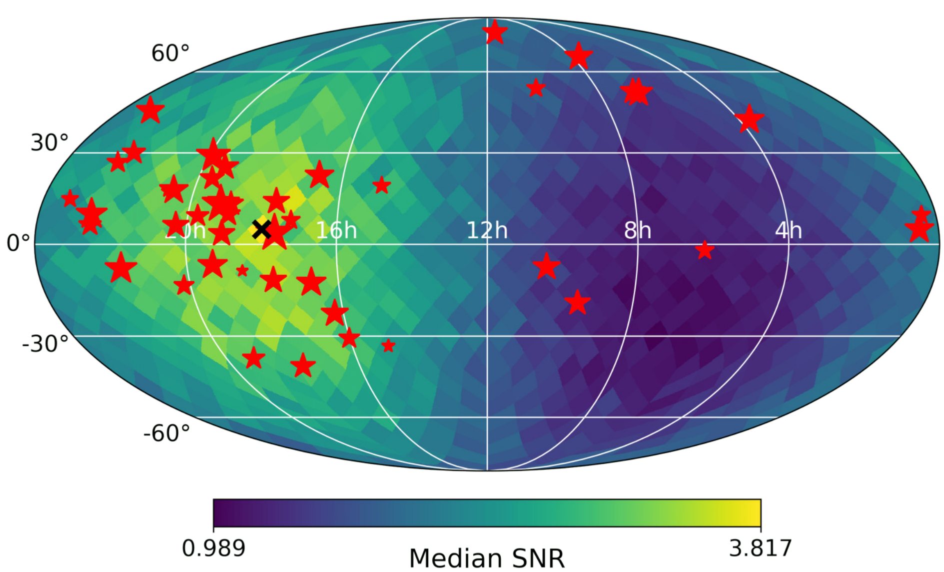


Figure 2.11: Distribution of the median SNR over the sky for a particular source. Red stars indicate the locations of pulsars in our array (based on the NANOGrav 12.5-year dataset, see Ref. [51]), with their sizes corresponding to their sensitivity as defined in Eq. (2.14). The black cross indicates the sky pixel with the highest median SNR.

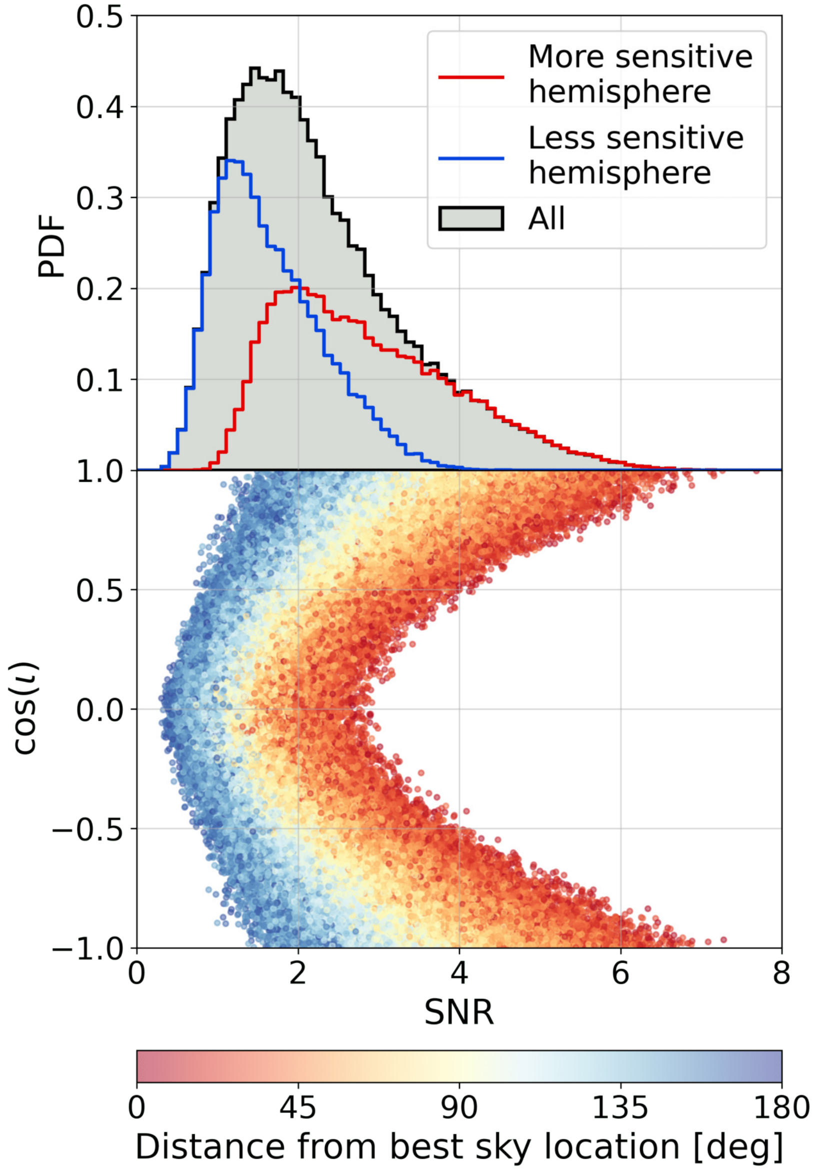


Figure 2.12: Distribution of SNRs for a particular source. The top panel shows the overall distribution (gray), along with histograms restricted to the half of the sky centered on the most sensitive sky location (red), and the antipodal point (blue). The bottom panel shows how SNR correlates with the orbital inclination (ι). The coloring indicates each source's angular distance from the most sensitive sky location (see black cross in Fig. 2.11).

The top panel of Figure 2.12 shows the distribution of SNRs for this particular source, marginalized over all external parameters. There is a significant variance in the SNR, with the 95% credible interval ranging from 0.8 to 4.9. We also show the distribution restricted to the more and less sensitive hemisphere, which we define relative to the most sensitive sky pixel marked with a black cross in Fig. 2.11. This shows that the sky location significantly contributes to the variance of the overall distribution. While most other extrinsic parameters do not show a clear correlation with SNR, the well-known effect of the inclination angle (ι) appears here as well. This is evident from the strong dependence of the GW amplitude on ι (see e.g. Eq. (13-14) of Ref. [31]). The bottom panel of Figure 2.12 shows the 2-dimensional distribution of $\cos\iota$ and SNR. Sources with face-on orientation ($\cos\iota = \pm 1$) produce significantly higher SNR values compared to edge-on systems ($\cos\iota = 0$). We also color-code the points on this panel with the source's angular distance from the most sensitive sky location marked in Fig. 2.11. We can see that orbital inclination and sky location represent the majority of the SNR variance. This means that PTAs are more likely to first see an individual SMBHB with approximately face-on orientation, located in the part of the sky where our PTA is most sensitive towards.

2.6.2 Expected properties of the highest-SNR SMBHB

We are interested in the expected properties of the CW signal that will first be detectable. To find those, we create 15 thousand realizations of a realistic GWB made up from 150 realizations of the SMBHB population with 100 different random extrinsic parameters each. We use a simulated PTA based on the NANOGrav 12.5-year dataset extended to a 15 year timespan [12]. Thus the following results can be treated as pessimistic predictions for the upcoming NANOGrav 15-year dataset, since we do not take into account the improved timing precision and the addition of new pulsars. For each realization we find the individual binary with the highest SNR looking through all the frequency bins. Figure

2.13 shows the location of these best sources on the h_c–$f_{\rm obs}$ plane, color-coded with their corresponding SNR. We also show the spectrum of the GWB for all realizations. We can see that the best sources are concentrated at moderate frequencies. The lack of high-SNR sources at higher frequencies is due to the white noise in the dataset, which has $h_c \sim f_{\rm obs}^{3/2}$. Note that a large fraction of the best sources lie below the characteristic strain spectrum of the GWB, which act as a noise source here. This is possible due to the fact that the SNR not only depends on h_c, but also on extrinsic parameters as we have seen in Section 2.6.1. Selecting the best source over all frequencies preferentially selects sources with the most favorable sky locations, inclinations, etc, thus resulting in significant SNRs even if the source has $h_c < h_c^{\rm GWB}$.

Figure 2.14 shows the 1 and 2-dimensional marginal distributions of the detector-frame GW frequency and chirp mass, luminosity distance, inclination angle, and the SNR for the best sources over 15 thousand realizations. The median $f_{\rm obs}$ is ~ 6 nHz, the median chirp mass is $\sim 5 \times 10^9\ M_\odot$, and the median luminosity distance is ~ 1.5 Gpc. This suggests that we will most likely first see a CW source at moderate frequencies with a very high chirp mass at a considerable distance. This is different from the common expectation of first seeing a nearby source (see e.g. Ref. [43]). Note however that this conclusion might be affected by the fact that our model does not include the local clustering of galaxies, and assumes a uniform-in-volume distribution instead. The distribution of $\cos\iota$ is significantly different from the flat distribution corresponding to isotropic inclination distribution. This is due to the increased SNR for face-on systems (see Fig. 2.12). The probability of the best source being within 30° of face-on is $\sim 26\%$ for the whole population and $\sim 36\%$ for the SNR > 5 subpopulation (compare with $\sim 13\%$ for the isotropic distribution).

We can also see on Figure 2.14 that the median SNR is 2.0, which is not expected to be detectable. However, there is about a 8% chance to get a source with SNR > 5. Depending on the details of the detection algorithm used, those might be detectable, which

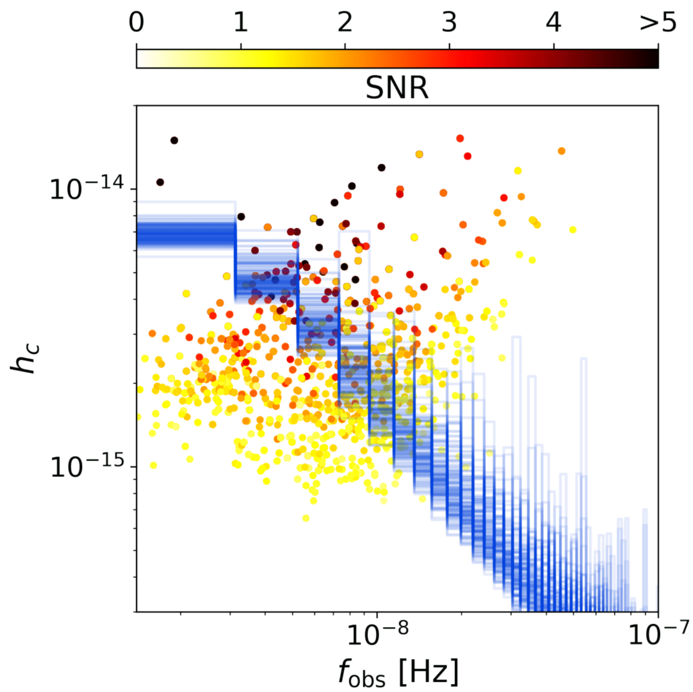


Figure 2.13: Location of the best sources on the h_c–$f_{\rm obs}$ plane over 150 realizations of the binary population with 100 random extrinsic parameters each (15 thousand in total). The color of each dot indicates the SNR of the best source in the particular realization. Note that the color scale was capped at SNR=5 for better visibility. We also show the GWB spectra for each realization in blue. Sources can have significant SNRs even if they are below the characteristic strain spectrum for the GWB if they have favorable extrinsic parameters (e.g. sky locations and inclinations).

is an interesting prospect for the search for individual binaries in the upcoming NANOGrav 15-year dataset. We also show histograms for these high-SNR sources in red. We can see that they largely follow the same distribution for $f_{\rm obs}$, $\mathcal{M}$, and $d_{\rm L}$ as all realizations. The same analysis extended to a 20 year timespan yields similar $f_{\rm obs}$, $\mathcal{M}$, and $d_{\rm L}$ values, a median SNR of 4.0, and SNR > 5 in about 32% of the realizations.

Note that another reason why these predictions are pessimistic is that we only check the binary with the highest h_c in each bin. In principle, binaries with lower h_c can end up producing the highest SNR signals. This means that by not taking those into account, our

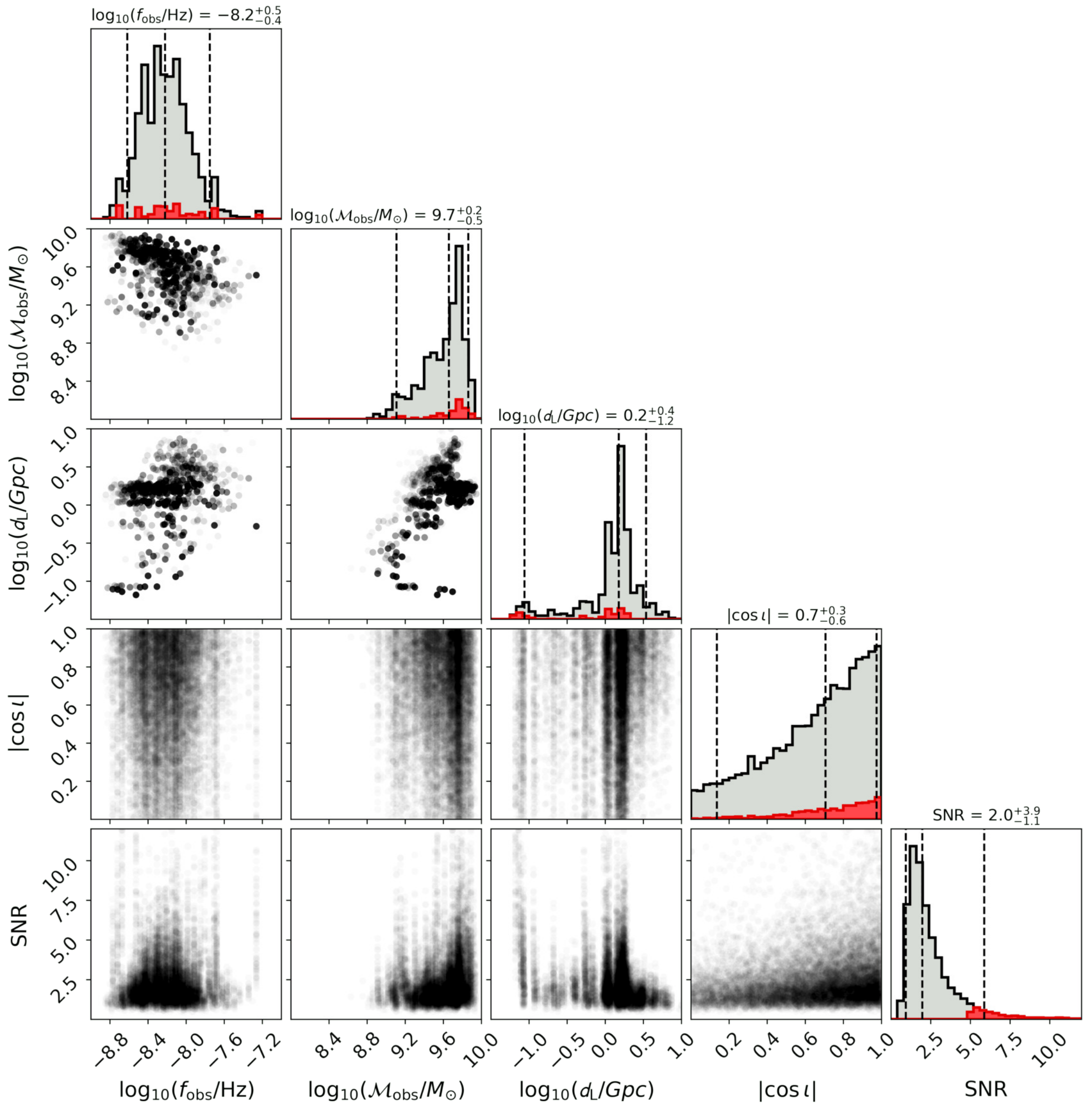


Figure 2.14: Distribution of the detector-frame GW frequency and chirp mass, luminosity distance, inclination angle, and the SNR of the best source over 150 binary population realizations with 100 realization each (15 thousand in total). Dashed lines and quoted values correspond to the 5th/50th/95th percentiles. Gray histograms show all realizations, while red histograms show only those where the best SNR is higher than 5 ($\sim 8\%$ of all realizations).

SNR distributions are biased low. On the other hand, these results do not marginalize over the uncertainty in the astrophysical models used for the SMBHB populations, which could have a large effect on these results. We will incorporate these additional details in a future study to provide more robust predictions for future PTA detection prospects.

2.7 Conclusion and future work

In this paper we presented a new approach to efficiently simulating realistic PTA datasets by modeling the signal as a combination of an isotropic Gaussian GWB and a few of the brightest binaries modeled individually. This produces datasets consistent with the naive method of directly modeling all binaries in a fraction of the time. We used this simulation technique to explore various properties of realistic PTA datasets. We show that the datasets are dominated by a small number of binaries at high frequencies, resulting in the well-known lack of GW power compared to a simple power-law model (see e.g. Ref. [34]). We test these datasets for statistical isotropy and find that they can be made isotropic by removing the couple brightest binaries at all except the highest frequencies. We calculate the mean and the variance of the spatial correlations in our simulated datasets and find good agreement with analytical results presented in Ref. [48].

Our methodology also allows for calculating SNRs of any binary in our datasets. We use that to calculate the distribution of the SNR of the brightest source over realizations for a simulated PTA based on the NANOGrav 12.5-year pulsars with time spans extended to 15 years. These calculations account for both pulsar noise and the confusion noise from the GWB. We find that the brightest binary tends to have moderate GW frequency, high chirp mass, large distance, and nearly face-on orientation. The median SNR of the brightest source is 2.0, and about 8% of the realizations produce a source with SNR higher than 5. These SNR values are pessimistic estimates for the upcoming NANOGrav 15-year dataset, since they do not take into account any potential improvements of the timing model solutions or the addition of new pulsars relative to the 12.5-year dataset. However, they rely on a particular model of the SMBHB population, which might introduce biases. If we further increase the observing timespan to 20 years, the fraction of realizations producing a binary with SNR greater than 5 increases to about 32%, and the median SNR increases to 4.0. We also calculate the SNR distribution for a given binary over different values of their extrinsic parameters (sky location, inclination and polarization angle, initial phase). We find that these parameters can have a significant effect on the SNR. The sky location and the inclination angle are particularly impactful parameters, as they can change the SNR by a factor of $2-4$.

The methods presented in this paper will continue to be useful tools to understand the interplay between the stochastic background and individual sources. In the future, we plan to incorporate different models of the SMBHB population, allowing us to produce simulated datasets using different assumptions about SMBHB formation and evolution. The software used for this paper was made to be compatible with the `holodeck` software package, allowing us to use any SMBHB population model developed there. That will also allow us to make our detection prospect predictions marginalized over astrophysical uncertainties. In addition, we also plan to run PTA detection pipelines on realistic datasets produced by the methods presented here. In particular, analyzing such datasets with the `QuickCW`[5] pipeline [52], would allow us to make more accurate predictions than the simple SNR calculations presented in this paper. We also plan to analyze such datasets with `BayesHopper` to investigate how PTAs will be able to detecting multiple individual sources in the future [32].

[5]Publicly available at: `https://github.com/bencebecsy/QuickCW`

ACKNOWLEDGEMENTS

The authors thank Stephen Taylor and Nihan Pol for providing access to software used in Ref. [12] to create simulated datasets based on measured noise properties of NANOGrav pulsars. We thank Chiara Mingarelli for feedback on the manuscript. We also thank Bruce Allen and Joseph Romano for useful discussions about the variance of HD correlations. We appreciate the support of the NSF Physics Frontiers Center Award PFC-1430284.

Software: `healpy` [53], `HEALPix` [42], `ENTERPRISE` [50], `corner` [54], `Numba` [55, 56]

CHAPTER THREE

FAST BAYESIAN ANALYSIS OF INDIVIDUAL BINARIES IN PULSAR TIMING ARRAY DATA

3.1 Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Bence Bécsy

Contributions: Developed the `QuickCW` software presented in the paper. Introduced a novel sampling scheme. Performed the analysis. Wrote the first draft of the manuscript based on a document detailing the methodology.

Author: Neil J. Cornish

Contributions: Conceived the algorithmic idea behind the method. Provided guidance on the analysis. Drafted a document describing the method, which was used as a basis for the manuscript. Edited the manuscript.

Author: Matthew C. Digman

Contributions: Contributed to developing the `QuickCW` software package described in the paper, including significantly increasing computational efficiency. Provided feedback on the analysis and the paper draft.

3.2 Manuscript Information

Bence Bécsy, Neil J. Cornish, Matthew C. Digman

Physical Review D

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
_X_ Published in a peer-reviewed journal

American Physical Society
Published 17 June 2022
Volume 105, Issue 12
10.1103/PhysRevD.105.122003

3.3 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. [15]).

All three major PTA experiments recently detected a common red noise process: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav [7]); the European Pulsar Timing Array (EPTA [8]); and the Parkes Pulsar Timing Array (PPTA [9]). The International Pulsar Timing Array (IPTA) also found strong support for such a signal in the combined dataset from these regional PTAs [10]. A common red noise process is expected to be the first sign of the GWB [57, 12]. 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 [12].

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. [58, 59]). 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 [60]). An individually resolved SMBHB would also provide a unique opportunity for joint electromagnetic and GW observations (see e.g. [61]). Several searches for individual SMBHBs have been carried out in the past, resulting in upper limits on the amplitude of GWs from SMBHBs [62, 63, 64].

Numerous analysis techniques have been developed to search for and characterize GWs from individual SMBHBs (see e.g. [65, 66, 67, 68]). 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 (non-stationary) noise rule out the fast Fourier domain methods used in the analysis of ground-based GW detectors (see [69] 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 [70, 71] or multiple binaries [32]. 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*, 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 [1]. Note that this approach is similar to the F-statistic method introduced in Ref. [72], 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. [68], where the likelihood function is either maximized or marginalized over the pulsar phase parameters.

The paper is organized as follows. In Section 3.4, 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 Section 3.5, 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 (Section 3.6) and the NANOGrav 11-year dataset [73, 62] (Section 3.7). We conclude and outline possible future directions in Section 3.8. Throughout this paper we use units where $G = c = 1$.

[1] Publicly available at: `https://github.com/bencebecsy/QuickCW`

3.4 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. [74, 75, 76, 77, 72]). 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 [2]:

$$h_{ab}(t,\hat{\Omega}) = e^{+}_{ab}(\hat{\Omega})\; h_{+}(t,\hat{\Omega}) + e^{\times}_{ab}(\hat{\Omega})\; h_{\times}(t,\hat{\Omega})\,, \tag{3.1}$$

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

$$e^{+}_{ab}(\hat{\Omega}) = \hat{m}_a\,\hat{m}_b - \hat{n}_a\,\hat{n}_b\,, \tag{3.2}$$

$$e^{\times}_{ab}(\hat{\Omega}) = \hat{m}_a\,\hat{n}_b + \hat{n}_a\,\hat{m}_b\,, \tag{3.3}$$

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

$$\hat{\Omega} = -\sin\theta\cos\phi\;\hat{x} - \sin\theta\sin\phi\;\hat{y} - \cos\theta\;\hat{z}\,, \tag{3.4}$$

$$\hat{m} = \sin\phi\;\hat{x} - \cos\phi\;\hat{y}\,, \tag{3.5}$$

$$\hat{n} = -\cos\theta\cos\phi\;\hat{x} - \cos\theta\sin\phi\;\hat{y} + \sin\theta\;\hat{z}\,. \tag{3.6}$$

The response of a pulsar to the source is described by the antenna pattern functions F^{+} and

[2]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

$F^{\times}$:

$$F^{+}(\hat{\Omega}) = \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}}\,, \tag{3.7}$$

$$F^{\times}(\hat{\Omega}) = \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})}{1+\hat{\Omega}\cdot\hat{p}}\,, \tag{3.8}$$

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:

$$s(t,\hat{\Omega}) = F^{+}(\hat{\Omega})\,\Delta s_{+}(t) + F^{\times}(\hat{\Omega})\,\Delta s_{\times}(t)\,, \tag{3.9}$$

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"),

$$\Delta s_{+,\times}(t) = s_{+,\times}(t_p) - s_{+,\times}(t)\,, \tag{3.10}$$

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:

$$t_p = t - L_p(1+\hat{\Omega}\cdot\hat{p})\,, \tag{3.11}$$

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:

$$\begin{aligned} s_{+}(t) &= \frac{\mathcal{M}^{5/3}}{d_L\,\omega(t)^{1/3}}\left[\sin 2\Phi(t)\,\left(1+\cos^2\iota\right)\,\cos 2\psi\right. \\ &\quad \left. +2\cos 2\Phi(t)\,\cos\iota\,\sin 2\psi\right]\,, \end{aligned} \tag{3.12}$$

$$\begin{aligned} s_{\times}(t) &= \frac{\mathcal{M}^{5/3}}{d_L\,\omega(t)^{1/3}}\left[-\sin 2\Phi(t)\,\left(1+\cos^2\iota\right)\,\sin 2\psi\right. \\ &\quad \left. +2\cos 2\Phi(t)\,\cos\iota\,\cos 2\psi\right]\,, \end{aligned} \tag{3.13}$$

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. (3.9)):

$$s_i(t,\hat{\Omega}) = \sum_{j=1}^{4} \sigma_{4(i-1)+j}(\hat{\Omega}, A_e, \iota, \psi, \Phi_0, \Phi_i) S^{4(i-1)+j}(t), \tag{3.14}$$

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

$$\begin{aligned}
S^{4(i-1)+1}(t) &= [\omega_0/\omega(t)]^{1/3} \cos 2\Phi(t)\,, \\
S^{4(i-1)+2}(t) &= [\omega_0/\omega(t)]^{1/3} \sin 2\Phi(t)\,, \\
S^{4(i-1)+3}(t) &= [\omega_i/\omega(t_{p_i})]^{1/3} \cos 2\Phi(t_{p_i})\,, \\
S^{4(i-1)+4}(t) &= [\omega_i/\omega(t_{p_i})]^{1/3} \sin 2\Phi(t_{p_i})\,,
\end{aligned} \tag{3.15}$$

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 timespan. 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:

$$\omega(t) = \omega_0 \left[1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}(t-t_0)\right]^{-3/8}. \tag{3.16}$$

The phases in eq. (3.15) are given by:

$$\begin{aligned}
\Phi(t) &= \Phi_0 + \frac{1}{32}\mathcal{M}^{-5/3}\left[\omega_0^{-5/3} - \omega(t)^{-5/3}\right], \\
\Phi(t_{p_i}) &= \Phi_i + \frac{1}{32}\mathcal{M}^{-5/3}\left[\omega_i^{-5/3} - \omega(t_{p_i})^{-5/3}\right]. \qquad (3.17)
\end{aligned}$$

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:

$$\begin{aligned}
\sigma_{4(i-1)+1} &= -A_{\rm e}\omega_0^{-1}\Big[\cos 2\Phi_0(1+\cos^2\iota)(\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) - \\
&\qquad 2\sin 2\Phi_0 \cos\iota(\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)\Big], \\
\sigma_{4(i-1)+2} &= -A_{\rm e}\omega_0^{-1}\Big[\sin 2\Phi_0(1+\cos^2\iota)(\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) + \\
&\qquad 2\cos 2\Phi_0 \cos\iota(\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)\Big], \\
\sigma_{4(i-1)+3} &= A_i\omega_0^{-1}\Big[\cos 2\Phi_i(1+\cos^2\iota)(\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) - \\
&\qquad 2\sin 2\Phi_i \cos\iota(\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)\Big], \\
\sigma_{4(i-1)+3} &= A_i\omega_0^{-1}\Big[\sin 2\Phi_i(1+\cos^2\iota)(\cos 2\psi F_i^+ - \sin 2\psi F_i^\times) + \\
&\qquad 2\cos 2\Phi_i \cos\iota(\sin 2\psi F_i^+ + \cos 2\psi F_i^\times)\Big], \qquad (3.18)
\end{aligned}$$

where $A_{\rm e} = \mathcal{M}^{5/3}d_L^{-1}\omega_0^{2/3}$ 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 $A_{\rm e}$, ω_0, and ω_i. The log likelihood can be written as:

$$\log L = -\frac{1}{2}(\delta t - s|\delta t - s) - \frac{1}{2}\log\det(2\pi C)\,, \qquad (3.19)$$

where:

$$(a|b) = a^T C^{-1} b\,. \tag{3.20}$$

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. [15]). Using the Woodbury matrix identity, we can express the inverse of C as:

$$C^{-1} = (N + TBT^T)^{-1} = N^{-1} - N^{-1}T\Sigma^{-1}T^T N^{-1}\,, \tag{3.21}$$

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

$$\begin{aligned} \log L \;\; &= \;\; -\frac{1}{2}(\delta t|\delta t) - \frac{1}{2}\log\det(2\pi C) \\ &\quad + \sum_{k=1}^{4N_{\rm p}} \sigma_k N^k - \frac{1}{2}\sum_{k=1}^{4N_{\rm p}}\sum_{l=1}^{4N_{\rm p}} \sigma_k \sigma_l M^{kl}\,, \end{aligned} \tag{3.22}$$

where $N^k = (\delta t|S^k)$, $M^{kl} = (S^k|S^l)$, and $N_{\rm p}$ 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 3.1) 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 un-evenly sampled and the orbital periods are generally not integer sub multiples of the observation time. The band-diagonal structure of the M^{kl} mean that it can computed and stored in an array of size $10N_{\rm p}$, rather than the naive $(4N_{\rm p})^2$. 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 B 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.

Table 3.1: List of shape and projection parameters and their numbers as a function of the number of pulsars (N_p).

Shape parameters	Projection parameters
$(4 + 3 \times N_\mathrm{p})$	$(4 + N_\mathrm{p})$
θ, ϕ, f_GW, $\mathcal{M}$, L_{p_i}, γ_{p_i}, A_{RN,p_i}	ι, A_e, Φ_0, Ψ, Φ_i

For a fixed set of noise parameters and shape parameters, we can compute the likelihood for *any* set of projection parameters (see Table 3.1) in essentially zero time. See details of how we can take full advantage of the speedup by an implementation relying on `Numba` in Appendix C. 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. (3.11) depends on the sky location.

We can see from eq. (3.16) that the initial angular frequency at each pulsar is given by:

$$\omega_i = \omega_0 \left(1 + \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}L_{p_i}(1 + \hat{\Omega}\cdot\hat{p}_i)\right)^{-3/8} . \tag{3.23}$$

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

$$\begin{aligned} \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}L_{p_i} &= 10.16\left(\frac{\mathcal{M}}{10^9\,M_\odot}\right)^{5/3} \\ &\quad \times \left(\frac{\omega_0}{2\pi/\mathrm{yr}}\right)^{8/3}\left(\frac{L_{p_i}}{1\,\mathrm{kpc}}\right) . \end{aligned} \tag{3.24}$$

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:

$$\frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}t = 0.031\left(\frac{\mathcal{M}}{10^9\,M_\odot}\right)^{5/3} \times\left(\frac{\omega_0}{2\pi/\text{yr}}\right)^{8/3}\left(\frac{t}{10\text{ yr}}\right). \tag{3.25}$$

3.5 Metropolis-within-Gibbs sampling

With the new formulation of the likelihood in eq. (3.22) 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 [78] 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, see e.g. [79, 80]). 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 $N \to \infty$, 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 [81, 82].

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 ith iteration as:

$$\boldsymbol{\theta}_i = \left\{\boldsymbol{\theta}_i^{(\mathrm{s})}, \boldsymbol{\theta}_i^{(\mathrm{p})}\right\}, \tag{3.26}$$

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

1. Draw a set of new shape parameters $\boldsymbol{\theta}^{(\mathrm{s})}$ from the proposal $q^{(\mathrm{s})}\left(\boldsymbol{\theta}^{(\mathrm{s})}\middle|\boldsymbol{\theta}_i^{(\mathrm{s})}\right)$.

2. Draw N different sets of projection parameters, $\boldsymbol{\theta}_{(1)}^{(\mathrm{p})}, ..., \boldsymbol{\theta}_{(k)}^{(\mathrm{p})}, ..., \boldsymbol{\theta}_{(N)}^{(\mathrm{p})}$, from the proposals $q_k^{(\mathrm{p})}\left(\boldsymbol{\theta}_{(k)}^{(\mathrm{p})}\middle|\boldsymbol{\theta}_i^{(\mathrm{p})}\right)$.

3. Randomly select $\boldsymbol{\theta}_{(j)} = \left\{\boldsymbol{\theta}^{(\mathrm{s})}, \boldsymbol{\theta}_{(j)}^{(\mathrm{p})}\right\}$, according to the probability mass function:

$$p(\boldsymbol{\theta}_{(j)}) = \frac{L(\boldsymbol{\theta}_{(j)})}{\sum_{k=1}^{N} L(\boldsymbol{\theta}_{(k)})}. \tag{3.27}$$

4. Form N auxiliary samples:

$$\boldsymbol{\Theta}_{(k)} = \begin{cases} \left\{\boldsymbol{\theta}_i^{(\mathrm{s})}, \boldsymbol{\theta}_{(k)}^{(\mathrm{p})}\right\} & \text{if } k \neq j \\ \boldsymbol{\theta}_i & \text{if } k = j. \end{cases} \tag{3.28}$$

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

$$\min\left(1, \frac{\sum_{k=1}^{N} L(\boldsymbol{\theta}_{(k)})}{\sum_{k=1}^{N} L(\boldsymbol{\Theta}_{(k)})}\right), \tag{3.29}$$

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

Note that eqs. (3.27) and (3.29) 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. (3.29) reduces to the usual Metropolis-Hastings acceptance probability. Also note that since we are selecting the proposed projection parameters, $\boldsymbol{\theta}_{(j)}^{(\mathrm{p})}$, according to eq. (3.27), as $N \to \infty$ we are drawing independent samples from the conditional likelihood at the new shape parameters, $L(\boldsymbol{\theta}^{(\mathrm{p})}_{(j)}|\boldsymbol{\theta}^{(\mathrm{s})})$. At this large N limit, we can also interpret the acceptance probability in eq. (3.29) as comparing the likelihood at the new and old shape parameters marginalized over the 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 C.

In general, MTMCMC algorithms provide better mixing at the cost of additional likelihood evaluations needed in eqs. (3.27) and (3.29). 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.

3.6 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 [3]. 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 3.2 and Section 3.6.1.1); a rapidly evolving signal where the Earth and pulsar terms have significantly different frequencies (see Dataset 2 in Table 3.2 and Section 3.6.1.2); 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 3.2 and Section 3.6.1.3). 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 Section 3.6.2).

3.6.1 Detection analyses

3.6.1.1 Slowly evolving signal We first analyzed a signal with a relatively low chirp mass and frequency (see Dataset 1 in Table 3.2) resulting in only slightly different Earth term and pulsar term frequencies (cf. eq.(3.24)). We chose the amplitude of the signal to achieve a moderate SNR of 10.4. Fig. 3.1 shows how the square of the SNR (a good proxy

[3] Note that d_L is not an independent parameter we fit for. It is completely determined by A_{e}, f_{GW}, and $\mathcal{M}$.

Table 3.2: 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 [3]. The GW phase at each pulsar was determined from Φ_0, the light travel time to each pulsar, and $f_{\rm GW}$.

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

for detectability) is distributed among the 45 pulsars in the array. The highly 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. 3.2 shows the one-dimensional and two-dimensional marginal distributions of the 8 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

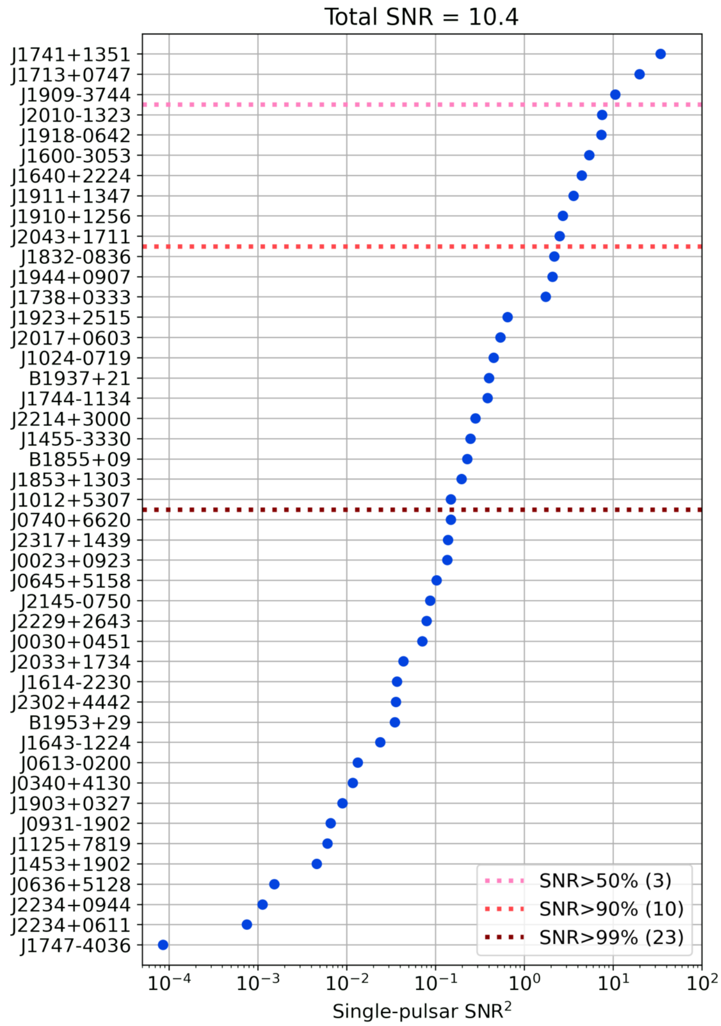


Figure 3.1: SNR^2 values in datastreams of each simulated pulsar for slowly a evolving signal (Dataset 1 in Table 3.2). Horizontal dashed lines indicate the minumum number of pulsars needed to reach 50/90/99% of the total SNR^2.

in a detectable frequency evolution. The amplitude is highly correlated with a number of nuisance parameters, most notably the inclination angle (ι), 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,

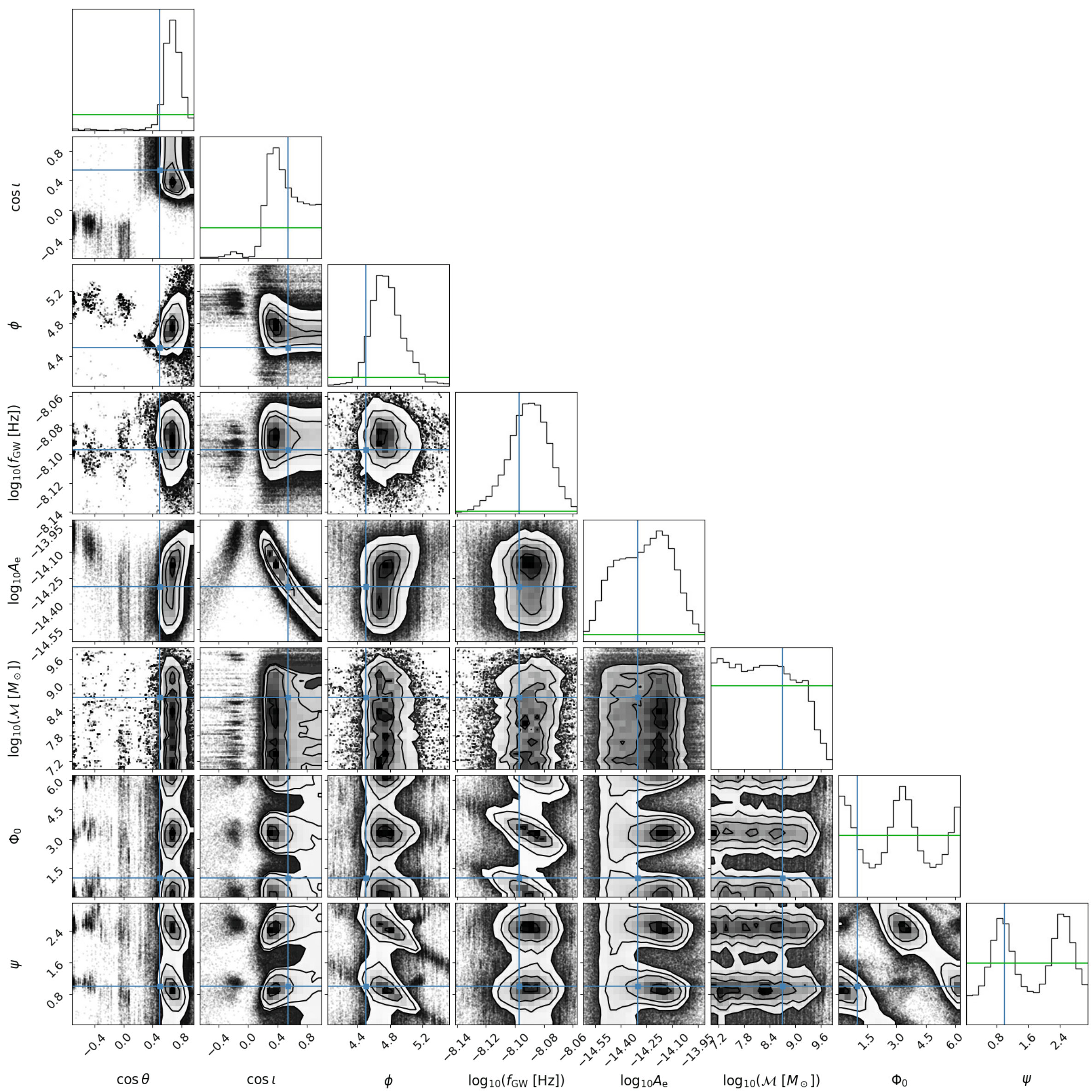


Figure 3.2: Corner plot of parameters common to all pulsars for the slowly evolving signal (SNR=10.4, Dataset 1 in Table 3.2). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the 1/2/3-σ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

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. Fig. 3.3 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.

3.6.1.2 Fast evolving signal The next signal we analyzed has higher frequency and chirp mass than the previous one (see Dataset 2 in Table 3.2), 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. (3.25)). 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 pulsars in Fig. 3.4. Unlike in the slowly evolving case, the observed frequency evolution

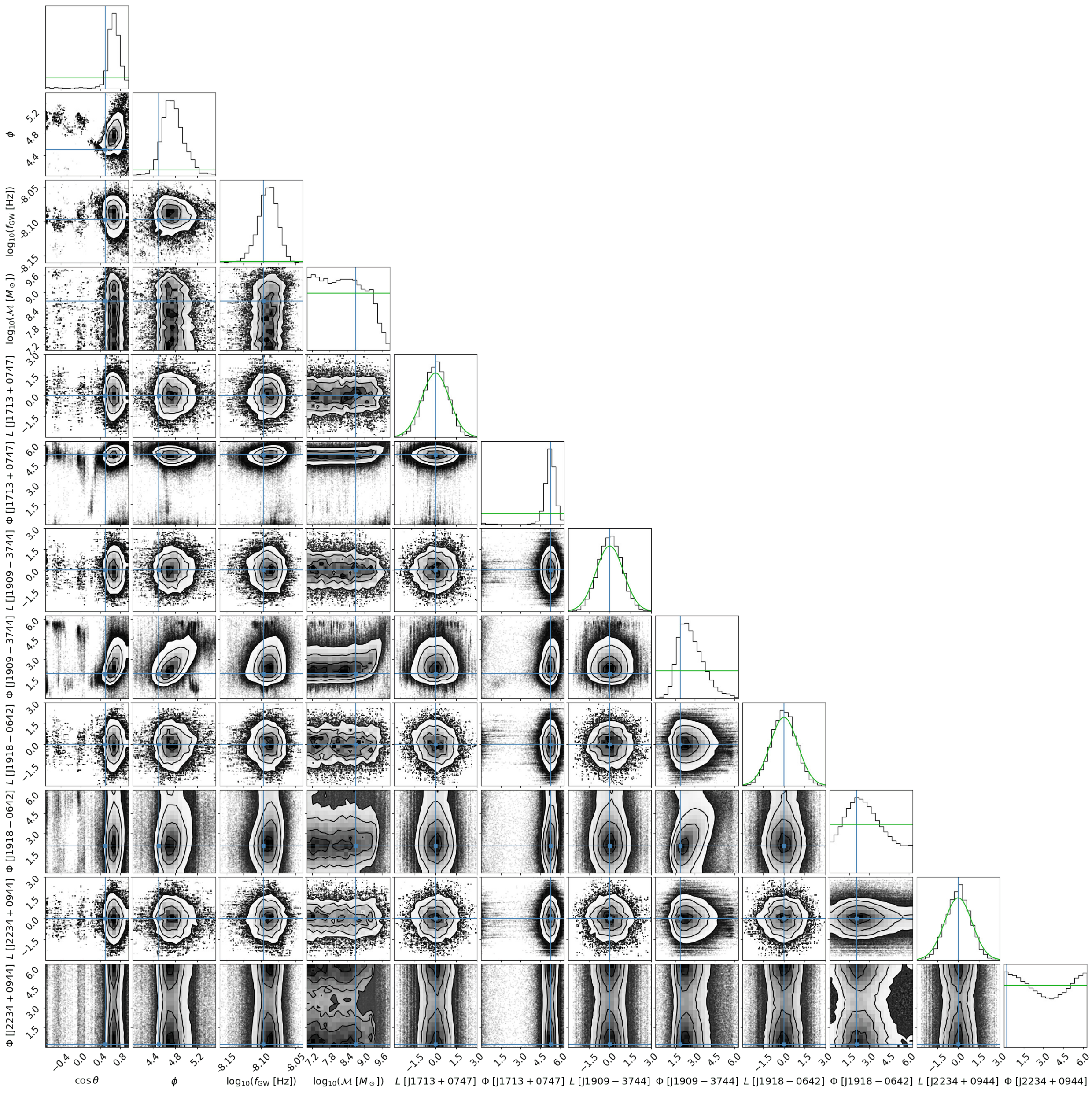


Figure 3.3: Corner plot of four common shape parameters and selected pulsar distances and phases for the slowly evolving signal (SNR=10.4, 1 in Table 3.2). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the 1/2/3-σ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

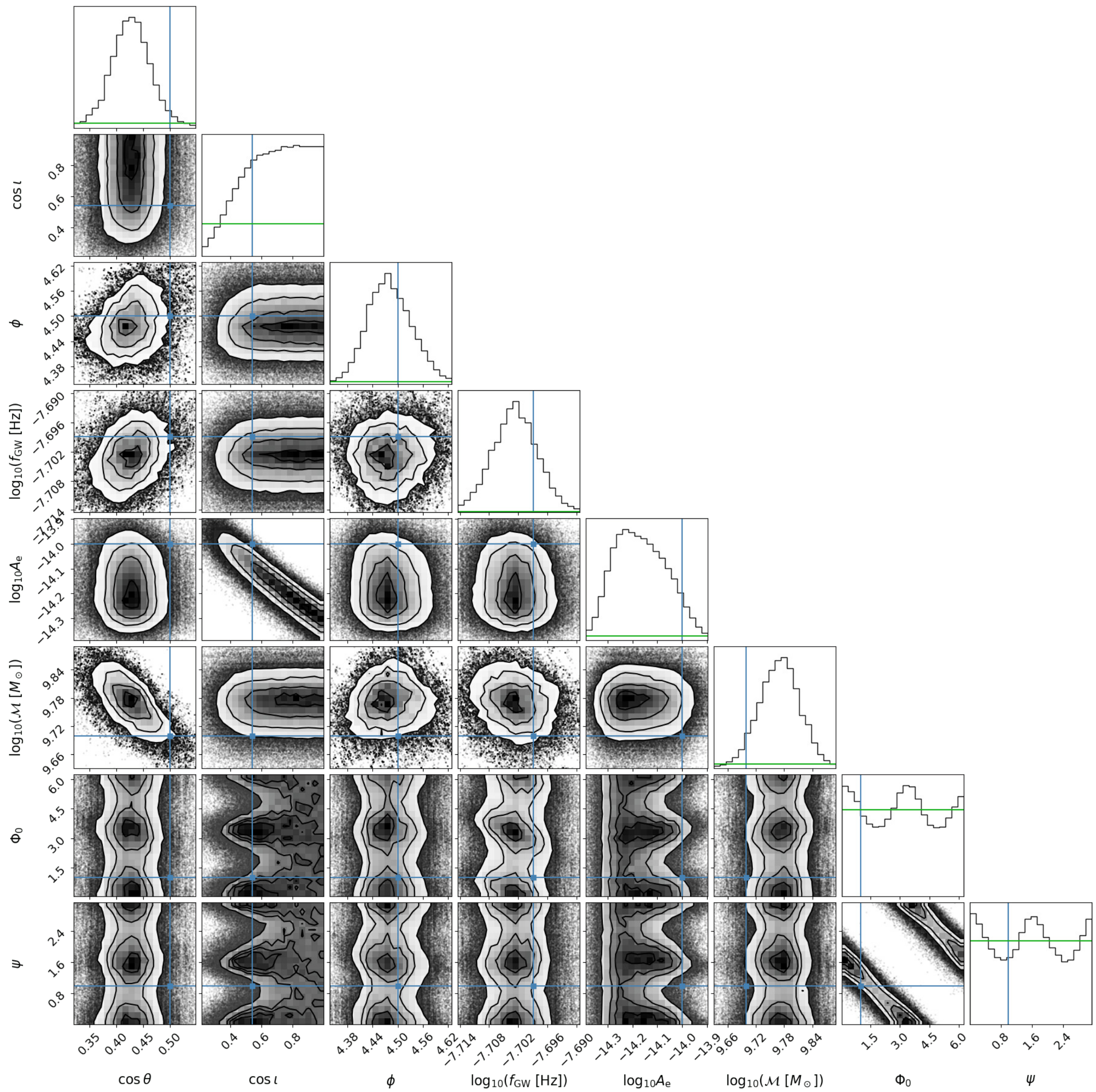


Figure 3.4: Corner plot of parameters common to all pulsars for the fast evolving signal (SNR=13, Dataset 2 in Table 3.2). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the 1/2/3-σ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

results in an informative chirp mass posterior. This moderate-SNR source results in a chirp mass measurement with a 1-σ 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. [83, 84, 85]). The intricate multimodal structure we have seen for the slowly evolving signal in Fig. 3.2 is much less prominent for this quickly evolving signal. Many of the one-dimensional marginal distributions we see in Fig. 3.4 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.

We show the posterior distributions of phases and distances for four selected pulsars, and their correlations with the four common shape parameters in Fig. 3.5. We show these for the same four pulsars as in Fig. 3.3. 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. 3.3, 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.

3.6.1.3 Marginally detectable signal We also analyzed a marginally detectable signal with a total SNR of 4.3, with 50/90/99% of the total SNR^2 coming from 3/12/22 pulsars (see Dataset 3 in Table 3.2). The sky location and frequency ($f_{\mathrm{GW}} = 8$ nHz) of the signal were chosen so that it roughly corresponds to the most sensitive part of the parameter space (see Section 3.6.2) 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 higher chirp mass systems have higher GW amplitudes, but

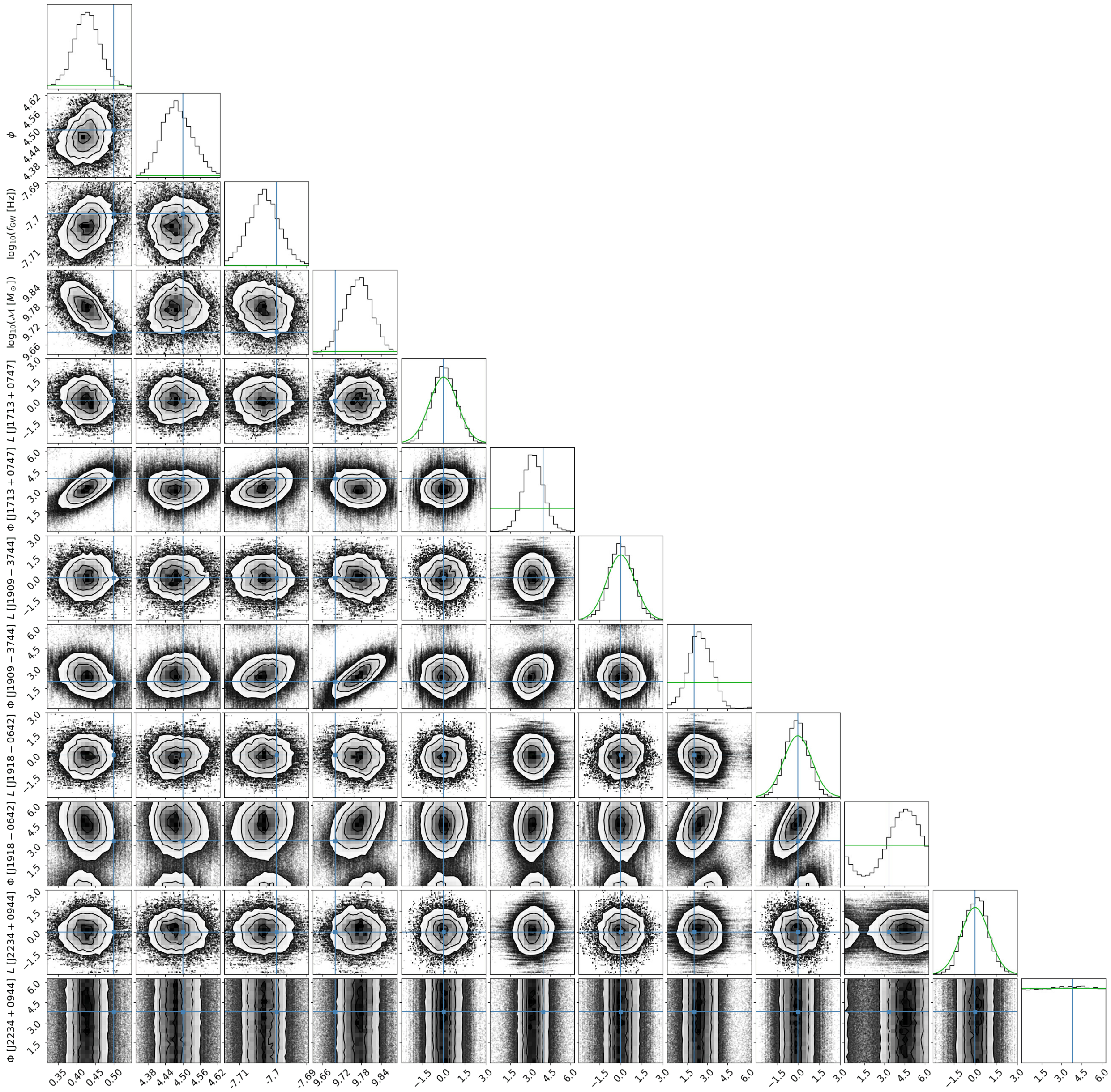


Figure 3.5: Corner plot of four common shape parameters and selected pulsar distances and phases for the fast evolving signal (SNR=13, Dataset 2 in Table 3.2). We show the posterior with black, the prior with green and the true values of parameters with blue. Contour lines represent the 1/2/3-σ levels in two dimensions, which correspond to 39.3%/67.5%/86.5% credible regions.

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

Fig. 3.6 shows the one-dimensional and 2-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 $f_{\rm GW}$ and $A_{\rm e}$, both of which have a posterior peaked at the true location of the parameter. The posterior for $f_{\rm GW}$ also has non-negligible support over the entire prior range. $A_{\rm e}$ 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 $f_{\rm GW}$ and $A_{\rm e}$, while basically recovering the priors of other parameters. These other parameters will start to have more informative posteriors as we gather more SNR (as we have seen on Figs. 3.2 and 3.4). 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 Section 3.6.2, we also investigate how the upper limit can be affected by a marginally detectable signal.

3.6.2 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.

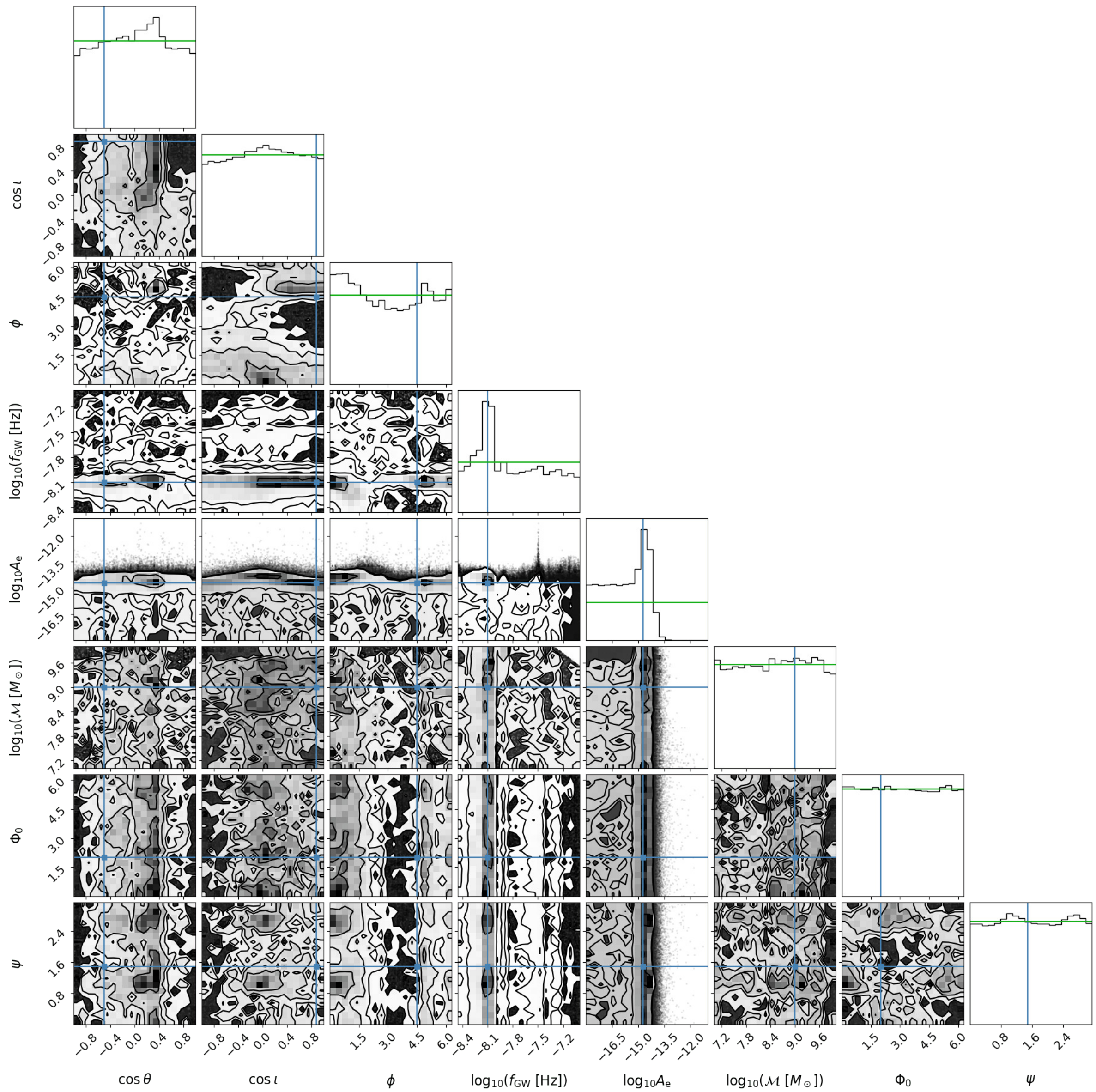


Figure 3.6: Corner plot of parameters common to all pulsars for marginally detectable signal (SNR=4.3, Dataset 3 in Table 3.2). We show the posterior with black, the prior with green and the true values of parameters with blue.

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. 3.7 shows the upper limit on $\log_{10} A_{\rm e}$ as a function of $\log_{10} f_{\rm GW}$ for this simulated dataset. We also show the number of independent samples in each bin on the $\log_{10} f_{\rm GW}$–$\log_{10} A_{\rm e}$ plane. The upper limit is calculated by binning the samples in $\log_{10} f_{\rm GW}$ and finding the 95th percentile of the amplitude in each bin. As expected, the dataset is most sensitive around 8 nHz ($\log_{10} f_{\rm GW} \simeq$ -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 $f_{\rm GW} = 1/\mathrm{yr}$ 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 $(\log_{10} f_{\rm GW}, \log_{10} A_{\rm e}) \simeq$ (-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. Fig. 3.8 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 [86, 87], and since it destroys the coherence of the signal, it could be used the build a null distribution. Comparing the non-scrambled 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

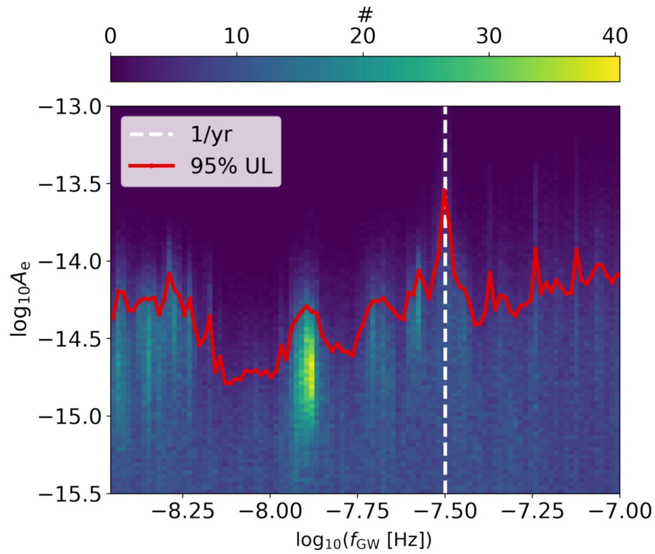


Figure 3.7: GW amplitude upper limit as a function of GW frequency (red line). We also show the number of independent samples on the $\log_{10} f_{\rm GW}$–$\log_{10} A_{\rm e}$ plane. The vertical dashed line indicates the frequency of 1/yr, where we expect a significantly reduced sensitivity.

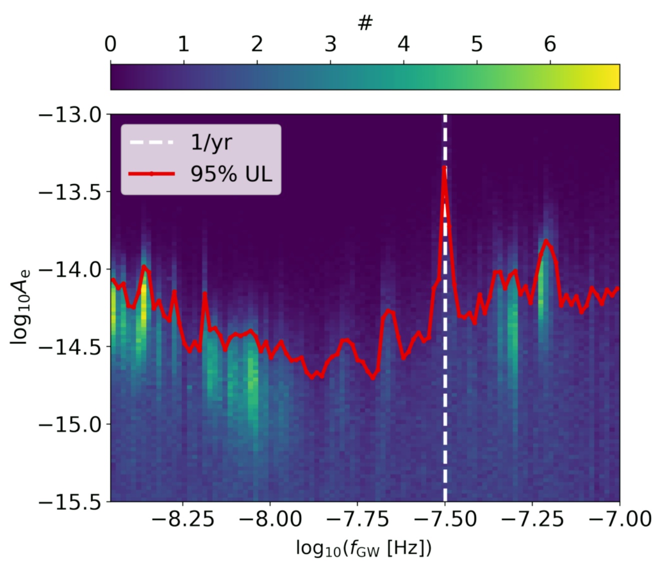


Figure 3.8: Same as Fig. 3.7, but with a different noise realization. Note that bumps in the upper limit due to noise fluctuations appear at different random frequencies.

candidate vetting.

We also investigated how the upper limit changes in the presence of a marginal GW signal. Fig. 3.9 shows the amplitude upper limit as a function of $\log_{10} f_{\rm GW}$ for Dataset 3 from Table 3.2. As a comparison, we also show the upper 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. 3.7 due to the noise fluctuation.

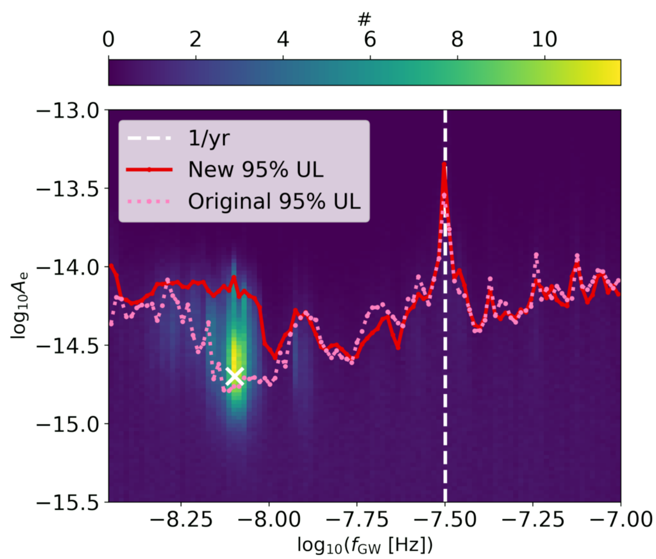


Figure 3.9: Same as Fig. 3.7 but in the presence of a marginally detectable signal at $\log_{10} f_{\rm GW} \simeq -8.1$. The red solid line shows the upper limit we get from analyzing this dataset, which is elevated around the signal compared to the original upper limit shown by the pink dotted line. The white cross indicates the true parameters of the signal.

Fig. 3.10 shows the frequency-marginalized upper limit we can place on $\log_{10} A_{\rm e}$ 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 non-isotropic distribution on the sky. Such anisotropy is expected for any pulsar timing array, because there are more pulsars towards the galactic center than in the antipodal direction.

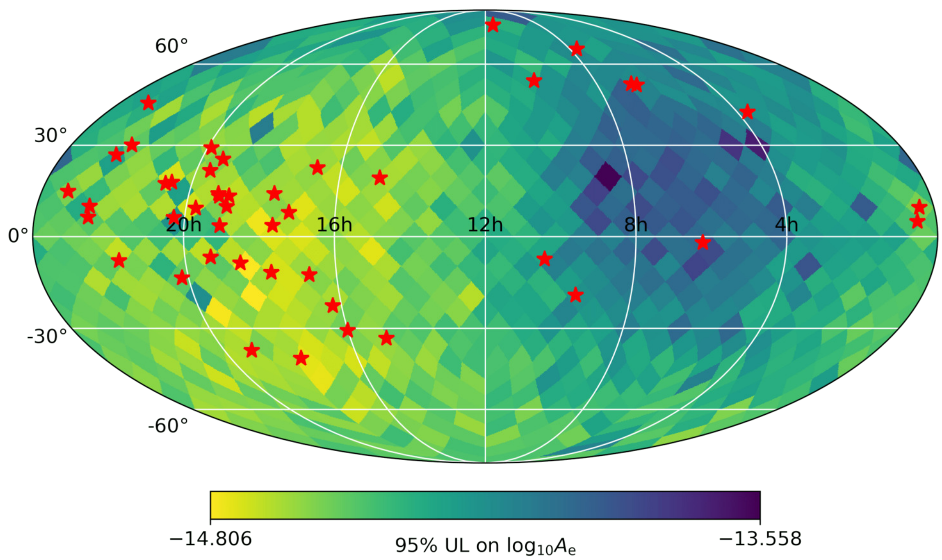


Figure 3.10: Upper limit on $\log_{10} A_\mathrm{e}$ as a function of sky location. Red stars indicate locations of the pulsars in the array.

3.7 Results with real data

To validate our new methods with real data, we analyzed the NANOGrav 11-year dataset. Fig. 3.11 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 [62]. 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 $f_{\rm GW} = 1/{\rm yr}$). 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 in the official analysis [62][4]. The bug resulted in a small overestimation of the red noise in some pulsars, which in turn meant that the upper limit on $A_{\rm e}$ was underestimated at low frequencies, where there is a strong correlation between the red noise and the individual binary models.

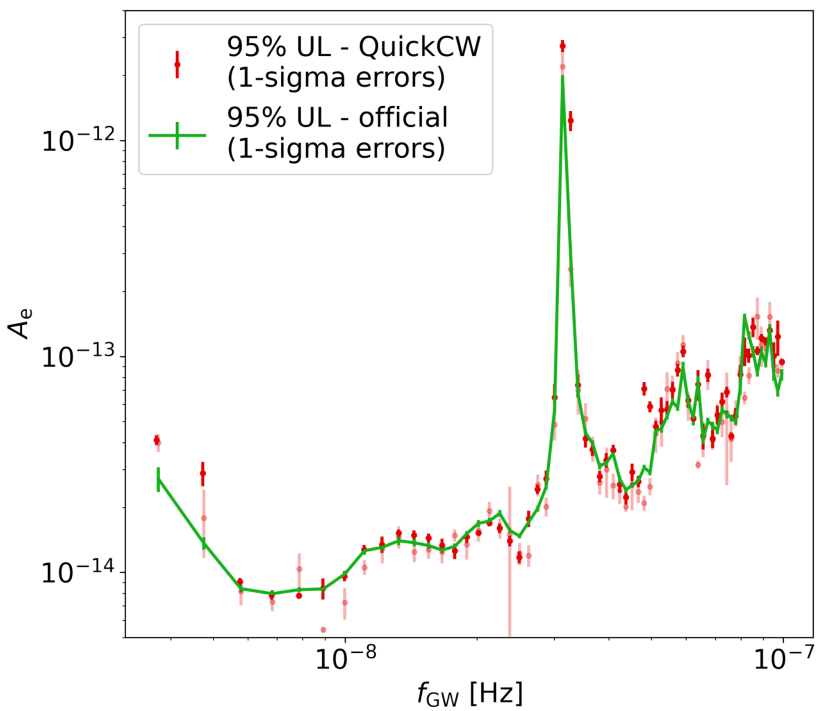


Figure 3.11: GW amplitude upper limit as a function of GW frequency for the NANOGrav 11-year dataset analyzed by `QuickCW` (red), and as reported in Ref. [62] (green). We also show our results using narrow frequency bins centered at the fixed frequencies used in the official NANOGrav analysis (pink).

[4]Note that this issue was recently resolved here: `https://github.com/nanograv/enterprise_extensions/commit/3a17943dddfa005867cdbaa960d4e72b747eb373`

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 sub-optimal, since if we require a fixed level of convrgence 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 pseudo-prior on $f_{\rm GW}$, which down-weights regions with many samples based on a pilot run. The pseudo-prior would then ensure a more uniform number of samples at all frequencies while preserving good mixing. The undesired bias from the presude-prior can be canceled in post processing by reweighting the posterior samples. This technique is similar to umbrella sampling [88], and the pseudo-prior used when calculating Bayes factors with the product space method [89].

3.8 Conclusion

In this paper, we presented a new formulation of the likelihood function (see eq. (3.22)), 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 [7, 8, 9, 10], 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. (3.20), 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 B 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` [32], 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 [90]. It might also be worthwhile to extend these methods to sine-Gaussian wavelets in an effort to speedup `BayesHopperBurst` [91], a pipeline to search for generic GW bursts in PTA data.

ACKNOWLEDGEMENTS

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` [54] software package to produce the corner plots presented in this paper. Some of the results in this paper have been derived using the `healpy` [53] and `HEALPix` [42] package. We appreciate the support of the NSF Physics Frontiers Center Award PFC-1430284 and NASA LISA foundation Science Grant 80NSSC19K0320.

CHAPTER FOUR

JOINT SEARCH FOR ISOLATED SOURCES AND AN UNRESOLVED CONFUSION BACKGROUND IN PULSAR TIMING ARRAY DATA

4.1 Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Bence Bécsy

Contributions: Developed the `BayesHopper` software presented in the paper. Performed the analysis. Wrote the first draft of the manuscript.

Author: Neil J. Cornish

Contributions: Conceived the study design. Provided guidance on the analysis. Supplied sample code to calculate Fisher proposals. Edited the manuscript.

4.2 Manuscript Information

Bence Bécsy, Neil J. Cornish

Classical and Quantum Gravity

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
__X__ Published in a peer-reviewed journal

IOP Publishing Ltd
Published 12 June 2020
Volume 37, Issue 13, id.135011
10.1088/1361-6382/ab8bbd

4.3 Introduction

Pulsar Timing Arrays (PTAs) are designed to detect gravitational waves (GWs) in the nanohertz regime by looking for signatures of a GW signal in the pulse arrival times of regularly observed millisecond pulsars (see e.g. [14]). Currently there are three major PTA projects: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav, [92]), the European Pulsar Timing Array (EPTA, [4]), and the Parkes Pulsar Timing Array (PPTA, [5]). These groups collaborate under the aegis of the International Pulsar Timing Array (IPTA, [6, 93]).

The dominant GW sources in the frequency band of PTAs are expected to be supermassive black-hole binaries (SMBHBs). The collection of these SMBHBs is expected to give rise to a stochastic gravitational wave background (GWB), which has been searched for in previous PTA datasets (see e.g. [94, 95]). The scarcity of SMBHBs in the local Universe can introduce anisotropy in the GWB [96, 97, 98, 99], which makes the standard isotropic searches sub-optimal, thus requiring changes to those standard algorithms [100, 101, 102].

In addition, some of the brightest sources might be individually detected as so called continuous waves (CWs), referring to the fact that the frequency of GWs from an SMBHB does not change appreciably during the observation time of PTAs. While it is more likely that PTAs will detect the GWB before detecting any CW signals [58], searches have been carried out resulting in upper limits on the possible amplitude of CWs (see e.g. [103, 104]).

In the context of the Laser Interferometer Space Antenna (LISA, [105]), it has been shown that removing the brightest sources results in a Gaussian background from the unresolved sources [106]. For PTAs, it has been shown that removing the brightest CW signal significantly reduces the anisotropy of the GWB [99]. Describing a signal as either a collection of individual components or as a stochastic background has been considered in the context of a toy model using a pair of co-located LIGO [107] detectors and a signal formed from multiple sinusoids [108]. The study considered three different models: i) a collection of sinusoidal GWs; ii) a Gaussian stochastic GWB; iii) the combination of i) and ii). It was found that model iii) was almost always preferred, despite the fact that the dataset was generated with the assumption of model i). This is because of the natural parsimony of Bayesian statistics preferring a model with fewer parameters even if it fits the data slightly worse. Non-Bayesian methods have been proposed to search for multiple CWs in PTA data in [109, 110].

These previous studies motivate a joint search for CWs and an isotropic GWB. Here we present a method for such a search using Bayesian inference and a trans-dimensional Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler [111]. RJMCMC methods are used in other GW data analysis problems as well, like the detection of unmodeled transient GWs in LIGO data by the BayesWave algorithm [112], and the detection of galactic binaries in LISA data [113, 114, 115, 116]. Such a sampler is able to explore a parameter space of varying dimension, allowing it to decide what to include in the model based on the data. In our case it can turn the GWB component on or off, and can also change how many CWs are in the model (including the possibility of zero CWs).

The paper is organized as follows: Section 4.4 introduces the models we use to describe our PTA dataset and how our RJMCMC sampler explores the whole parameter space including different model dimensions. Section 4.5 presents the results of analyzing various simulated datasets, which we carried out to assess the performance of our algorithm. We provide our conclusions and discuss future work in Section 4.6.

4.4 Methods

In this section, we discuss our signal and noise model, and introduce the RJMCMC sampler we developed to explore the variable-dimension parameter space in question.

4.4.1 Models

We use the `enterprise`[1] software package for handling the pulse time-of-arrival (TOA) data and for calculating the likelihood for our Bayesian analysis. Timing models were produced by `libstempo`[2], which is a `python` wrapper for the `tempo2`[3] timing package. After applying the timing model, we get a set of timing residuals for each pulsar. We model the ith residual in the kth pulsar as:

$$\delta t_{ki} = \sum_l M_{kil} \delta\xi_{kl} + n_{ki} + g_{ki} + s_{ki}(\boldsymbol{\theta}), \tag{4.1}$$

where $\delta\xi_{kl}$ is the offset between the true timing model parameters and their best fit parameters, and M_{kil} is the design matrix, which represents a timing model linearized around the best fit parameters. n_{ki} is the noise present in the TOAs, g_{ki} is the contribution of the isotropic GWB, while $s_{ki}(\boldsymbol{\theta})$ describes the response to CWs, where $\boldsymbol{\theta}$ is the set of parameters describing the CWs. From these four components, three of them are modeled directly (see following subsections), while the timing model component is analytically marginalized over [117]. In the following sections we supress the i index to simplify the notation.

4.4.1.1 Noise model The noise in the TOAs is assumed to be Gaussian, uncorrelated between pulsars, and it can be decomposed in the kth pulsar as:

$$n_k = n_k^{\mathrm{WN}} + n_k^{\mathrm{RN}}, \tag{4.2}$$

[1] `https://github.com/nanograv/enterprise`
[2] `https://github.com/vallis/libstempo`
[3] `https://bitbucket.org/psrsoft/tempo2/`

where n_k^{WN} is the white noise (WN) and n_k^{RN} is the red noise (RN) in the kth pulsar. We model the covariance matrix of n_k^{WN} as:

$$C_{k,ij}^{\mathrm{WN}} = E_k^2 W_{ij}, \tag{4.3}$$

where E_k is the "EFAC" parameter from `TEMPO2`, and $W = \mathrm{diag}\,\sigma_i^2$, where σ_i is the nominal uncertainty of the ith TOA. We used a uniform prior on each "EFAC" value in the range of $[0.01, 10]$. This is a simplified version of the WN model commonly used in PTA data analysis (see e.g. [118]), which can be extended easily if we want to analyze real PTA data.

To model n_k^{RN}, we follow the procedure commonly used by NANOGrav (see e.g. [103]). We describe the RN power spectral density (PSD) with a power-law model:

$$P_k^{\mathrm{RN}}(f) = \left(A_k^{\mathrm{RN}}\right)^2 \left(\frac{f}{f_{\mathrm{yr}}}\right)^{-\gamma_k}, \tag{4.4}$$

where A_k^{RN} is the RN amplitude, and γ_k is the spectral index in the kth pulsar, while $f_{\mathrm{yr}} \equiv 1/(1\mathrm{yr})$. We use 30 frequency bins spaced linearly between $1/T_{\mathrm{obs}}$ and $30/T_{\mathrm{obs}}$, where T_{obs} is the total observation time. Note that $30/T_{\mathrm{obs}}$ is smaller than our Nyquist frequency, but we only need to model RN at low frequencies, because WN will dominate at higher frequencies. Since the simulated datasets used in this study (see Section 4.5) does not include RN, including it in the model is only motivated by the fact that it is similar to the GWB model (the difference being in the correlations, see Section 4.4.1.2), so it will provide a competing model. This motivates us to simplify the RN model by requiring it to be a common red process for all pulsars, i.e. $A_k^{\mathrm{RN}} = A^{\mathrm{RN}}$ and $\gamma_k = \gamma$ for all k. This assumption will simplify our modeling, and speed up sampling, but can be easily relaxed in the future when analyzing actual PTA data. We used a uniform prior on the amplitude $A^{\mathrm{RN}} \in [10^{-18}, 10^{-13}]$ and the spectral index $\gamma \in [0, 7]$.

4.4.1.2 GWB model The timing residuals resulting from a GWB are described by a Gaussian process, which is correlated between pulsars, and described by the following cross-power spectral density:

$$S_{kl}^{\rm GWB}(f) = \Gamma_{kl}\frac{\left(A^{\rm GWB}\right)^2}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma}, \tag{4.5}$$

where Γ_{kl} is the overlap reduction function characterizing the correlations between the kth and lth pulsars. In our case, where we consider an isotropic GWB produced by SMBHBs, the overlap reduction function is given by the Hellings and Downs curve [119]. $A^{\rm GWB}$ is the amplitude of the GWB, and $\gamma = 13/3$ is the expected spectral index of the GWB [120, 121], which we keep fixed in our analysis. However, our analysis can be easily extended to vary γ. We use the same 30 frequency bins as for the RN. We used a uniform prior on the amplitude $A^{\rm GWB} \in [10^{-18}, 10^{-13}]$.

4.4.1.3 CW model In the kth pulsar, we model multiple CW signals as:

$$s_k(t;\boldsymbol{\theta}) = \sum_{m=1}^{N_{\rm CW}}\left[F^{+}(\boldsymbol{\theta}_m,\theta_k,\phi_k)\Delta s_k^{+}(t;\boldsymbol{\theta}_m) + F^{\times}(\boldsymbol{\theta}_m,\theta_k,\phi_k)\Delta s_k^{\times}(t;\boldsymbol{\theta}_m)\right], \tag{4.6}$$

where $N_{\rm CW}$ is the number of CW sources included in the model, $\boldsymbol{\theta}_m$ are the parameters describing the mth CW source, $F^{+,\times}(\boldsymbol{\theta}_m,\theta_k,\phi_k) = F^{+,\times}(\theta_m^{\rm GW},\phi_m^{\rm GW},\theta_k,\phi_k)$ are the antenna pattern functions given e.g. by Eq. (7) and (8) in [103], where $\theta_m^{\rm GW}$ and $\phi_m^{\rm GW}$ are spherical coordinates of the mth CW source, while θ_k and ϕ_k are spherical coordinates of the kth pulsar. $\Delta s_k^{+,\times}$ is the difference between the metric perturbation at the Earth ("Earth term") and at the kth pulsar ("pulsar term"). Due to different pulsar distances, the pulsar terms do not add up as coherently as the Earth terms, so for simplicity, we only include the Earth terms in this study. Pulsar terms can be easily included in our analysis, with the expense of a few additional parameters we need to sample. This simplifications makes $\Delta s_k^{+,\times}(t;\boldsymbol{\theta}_m) = -s^{+,\times}(t; A_{\rm CW}, \omega^{\rm GW}, \iota, \psi, \Phi_0)$, where $s^{+,\times}$ are the $+$ and $\times$ polarization Earth terms given by:

$$s^{+}(t) = \frac{A_{\rm CW}}{\omega^{\rm GW}}\left[-\sin(\omega^{\rm GW}t + 2\Phi_0)\left(1+\cos^2\iota\right)\cos 2\psi - 2\cos(\omega^{\rm GW}t + 2\Phi_0)\cos\iota\sin 2\psi\right], \quad (4.7)$$

$$s^{\times}(t) = \frac{A_{\rm CW}}{\omega^{\rm GW}}\left[-\sin(\omega^{\rm GW}t + 2\Phi_0)\left(1+\cos^2\iota\right)\sin 2\psi + 2\cos(\omega^{\rm GW}t + 2\Phi_0)\cos\iota\cos 2\psi\right], \quad (4.8)$$

where $A_{\rm CW}$ is the GW amplitude, $\omega^{\rm GW} = 2\pi f^{\rm GW}$ is the GW angular frequency, ι is the inclination angle, and ψ is the polarization angle, and Φ_0 is the initial orbital phase of a CW source. We can see that our CW model is described by 7 $N_{\rm CW}$ parameters, and the parameters of a single source are $\boldsymbol{\theta}_m = \{A_{{\rm CW},m}, f_m^{\rm GW}, \iota_m, \psi_m, \Phi_{0,m}, \theta_m^{\rm GW}, \phi_m^{\rm GW}\}$. We used a uniform prior on the amplitude $A_{\rm CW} \in [10^{-18}, 10^{-13}]$, the frequncy $f^{\rm GW} \in [3.5\text{ nHz}, 100\text{ nHz}]$, the polarization angle $\psi \in [0, \pi]$, the initial phase $\Phi_0 \in [0, 2\pi]$, and the cosine of the inclination angle $\cos\iota \in [-1, 1]$. We also used a uniform on the sky prior resulting in a uniform prior on $\phi^{\rm GW} \in [0, 2\pi]$ and on $\cos\theta^{\rm GW} \in [-1, 1]$.

4.4.2 RJMCMC

As described in Section 4.4.1, the number of parameters in our model is:

$$N(N_{\rm CW}, N_{\rm GWB}) = N_{\rm RN} + N_{\rm psr} + 7N_{\rm CW} + N_{\rm GWB}, \quad (4.9)$$

where $N_{\rm RN} = 2$ is the number of parameters used to describe RN, $N_{\rm psr}$ is the number of pulsars in the array, which corresponds to the number of parameters used in the WN model, $N_{\rm CW} \in [0, N_{\rm CW}^{\rm max}]$ is the number of CWs included in the model, where $N_{\rm CW}^{\rm max}$ is the maximum number of CWs allowed, and $N_{\rm GWB} = \{0, 1\}$ indicates if a GWB is included or not. The GWB is described by only one parameter, $A^{\rm GWB}$, because we fix $\gamma = 13/3$. Note that $N_{\rm CW}$ and $N_{\rm GWB}$ are allowed to vary, so the dimensionality of our parameter space is not fixed. We use a uniform prior on both $N_{\rm CW}$ and $N_{\rm GWB}$.

To sample this variable-dimension parameter space we implemented an RJMCMC (also called trans-dimensional MCMC, see e.g. [111, 122, 123, 124]) sampler called `BayesHopper`[4], which treats $N_{\rm CW}$ and $N_{\rm GWB}$ as free parameters allowing to sample the full parameter space. In the rest of this section, we describe different components of this algorithm.

Note that a similar data analysis problem arises in the context of LISA, where one needs to model a collection of signals from tens of thousands of galactic binaries (see e.g. [113, 114, 115, 116]). As a result, the sampling techniques we use show many similarities to those used in [116].

4.4.2.1 Fisher matrix proposals To achieve good mixing within a given model, we use a proposal based on the Fisher information matrix, defined as:

$$\Gamma_{ij} = -\partial_i \partial_j \ln p(d|\boldsymbol{\theta}), \tag{4.10}$$

where $p(d|\boldsymbol{\theta})$ is the likelihood, and ∂_i indicates partial derivative with respect to θ_i. We calculate Γ_{ij} every few hundred iterations numerically, since it is changing as we move around in the parameter space. If $p(d|\boldsymbol{\theta})$ is exactly Gaussian, Γ_{ij} gives the inverse of the covariance matrix, but the proposal works well even for non-Gaussian distributions.

When attempting a Fisher matrix based step, the proposed $(i+1)$th sample is given by:

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \xi \boldsymbol{e}_j, \tag{4.11}$$

where $\boldsymbol{e}_j$ is the jth normalized eigenvector of Γ^{-1}_{ij}, and $\xi \sim \mathcal{N}(0, \Delta_j^2)$, where Δ_j is the jth eigenvalue of Γ^{-1}_{ij}. We choose j randomly for each proposed step. If the target distribution is not too different from a multivariate Gaussian, this proposal is practically a draw from the marginal distribution along a principal axis, thus we get optimal acceptance and mixing of

[4]`https://github.com/bencebecsy/BayesHopper`

the chain. We find that in this application, the posterior distributions around the maximum are close enough to being Gaussian that this proposal results in very good mixing.

4.4.2.2 Global proposal using the $\mathcal{F}_e$-statistic The CW model has 7 parameters, some of which are very well constrained. This means that starting from a random point in the parameter space, it takes many iterations to find the maximum of the posterior using our Fisher matrix based proposal (or any local proposal). This is especially problematic in an RJMCMC such as this, because the sampler might only spend a few hundred samples in one model before changing the number of parameters, so the sampler should converge quickly.

To help with convergence we introduce a global proposal based on the Earth-term $\mathcal{F}$-statistic ($\mathcal{F}_e$, see [125]). $\mathcal{F}_e$ is the CW likelihood ratio analytically maximized over h, Φ_0, ψ, and ι. As such, it only depends on $f^{\rm GW}$, $\theta^{\rm GW}$, and $\phi^{\rm GW}$, and gives a good idea of how likely it is to have a CW source at a given frequency and sky location. Figure 4.1 shows $\mathcal{F}_e$ for the dataset with an $A_{\rm CW} = 7.08 \times 10^{-15}$ simulated CW signal used in Section 4.5.2. Note how $\mathcal{F}_e$ peaks at the true sky location and frequency of the simulated CW source, and that its spread is relatively small compared to the whole range of parameters. This indicates that by using $\mathcal{F}_e$ as a proposal we can greatly improve convergence, since we will effectively always propose a CW source to the close proximity of the true parameters. This results in at least two orders of magnitude improvement in convergence speed. The maximized values of h, Φ_0, ψ, and ι can also be calculated. These can be used to construct informed proposals for these parameters too, however currently we draw those parameters from their prior distributions.

4.4.2.3 Trans-dimensional proposals The most important proposal in our RJMCMC is the one changing $N_{\rm CW}$ or $N_{\rm GWB}$. We handle these separately, so we have a proposal which adds or removes a CW signal, and one which adds or removes the GWB from the model. For the GWB adding proposal we draw the GWB amplitude from the prior, while for adding a CW we use $\mathcal{F}_e$ described above in Section 4.4.2.2. For removing a CW, we randomly select

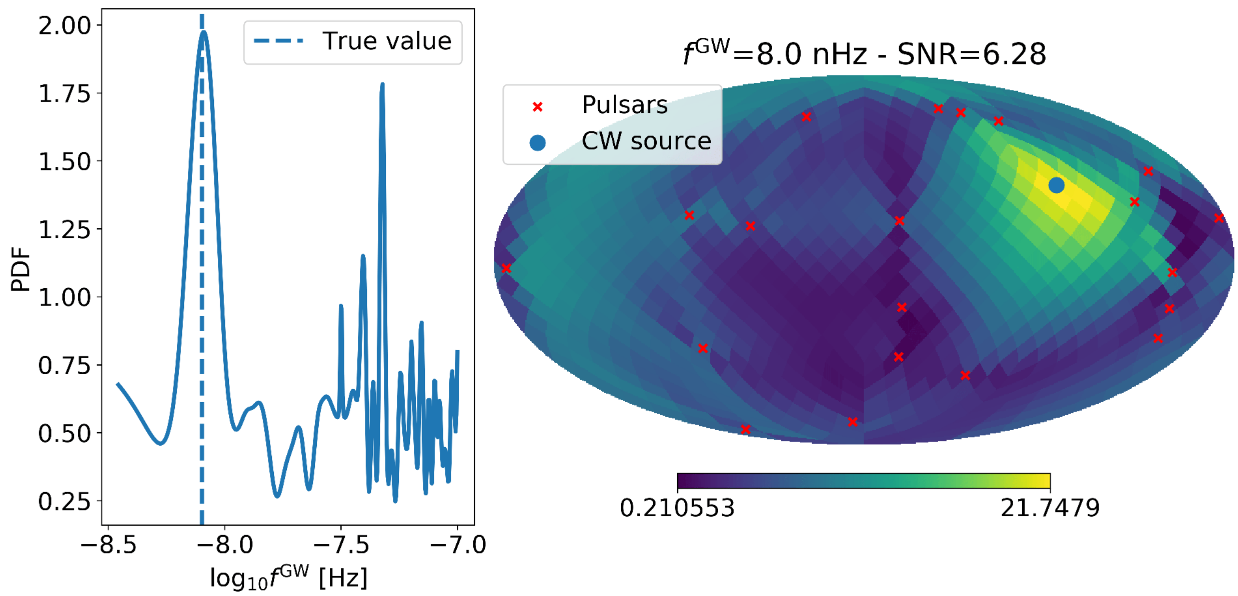


Figure 4.1: Global proposal for the dataset with an $A_{\rm CW} = 7.08 \times 10^{-15}$ simulated CW signal used in Section 4.5.2. *Left:* $\mathcal{F}_e$ marginalized over sky location showing the PDF of proposing a given value of $\log_{10} f^{\rm GW}$. *Right:* $\mathcal{F}_e$ at the true frequency. Positions of the simulated source and pulsars in the array are indicated by the blue circle and red crosses, respectively. Note how strongly $\mathcal{F}_e$ concentrates around the true sky location and true frequency of the CW source, indicating that it is much more efficient as a proposal than just drawing sky location and frequency from a uniform distribution.

one of the active CW sources to take out from the model.

In practice, these trans-dimensional proposals are much like the within-model ones, with the exception that one should make sure that the proposal densities and priors are properly normalized. While for within-model proposals, an overall normalization of these cancels out in the acceptance probability, having an error in normalizations in a trans-dimensional proposal can lead to biasing our posteriors on $N_{\rm CW}$ or $N_{\rm GWB}$. To make sure that is not the case, we performed prior recovery tests, i.e. we set the likelihood to a constant and checked if we recover the priors. `BayesHopper` passed all such tests.

4.4.2.4 Parallel tempering Parallel tempering [126] (also called Metropolis-coupled MCMC or replica exchange) is commonly used in MCMCs to help explore a multimodal distribution, and it also helps mixing in RJMCMCs (see e.g. [112]). We run multiple chains in parallel, with the ith chain using a modified likelihood of $p(d|\boldsymbol{\theta})^{1/T_i}$, where T_i is the temperature of the ith chain. A geometrically spaced temperature ladder is commonly used, where $T_i = c^i$, and c is a constant usually chosen to be between 1.2 and 2. The chains with higher temperature can explore the whole parameter space more easily, and by proposing swaps between chains adjacent on the temperature ladder we help the mixing of the first chain. Eventually, we only record samples from the first chain which has a temperature of 1, because that is the only chain that actually samples from the posterior.

We have found that parallel tempering is essential for exploring different values of $N_{\rm CW}$. It results in significantly more $N_{\rm CW}$ changing steps in the $T = 1$ chain than our dedicated trans-dimensional proposal (see Section 4.4.2.3). This is especially true for high signal-to-noise ratio (SNR) sources, where it gets hard to let go of a source once the sampler locks onto it. However, a dedicated trans-dimensional step is still needed to effectively explore different $N_{\rm CW}$ values at higher temperature chains. This is because parallel tempering can only change $N_{\rm CW}$ at a given temperature to its value at an adjacent temperature. E.g. if a given $N_{\rm CW}$ value is not present throughout the temperature ladder, parallel tempering cannot introduce it, but the trans-dimensional proposal can.

4.4.2.5 Trans-dimensional post processing for multiple CWs At the ith iteration, the sampler models the dataset with $N_{\rm CW,i}$ CWs and labels them $\rm CW_1$, $\rm CW_2$, ..., $\rm CW_{N_{\rm CW,i}}$. However, sources with the same label at different iterations do not necessarily correspond to the same physical source, e.g. $\rm CW_2$ from iteration $i = 2$ might not correspond to the same physical source as $\rm CW_2$ from iteration $i = 1$, but might correspond to the same physical source as $\rm CW_4$ from iteration $i = 5$. To produce Bayes factors and posterior distributions for unique sources, we need a post processing step, which looks through all the samples, identifies individual sources and associates samples with these identified sources. Our post processing procedure works as follows:

1. We scan through all the samples and find the one with the highest likelihood (L_i). Within the sample with the highest likelihood, we calculate the change in the log likelihood if the given source (labeled by j) is turned off ($\Delta \ln L_{i,j}$). The source with the highest $\Delta \ln L_{i,j}$ will be our reference source, denoted with $\bar{\boldsymbol{h}}$.

2. We scan through all the sources in all samples and calculate our distance measure between each source ($\boldsymbol{h}$) and the reference source, defined as:

$$\mathcal{D}(\bar{\boldsymbol{h}}, \boldsymbol{h}) = \left(\bar{\boldsymbol{h}} - \boldsymbol{h} | \bar{\boldsymbol{h}} - \boldsymbol{h}\right), \tag{4.12}$$

where $(\boldsymbol{a}|\boldsymbol{b})$ is the noise-weighted inner product between $\boldsymbol{a}$ and $\boldsymbol{b}$, defined as:

$$(\boldsymbol{a}|\boldsymbol{b}) = \boldsymbol{a}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{b}, \tag{4.13}$$

where $\boldsymbol{\Sigma}$ is the post-fit noise covariance matrix of the PTA. We then associate sources to the reference source, if $\mathcal{D}(\bar{\boldsymbol{h}}, \boldsymbol{h}) > \hat{\mathcal{D}}$. The optimal value of the $\hat{\mathcal{D}}$ threshold can be determined knowing that [127]:

$$\mathcal{D}(\bar{\boldsymbol{h}}, \boldsymbol{h}) \sim \chi^2_{D-1}, \tag{4.14}$$

where $D = 7$ is the number of signal parameters. To make sure we include all samples corresponding to a given source, we can take the 99.9th percentile of the χ^2_{D-1} distribution, which sets $\hat{\mathcal{D}} \simeq 25$.

3. We repeat (i) and (ii) with already assigned sources excluded until all sources are

associated with a physical source. We skip any reference source with SNR<5, since even if $\boldsymbol{h}=0$, $\mathcal{D}(\bar{\boldsymbol{h}},\boldsymbol{h}) = (\bar{\boldsymbol{h}}|\bar{\boldsymbol{h}}) = \mathrm{SNR}^2$, so with a reference source below an SNR of 5, all low SNR samples would be associated together.

This procedure results in a number of identified physical sources, each with $\bar{n}_{\mathrm{CW}}$ source samples associated with them. Comparing $\bar{n}_{\mathrm{CW}}$ to the total number of independent samples, we can calculate Bayes factors with the RJ method (see Section 4.4.2.6) for each physical source. Analyzing the set of $\bar{n}_{\mathrm{CW}}$ source samples for each identified source, we can perform source-by-source parameter estimation.

4.4.2.6 Calculating Bayes factors An RJMCMC algorithm provides a natural way of calculating Bayes factors between different models, by comparing the time the sampler spent at each model. So the Bayes factor between model $\mathcal{A}$ and $\mathcal{B}$ is given by:

$$\mathrm{BF}_{\mathcal{A}-\mathcal{B}} = \frac{n_{\mathcal{A}}}{n_{\mathcal{B}}}, \tag{4.15}$$

where $n_{\mathcal{A}}$ and $n_{\mathcal{B}}$ are the number of independent samples from model $\mathcal{A}$ and $\mathcal{B}$, respectively. We will refer to this method of calculating Bayes factors as RJ method. It works well if $0.01 \lesssim \mathrm{BF}_{\mathcal{A}-\mathcal{B}} \lesssim 100$, but otherwise the sampler spends most of the time in one of the models and the estimation becomes unreliable. This can be remedied by the prior re-weighting method, i.e. by introducing an artificial prior such that $n_{\mathcal{A}} \simeq n_{\mathcal{B}}$. The ideal prior can be determined from a pilot run, and the final result can be corrected for this artificial prior. This method theoretically allows us to calculate Bayes factors of any magnitude, but in practice it only helps a few orders of magnitude, beyond which multiple pilot runs may be needed to find the ideal value of the artificial prior. The statistical uncertainty of $\mathrm{BF}_{\mathcal{A}-\mathcal{B}}$ can be estimated by assuming that $n_{\mathcal{A}}$ and $n_{\mathcal{B}}$ follow a Poisson distribution.

The Bayes factors between models $\mathcal{A}:\{N_{\mathrm{GWB}}=1\}$ and $\mathcal{B}:\{N_{\mathrm{GWB}}=0\}$, or for

$\mathcal{A}:\{N_{\text{CW}}=1\}$ and $\mathcal{B}:\{N_{\text{CW}}=0\}$ can also be calculated using the Savage-Dickey density ratio (SD method). The Bayes factor in this case is given by:

$$\text{BF}_{\mathcal{A}-\mathcal{B}} = \frac{p(A=0|\mathcal{A})}{p(A=0|d,\mathcal{A})}, \tag{4.16}$$

where A is the amplitude parameter of model $\mathcal{A}$, such that $A=0$ is equivalent of model $\mathcal{B}$, $p(A=0|\mathcal{A})$ is the prior at $A=0$, and $p(A=0|d,\mathcal{A})$ is the posterior at $A=0$. In practice, the posterior of A is only available as a histogram from posterior samples, and we approximate $p(A=0|d,\mathcal{A})$ with the lowest amplitude bin[5]. This also means that we have to decide what bin size to use, which will affect our estimation of $\text{BF}_{\mathcal{A}-\mathcal{B}}$. The smaller the bin size is, the closer we are to truly measuring the posterior at $A=0$, but the fewer independent samples we have in the lowest bin. Another issue arises as $\text{BF}_{\mathcal{A}-\mathcal{B}}$ gets higher, because $p(A|d,\mathcal{A})$ will peak at a non-zero value of A, and it will tail off at low values of A providing very few samples in the lowest bin. A solution to that is doing "zoom-in" reruns [128], where we rerun the sampler with a reduced prior range on A. The rerun will provide more samples at the lowest bin, and we can apply a correction to the Bayes factor, because we know the fraction of samples falling into the reduced prior range from the first run. This procedure can be repeated multiple times, until we get a sufficient number of samples at the lowest bin. The statistical uncertainty on the SD Bayes factor can be estimated by assuming that the number of independent samples in the lowest bin follows a Poisson distribution.

4.5 Results

We tested `BayesHopper` on different simulated datasets to demonstrate its capabilities. All datasets correspond to a simulated pulsar timing array with 20 pulsars at random sky

[5]Note that technically our prior does not include $A=0$ (see Sections 4.4.1.2 and 4.4.1.3), but the lower limit is typically much smaller than the bin size, so the error introduced by this is much smaller than the effect of having a finite bin size.

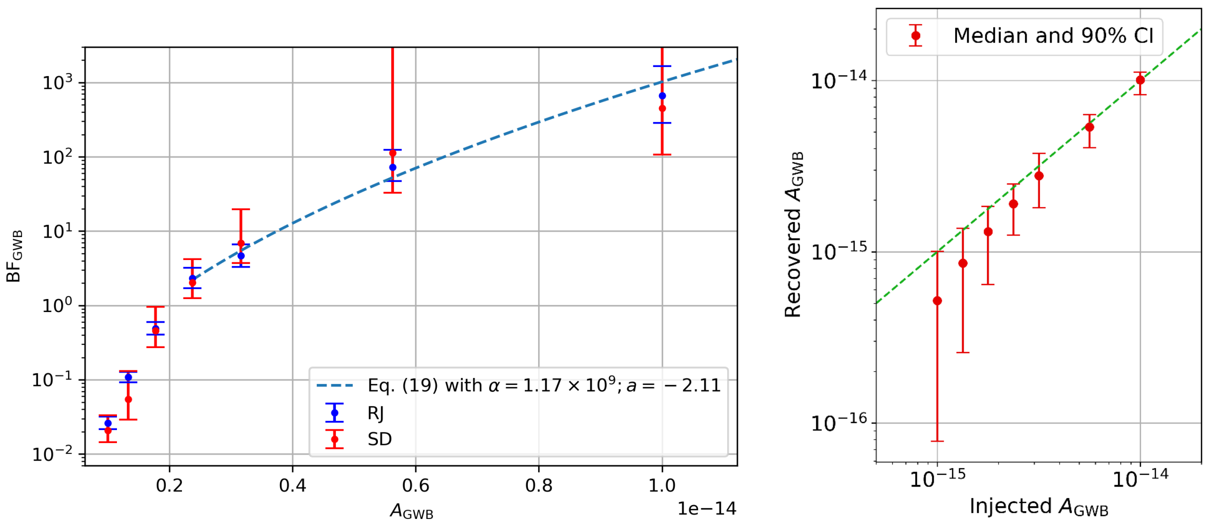
locations (see red crosses on Figure 4.1) and the same white nose level of $\sigma = 0.5$ μs for each pulsar. Our datasets does not include RN. All pulsars of the simulated dataset were observed for 10 years, evenly spaced at every 30 days. We added different signals to this dataset to test different parts of our algorithm. First, we looked at datasets with only a GWB (see Section 4.5.1) or only a single CW in them (see Section 4.5.2). Our results for a dataset with three CWs is presented in Section 4.5.3. Finally, Section 4.5.4 describes our analysis of a dataset with both a GWB and three CWs in it.

4.5.1 Stochastic background

We simulated a GWB with different amplitudes resulting in 7 datasets. We ran `BayesHopper` on these datasets in GWB-only mode, i.e. allowing $N_{\rm GWB}$ to be either 1 or 0, but requiring $N_{\rm CW} = 0$. We then calculated the Bayes factor between the $N_{\rm GWB} = 1$ and $N_{\rm GWB} = 0$ models, denoted as $\rm BF_{GWB}$, using both RJ and SD methods described in Section 4.4.2.6.

Figure 4.2a shows $\rm BF_{GWB}$ as a function of the amplitude of the simulated GWB ($A_{\rm GWB}$). Bayes factors obtained using Savage-Dickey density ratio are labeled with red, while Bayes factors calculated using RJMCMC methods are blue. Error bars indicate 90% credible intervals obtained by methods described in Section 4.4.2.6. Note that $\rm BF_{GWB}$ calculated with the RJ method, our primary method of calculating Bayes factors in an RJMCMC, are consistent with the ones calculated using the SD method, which is widely used in analysis of pulsar timing arrays.

For the two highest amplitude GWBs, our $\rm BF_{GWB}$ calculated using the SD method is consistent with it being infinity. This is due to the fact that for high-amplitude signals, we get an increasingly small amount of samples at the lowest $A_{\rm GWB}$ bin, which is where the SD method estimates the Bayes factor. In this case, even after applying the "zoom-in" procedure described in Section 4.4.2.6, we still only get a few samples in the lowest amplitude



(a) (b)

Figure 4.2: (a) Bayes factors for a GWB signal for different levels of the background. Bayes factors obtained using Savage-Dickey density ratio are labeled with red, while Bayes factors calculated using RJMCMC methods are blue. Note that the two methods give consistent results. Error bars indicate 90% credible intervals obtained by methods described in Section 4.4.2.6. The dashed line shows a fit in the high SNR regime using the theoretical scaling described by (4.19). (b) Recovered values of A_{GWB} as a function of the injected values present in the dataset. Medians and 90% credible intervals are shown for each dataset. The dashed line indicates perfect recovery, i.e. where the injected and recovered amplitudes are the same.

bin, which is consistent with being zero at the 90% credibility level, so the resulting Bayes factor is consistent with infinity. One can do multiple "zoom-in" reruns to remedy this, but we decided not to carry out those, since our SD Bayes factors are just a secondary product meant to be a sanity check of our RJ Bayes factors.

To check if the $\mathrm{BF}_{\mathrm{GWB}}$ values we get are consistent with what we would expect for a GWB observed by a pulsar timing array, we derive the theoretically expected scaling of $\mathrm{BF}_{\mathrm{GWB}}$ with A_{GWB}, and compare it to our results. Using the Laplace approximation the

Bayes factor can be written as:

$$\ln \mathrm{BF} = \ln \mathcal{L}_{\mathrm{max}} + \ln \left[p(\theta_{\mathrm{ML}}) V_{\mathrm{post}} \right], \tag{4.17}$$

where $\mathcal{L}_{\mathrm{max}}$ is the maximum likelihood ratio, and the second term is the Occam penalty, where $p(\theta_{\mathrm{ML}})$ is the prior at the maximum likelihood parameters, and V_{post} is the posterior volume. Note that if $p(\theta)$ is uniform in all parameters, then the second term in (4.17) is just the logarithm of the ratio between posterior and prior volumes.

For the GWB model, the maximum likelihood ratio is given by $\ln \mathcal{L}_{\mathrm{max}} = \rho^2/2$, where ρ is the optimal statistic SNR [129]. The scaling of ρ^2 with A_{GWB} is given by [130]:

$$\rho^2(A_{\mathrm{GWB}}; \alpha) = \alpha \left[{}_2F_1 \left(2, \gamma^{-1}; 1 + \gamma^{-1}; -\frac{12\pi^2 f^{\gamma} P_{\mathrm{WN}}}{A_{\mathrm{GWB}}^2 f_{\mathrm{yr}}^{4/3}} \right) f \right]_{f=f_{\mathrm{min}}}^{f_{\mathrm{max}}}, \tag{4.18}$$

where α is a constant independent of A_{GWB}, ${}_2F_1$ is the ordinary hypergeometric function, $\gamma = 13/3$ is the spectral slope of the GWB, and $P_{\mathrm{WN}} = 2\sigma^2 \Delta t$ is the PSD of the white noise, where $\Delta t = 30$ days is cadence. $f_{\mathrm{min}} = 1/T_{\mathrm{obs}}$ is the minimum, while $f_{\mathrm{max}} = 30/T_{\mathrm{obs}}$ is the maximum frequency used in our GWB model (see Section 4.4.1.2). In case of a fixed spectral index GWB model, we only have one parameter, so in the high SNR regime, the Occam penalty terms will not depend strongly on A_{GWB}, so we will assume they are constant. So we expect the Bayes factor to be given by:

$$\ln \mathrm{BF}_{\mathrm{GWB}} \approx \frac{\rho^2(A_{\mathrm{GWB}}; \alpha)}{2} + a, \tag{4.19}$$

where a is a free parameter describing the A_{GWB}-independent contribution from the Occam terms. Note that this expression has two free parameters, a and α from (4.18).

In Figure 4.2a we fit the highest 4 $\mathrm{BF}_{\mathrm{GWB}}$ values from the RJ method with this expected relation. We can see that (4.19) fits the results well, indicating that the Bayes factors we get are consistent with the theoretical expectations.

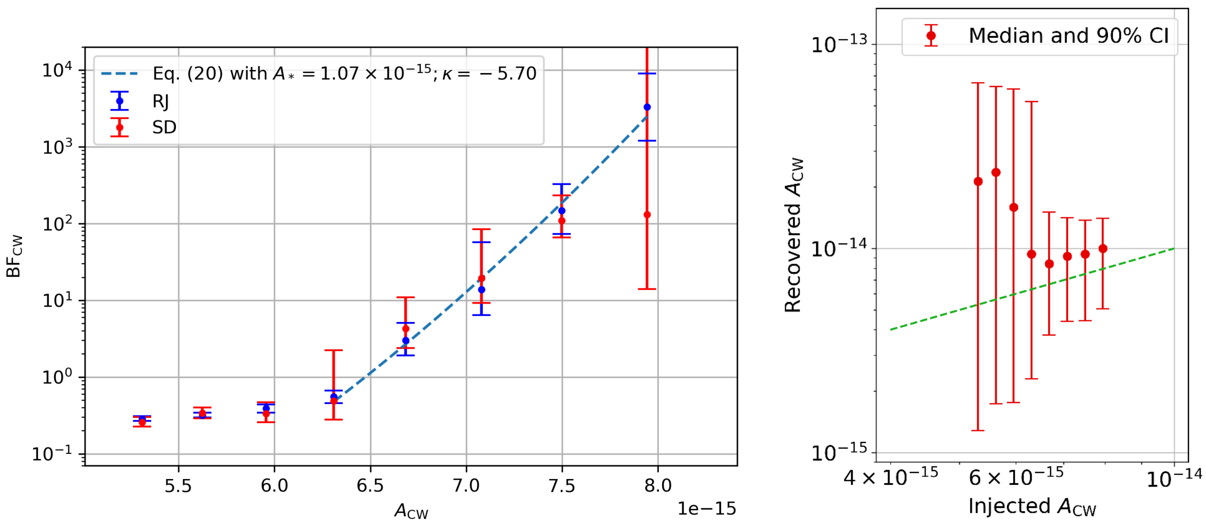
Figure 4.2b shows the median recovered $A_{\rm GWB}$ values with their 90% credible intervals as a function of the injected amplitude value present in the dataset. We can see that as $A_{\rm GWB}$ increases the recovered value approaches the true value, as expected when we transition from an upper-limit type of result to a detection. We can also see how the relative uncertainty on $A_{\rm GWB}$ decreases as the injected amplitude (and thus the SNR) increases.

4.5.2 Single continuous wave

We added a single CW signal to our dataset with different amplitudes ($A_{\rm CW}$), resulting in 8 different datasets. Other parameters of the source were kept fixed throughout these datasets, and were the same as for source #1 in Table 4.1. We ran `BayesHopper` on these datasets in CW-only mode with $N_{\rm CW}^{\rm max} = 1$, i.e. we fixed $N_{\rm GWB}$ to be zero, and allowed $N_{\rm CW}$ to be either 0 or 1. Then, we calculated Bayes factors between the $N_{\rm CW} = 1$ and $N_{\rm CW} = 0$ models, denoted as $\mathrm{BF}_{\rm CW}$, with both RJ and SD methods described in Section 4.4.2.6.

Figure 4.3a shows $\mathrm{BF}_{\rm CW}$ as a function of $A_{\rm CW}$, both for RJ (blue) and SD (red) methods. Error bars indicate 90% credible intervals obtained by methods described in Section 4.4.2.6. Note that $\mathrm{BF}_{\rm CW}$ calculated with the RJ method, our primary method of calculating Bayes factors in an RJMCMC, are consistent with the ones calculated using the SD method, which is widely used in data analysis of pulsar timing arrays.

To check if the $\mathrm{BF}_{\rm CW}$ values we get are consistent with what we would expect for a CW signal observed by a pulsar timing array, in the following we derive the theoretically expected scaling of $\mathrm{BF}_{\rm CW}$ with $A_{\rm CW}$, and compare it to our results. The general expression of a Bayes factor is given by (4.17). For the CW case, $\ln \mathcal{L}_{\rm max} = \mathrm{SNR}^2/2$. For slowly varying priors, and large SNR, the posterior volume can be approximated as $V_{\rm post} \approx (\det(\Gamma/2\pi))^{-1/2}$, where $\det(\Gamma)$ is the determinant of the Fisher information matrix, which scales with the SNR as $\det(\Gamma) \propto \mathrm{SNR}^{2D}$, where $D = 7$ is the model dimension. Since $\mathrm{SNR} \propto A_{\rm CW}$, we get the



(a) (b)

Figure 4.3: (a) Bayes factors of a CW signal for different injected amplitudes. Bayes factors obtained using Savage-Dickey density ratio are labeled with red, while Bayes factors calculated using RJMCMC methods are blue. Note that the two methods give consistent results. Error bars indicate 90% credible intervals obtained by methods described in Section 4.4.2.6. The dashed line shows a fit in the high SNR regime using the theoretical scaling of $\mathrm{BF_{CW}}$ from (4.20), where we find $A_* = 1.07 \times 10^{-15}$ and $\kappa = -5.7$. The good fit indicates that the Bayes factors we get are consistent with the theoretical scaling. (b) Recovered values of A_{CW} as a function of the injected values present in the dataset. Medians and 90% credible intervals are shown for each dataset. The dashed line indicates perfect recovery, i.e. where the injected and recovered amplitudes are the same.

following theoretical scaling of the CW Bayes factor:

$$\ln \mathrm{BF_{CW}} \approx \frac{1}{2}\left(\frac{A_{\mathrm{CW}}}{A_*}\right)^2 - D \ln\left(\frac{A_{\mathrm{CW}}}{A_*}\right) + \kappa, \tag{4.20}$$

where A_* fixes the scaling between amplitude and SNR and κ is a constant.

Figure 4.3a shows the fit to $\mathrm{BF_{CW}}$ using the scaling described by (4.20). We can see that the scaling fits well in the high SNR regime, indicating that our Bayes factors are consistent with the theoretical expectation.

Figure 4.3b shows the median recovered A_{CW} values with their 90% credible intervals as a function of the injected amplitude value present in the dataset. We can see that as A_{CW} increases the recovered value approaches the true value, as expected when we transition from an upper-limit type of result to a detection. We can also see how the relative uncertainty on A_{CW} decreases as the injected amplitude (and thus the SNR) increases.

4.5.3 Multiple continuous waves

In this section we analyzed a single dataset, where we added 3 different CW sources. Parameters of the three source are given in Table 4.1. We analyzed this dataset by `BayesHopper` in the CW-only mode, with $N_{\mathrm{CW}}^{\mathrm{max}} = 10$, i.e. we fixed N_{GWB} to be zero, and allowed N_{CW} to vary between 0 and 10, with a uniform prior on N_{CW}.

Table 4.1: Parameters of simulated CW signals.

Source number	f^{GW} [nHz]	A_{CW} [10^{-15}]	θ^{GW}	ϕ^{GW}	Φ_0	ψ	ι	SNR
#1	8	7.5	$\pi/3$	4.5	1.0	1.0	1.0	8.6
#2	20	16.4	$2\pi/3$	1.0	0.5	0.0	0.8	7.0
#3	60	34.1	$\pi/6$	2.0	0.0	1.2	0.3	6.9

We used the post processing procedure described in Section 4.4.2.5 to identify physical CW sources, and calculated the Bayes factors for each of those. Figure 4.4 shows $\mathrm{BF}_{\mathrm{CW}}$ for each identified source, with the inset showing $f_{\mathrm{GW}} - A_{\mathrm{CW}}$ posterior samples for each source. We can see that the source separation algorithm does a fair job separating the sources, and that the posterior coincides with the true values of f_{GW} and A_{CW} marked with $\times$ symbols. Note that the three most significant sources identified correspond to the three simulated CW signals added to the dataset. As expected, the $\mathrm{BF}_{\mathrm{CW}}$ values we get for these increase with the SNR of the simulated signal. The most significant source not associated with an injected source has a Bayes factor slightly below 1. We re-analyzed the dataset with a different realization of the WN, and found that this source disappeared, but another source with similar significance appeared. This suggests that this source corresponds to the sampler trying to fit some noise fluctuation with the CW model, and it is reassuring to see that we slightly disfavor it being a real CW signal.

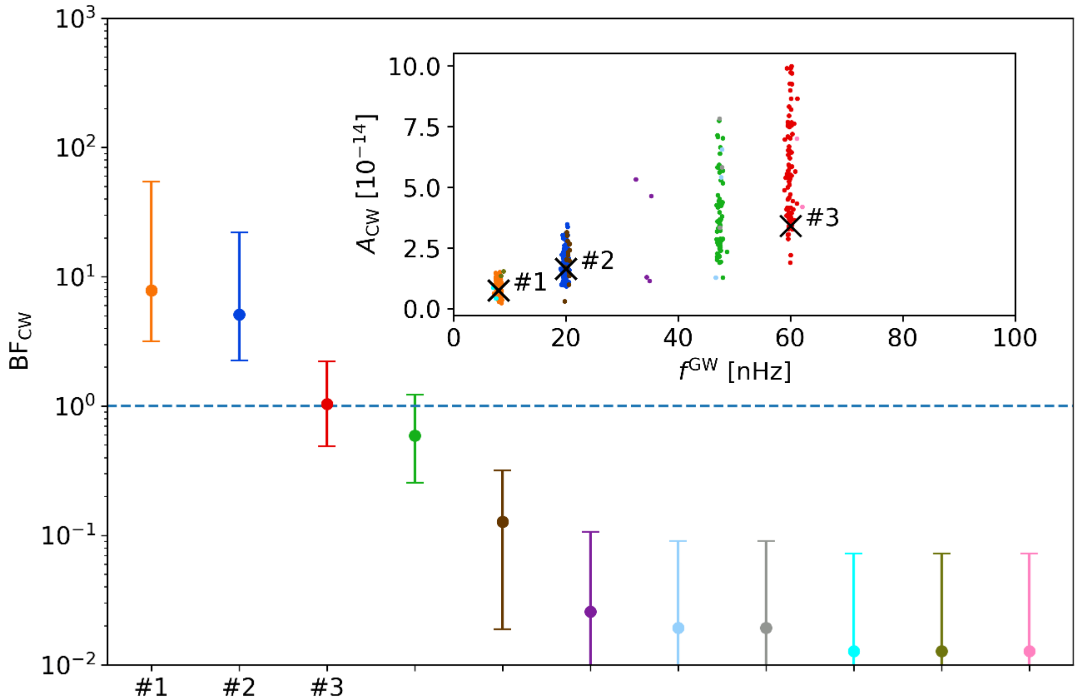


Figure 4.4: Bayes factors of different CW sources found by the source identification post processing step. Inset shows the distribution of f_{GW} and A_{CW} for each source, with true parameter values of the simulated sources indicated by $\times$ symbols. Note that the three simulated CWs (see Table 4.1) were clearly found with $\mathrm{BF}_{\mathrm{CW}}$ increasing with simulated SNRs.

4.5.4 Stochastic background and continuous waves

To assess how the GWB model in itself affects the CW model, we first analyzed the same dataset as in Section 4.5.3, but allowing the sampler to also vary N_{GWB}, instead of setting it to zero. We have found that the three simulated signals have $\mathrm{BF}_{\mathrm{CW}}$ values consistent with the CW-only run. So in this particular case, including a GWB signal in the model does not seem to affect our results for the CWs. This is not surprising, since we only have three CW sources, all of which are relatively strong. We expect the effect of the GWB model soaking up CW sources to play a stronger role when many low-SNR CWs are present.

To test BayesHopper on a dataset with both CWs and a GWB in it, we used the sources described in Table 4.1 and we also added an isotropic GWB signal with $A_{\rm GWB} = 3.98 \times 10^{-15}$. The GWB amplitude is near the threshold level identified in Figure 4.2a, and in the combined search yielded a Bayes factor of $\mathrm{BF}_{\rm GWB} = 5.18^{+4.82}_{-1.99}$. Figure 4.5 shows the identified sources with their $\mathrm{BF}_{\rm CW}$ values and $f_{\rm GW} - A_{\rm CW}$ posterior samples. The three most significant sources are still the three simulated CWs, similarly to what we have seen in Section 4.5.3, but their significance changed considerably. While the Bayes factor of source #2 stayed about the same, source #1 and #3 show a significant decrease in their Bayes factors. For source #1, this can be explained by the increased level of the noise at low frequencies due to the GWB. This explanation is also consistent with the fact that the $A_{\rm CW}$ posterior extends to much lower values for the dataset without GWB (compare inset figures of Figure 4.4 and 4.5). Source #3 is at much higher frequency, where the increase in the noise level due to the GWB is negligible. However, it was already a low-significance CW source, so in this case we think that the decrease in $\mathrm{BF}_{\rm CW}$ can be explained by the fact that the GWB model might be able to partially fit the contribution of source #3, which might be statistically preferred over the more complex CW model.

We have seen that the significance of CW sources can be affected by the presence of a GWB. To investigate whether CWs can affect the significance of GWBs, we analyze 2 datasets: one with no CWs and a GWB with $A_{\rm GWB} = 3.98 \times 10^{-15}$, and one with the same GWB and the three CWs described in Table 4.1. We analyze both of these datasets with two settings: one where our model can only contain a GWB signal, and one where it can contain a GWB signal and up to 10 CW signals. Table 4.2 shows the $\mathrm{BF}_{\rm GWB}$ values we get for each of these runs. We can see that when no CWs are present in the dataset, it does

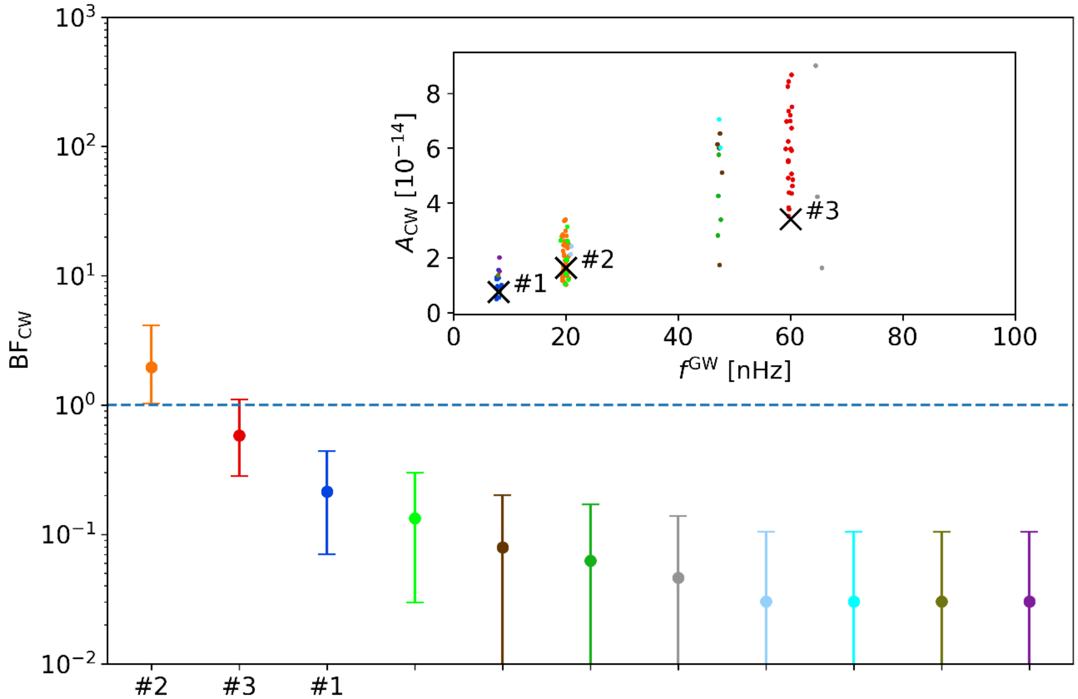


Figure 4.5: Bayes factors of different physical CW sources found by the source identification post processing step (see Section 4.4.2.5). Inset shows the distribution of $f_{\rm GW}$ and $A_{\rm CW}$ for each source, with true parameter values of the simulated sources indicated by $\times$ symbols. This dataset includes a simulated GWB with $A_{\rm GWB} = 3.98 \times 10^{-15}$ and CW sources described in Table 4.1. Note that only the moderate-SNR source at moderate frequency remained significant, while both the high-SNR source at low frequency and the moderate-SNR source at high frequency is unfavored when a GWB of this level is present.

not matter if we include CWs in the model or not, we get $\rm BF_{GWB} \sim 1$ in both cases. If CWs are present in the dataset and we include them in our model, we get a slightly higher $\rm BF_{GWB} \sim 4$. This is consistent with the fact that $\rm BF_{CW}$ decreased when we included GWBs, because part of the CW signals can be fit with the GWB model. Finally, if CWs are present in the dataset, but we do not model them, we see an even higher increase in the GWB significance ($\rm BF_{GWB} \sim 120$) indicating that in this case the GWB model tries to fit as much of the CW signals as it can. Figure 4.6 shows posterior distributions of $A_{\rm GWB}$ for the same runs. We see again that the GWB model takes some power from the CWs, increasing the significance of the GWB, and decreasing the significance of the CWs. Note that the change in the recovered amplitude is consistent with the bias we see in Figure 4.2b.

Table 4.2: Dependence of $\mathrm{BF}_{\mathrm{GWB}}$ on CWs. $\mathrm{BF}_{\mathrm{GWB}}$ values are shown for four different runs. Each dataset included a stochastic background with $A_{\mathrm{GWB}} = 3.98 \times 10^{-15}$, and datasets reported in the second column also included the 3 CW signals from Table 4.1. All runs included a GWB in the model, and runs reported in the second row also allowed for CW signals.

	CWs not in data	CWs in data
CWs not in model	$1.018^{+0.067}_{-0.063}$	124^{+35}_{-23}
CWs in model	$1.00^{+0.14}_{-0.12}$	$3.9^{+3.6}_{-1.5}$

4.6 Conclusion

We have presented `BayesHopper`, an RJMCMC algorithm for jointly searching for isolated CW sources and an isotropic GWB in PTA data. We have tested it on various simulated datasets, and have found that for the CW-only and GWB-only datasets it delivers results consistent with methods currently used in the PTA community. For datasets with multiple CWs in them, we have illustrated that an additional post processing step is needed to identify physical CW sources from the MCMC samples, and introduced a method which can separate these sources by finding the maximum likelihood parameters of sources and comparing it to other source parameters in the samples. For the full problem of analyzing a dataset with both a GWB and multiple CWs in it, we have seen some examples of how the CW and GWB models interact. While we have seen that allowing to fit for a GWB when no GWB is present does not change the CW results (at least in the one case we looked into), we have seen that the significance of CW sources can decrease in presence of a GWB because of two different reasons: i) the GWB increases the noise level, which can make a CW source at low frequency insignificant; ii) sometimes the GWB model can partially fit

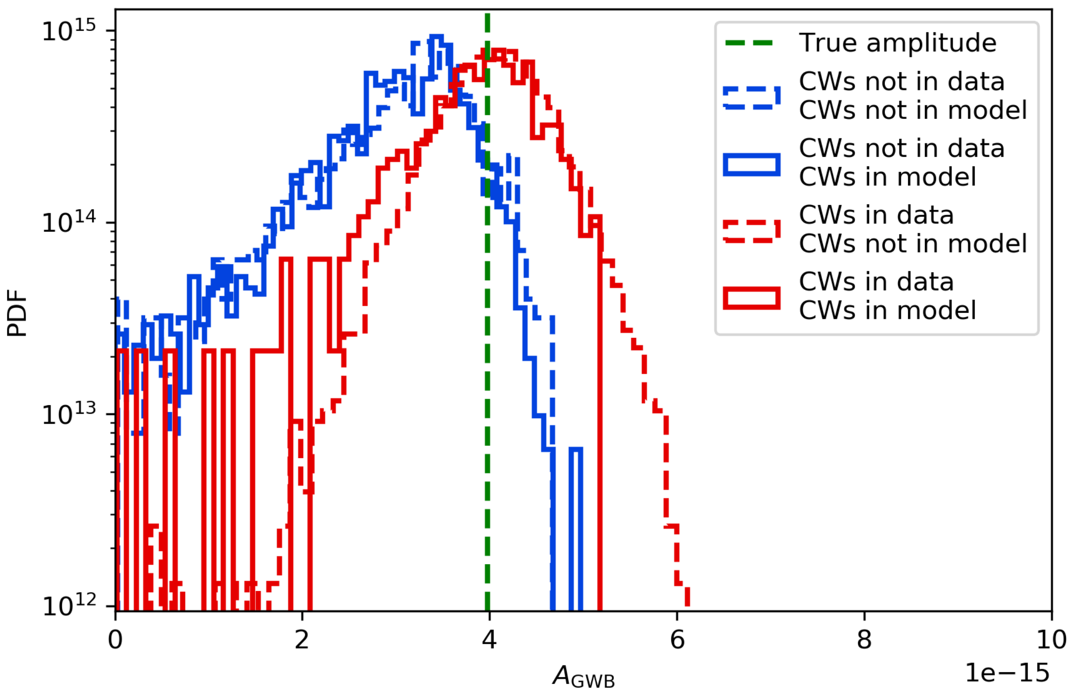


Figure 4.6: Posterior distribution of A_{GWB} for four different runs. Each dataset included a stochastic background with $A_{\mathrm{GWB}} = 3.98 \times 10^{-15}$. Blue histograms show results for a dataset with only this GWB signal, while red histograms are for a dataset that also included the 3 CW signals from Table 4.1. Solid lines show results for runs where we allowed the sampler to include CWs in the model, while dashed lines represent runs where we did not allow the sampler to include CWs in the model. We see that the GWB model absorbs some power from the CWs. The $\mathrm{BF}_{\mathrm{GWB}}$ values obtained for these runs are reported in Table 4.2

the CW signal, and having only 1 parameter, it can be preferred by the Bayesian analysis' natural parsimony, even if it provides a worse fit than the CW model, especially if the CW source had low-SNR to start with. This interaction between the GWB and CW models is what makes such a joint analysis interesting and worthwhile, and we plan to investigate such interactions in greater depth in upcoming studies.

Beyond being the first implementation of a joint analysis of CWs and a GWB in PTAs, `BayesHopper` is also the first algorithm which can calculate Bayes factors of a CW or a GWB based on an RJMCMC, instead of the Savage-Dickey density ratio method. It would be beneficial to run `BayesHopper` on real PTA data, to provide an independent measurement of the Bayes factors.

Work is currently underway in testing `BayesHopper` on a realistic cosmological simulation [59], where the GWB is not manually added as an isotropic background, but instead builds up from simulated BBHs in the Universe. This will provide a more realistic test of the algorithm, since the astrophysical background is expected to be built from a collection of individually unresolvable CWs. Indeed, the "true" model of the PTA dataset would be just a large collection of CWs, but due to the parsimony of Bayesian inference, a model with fewer parameters including a few CWs and an isotropic GWB will be preferred (see e.g. [108]). Such a study on realistic simulations will prepare us for analyzing real PTA datasets with our full model including multiple CWs and a GWB. Such an analysis of the NANOGrav 12.5-year dataset is planned for the near future.

We are also planning to build on the RJMCMC engine developed for `BayesHopper` to create a PTA version of the BayesWave [112] algorithm, which is used to search for unmodeled transient GWs in the data of ground-based interferometric GW detectors. This would only require changing the sinusoidals in our CW model to Morlet-Gabor wavelets used in BayesWave, and we would be in a position to search for unmodeled transients in PTA data.

ACKNOWLEDGEMENTS

We acknowledge the help of Logan O'Beirne in installing and using the `Enterprise` software package. We appreciate the support of the NSF Physics Frontiers Center Award PFC-1430284. Computations were performed on the Hyalite High Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure at Montana State University.

CHAPTER FIVE

BAYESIAN SEARCH FOR GRAVITATIONAL WAVE BURSTS IN PULSAR TIMING ARRAY DATA

5.1 Contribution of Authors and Co-Authors

Manuscript in Chapter 5

Author: Bence Bécsy

Contributions: Developed the `BayesHopperBurst` software presented in the paper. Performed the analysis. Wrote the first draft of the manuscript.

Author: Neil J. Cornish

Contributions: Conceived the study design. Provided guidance on the analysis. Edited the manuscript.

5.2 Manuscript Information

Bence Bécsy, Neil J. Cornish

Classical and Quantum Gravity

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
_X_ Published in a peer-reviewed journal

IOP Publishing Ltd
Published 14 April 2021
Volume 38, Issue 9, id.095012
10.1088/1361-6382/abf1c6

5.3 Introduction

The detection of nanohertz gravitational waves (GWs) is the main objective of pulsar timing arrays (PTAs). These large scale experiments regularly monitor a collection of millisecond pulsars to achieve this goal (see e.g. [14]). The three major pulsar timing arrays currently operating are the North American Nanohertz Observatory for Gravitational Waves (NANOGrav, [3]), the European Pulsar Timing Array (EPTA, [4]), and the Parkes Pulsar Timing Array (PPTA, [5]). In addition, there are emerging PTA efforts in India (InPTA, [131]), China (CPTA, [132]), and South Africa (SAPTA, [133]). These project are collaborating under the International Pulsar Timing Array (IPTA, [6, 93, 134]) consortium.

The most promising GW sources in the nHz regime are supermassive black hole binaries (SMBHBs). These can be detected through observing their collective effect as a red noise process with quadrupolar correlations between pulsars (see e.g. [7]). The most massive and nearby of these SMBHBs might also be detectable individually (see e.g. [62]). It has also been proposed that these two searches can be carried out simultaneously, thus accounting for any interaction between the two types of signatures [32]. Searches have also been carried out for bursts with memory [135], i.e., a permanent deformation of spacetime after a violent astrophysical event, like the merger of two black holes.

In this paper, we focus on searching for generic GW transients (a.k.a. GW bursts) in PTA data. The subject has been studied extensively in the context of ground-based interferometric GW detectors, where many different algorithms are in use [112, 136, 137, 138, 139, 140] and several searches have been carried out throughout different observing runs of the LIGO [141] and Virgo [142] detectors [143, 144]. The detection problem of GW bursts has also been considered for space-based detectors [145].

In the PTA context, no analysis of real data has been carried out so far. However, several methods have been suggested, including an analytical Bayesian framework [146], a Bayesian nonparametric approach [147], and a frequentist search working in the time-frequency domain [148]. Ref. [149] suggests a possible improvement by considering the coherence between the pulsar terms, which appears in some cases. In this paper we present `BayesHopperBurst`[1], a Bayesian search for GW bursts in PTA data, which is based on a trans-dimensional Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler [111, 122] akin to that used in the BayesWave algorithm [112] to search for and reconstruct GW bursts in the data of ground-based GW detectors. The work described in this paper is the continuation of that presented in [150], where the authors used similar methods to model noise transients in PTA data. One promising source of nHz GW bursts are parabolic (or highly eccentric) encounters of two SMBHs, which can occur in a hierarchical triple system of SMBHs [151]. Cosmic string kinks and cusps can also produce GW busts in this frequency range [152].

The paper is organized as follows. In Section 5.4 we describe the methods employed by `BayesHopperBurst` including the model used to describe transient signals and the sampling techniques that enables it to effectively explore the parameter space. In Section 5.5 we perform various injection tests, demonstrating that `BayesHopperBurst` recovers a wide variety of signals and noise transients. In Section 5.6 we analyze a simulated dataset similar to the latest NANOGrav data release [3] to make a prediction of what we can expect from analyzing the real dataset. In Section 5.7 we analyze B1855+09 from the NANOGrav 9-year dataset [153] to demonstrate our algorithm's noise transient modeling capabilities on real data. Finally, we offer concluding remarks in Section 5.8.

[1]`https://github.com/bencebecsy/BayesHopperBurst`

5.4 Methods

In this section, we introduce the model used to describe signals and noise transients with arbitrary morphology, and we provide some details of the sampling techniques employed to efficiently explore the resulting high-dimensional parameter space. We use the `enterprise`[2] [50] software package for handling PTA data and calculating the likelihood used in our Bayesian analysis. Timing models were produced by `libstempo`[3], which is a `python` wrapper for the `tempo2`[4] timing package. We also make use of the `la_forge`[5] [154] package for some of our figures.

5.4.1 Model

The model we use to describe our dataset has many similarities with the one used in the `BayesHopper` algorithm. The only fundamental difference is that the collection of sinusoids are replaced by a set of "signal" wavelets coherent across detectors and a set of "noise transient" wavelets which only appear in a given pulsar and describe transient noise features in the data. In this section we give a brief description of the model and focus on differences between `BayesHopper` and `BayesHopperBurst`. More details can be found in [32].

Consider a PTA consisting of N pulsars. The ith residual in the kth pulsar of the array is modeled as:

$$\delta t_{ki} = \sum_l M_{kil}\delta\xi_{kl} + n_{ki} + g_{ki} + w_{ki}(\boldsymbol{\theta}_\mathrm{s}) + v_{ki}(\boldsymbol{\theta}_\mathrm{n}), \tag{5.1}$$

where $\delta\xi_{kl}$ is the offset from the best fit value of the lth timing model parameter, and M_{kil} is the design matrix, which represents a timing model linearized around the best fit parameter values. n_{ki} represents all noise processes unique to each pulsar, while g_{ki} is the contribution

[2] https://github.com/nanograv/enterprise
[3] https://github.com/vallis/libstempo
[4] https://bitbucket.org/psrsoft/tempo2/
[5] https://github.com/Hazboun6/la_forge

of an isotropic stochastic GW background (GWB) which is correlated between different pulsars. The contribution of the transient GW signal is represented by $w_{ki}(\boldsymbol{\theta}_{\rm s})$, while that of incoherent transient noise features is described by $v_{ki}(\boldsymbol{\theta}_{\rm n})$, where $\boldsymbol{\theta}_{\rm s}$ and $\boldsymbol{\theta}_{\rm n}$ represent the parameters describing the GW signal and the transient noise respectively.

5.4.1.1 Transient noise model The contribution of noise transients to the timing residual $(v_{ki}(\boldsymbol{\theta}_{\rm n}))$ is modeled as a sum of sine-Gaussian (Morlet-Gabor) wavelets:

$$v_{ki}(\boldsymbol{\theta}_{\rm n}) = \sum_{j=1}^{M_k} \Psi(t_{ki}; \boldsymbol{\lambda}_{kj}), \tag{5.2}$$

where M_k is the number of wavelets used in the kth pulsar, t_{ki} is the ith observing time of the kth pulsar, and $\boldsymbol{\lambda}_{kj}$ is the parameter vector describing the jth wavelet in the kth pulsar, which is related to the full parameter vector as $\boldsymbol{\theta}_{\rm n} = (M_1, \boldsymbol{\lambda}_{11}, ..., \boldsymbol{\lambda}_{1M_1}; ...; M_N, \boldsymbol{\lambda}_{N1}, ..., \boldsymbol{\lambda}_{NM_N})$, and the wavelets are defined as:

$$\Psi(t_i; \boldsymbol{\lambda}_{kj}) = Ae^{(t-t_0)^2/\tau^2} \cos(2\pi f_0 (t - t_0) + \phi_0), \tag{5.3}$$

where A is the amplitude, t_0 is the central time, τ is the characteristic duration, f_0 is the central time, and ϕ_0 is the initial phase. We can see that a single wavelet can be described by 5 parameters, i.e. $\boldsymbol{\lambda}_{kj} = (A, t_0, \tau, f_0, \phi_0)$. We use Morlet-Gabor wavelets as they have the compelling property of having the smallest time-frequency area allowed by the Heisenberg–Gabor limit [155]. Other functions, such as shapelets [156], might be better suited to certain types of signals, so we plan to investigate using them in the future.

5.4.1.2 Signal model In order to impose the proper coherence between pulsars, in the signal model we model the GW waveform itself as a sum of wavelets, and then project that onto the pulsar lines of sight taking into account the corresponding antenna factors. Since

PTAs are sensitive to the time integral of the metric perturbation, we introduce:

$$H_{+,\times}(t) = \int h_{+,\times}(t)\mathrm{d}t, \tag{5.4}$$

where $h_{+,\times}$ are the usual components of the metric perturbation corresponding to plus and cross polarized GWs. We model $H_{+,\times}(t)$ as:

$$H_{+}(t) = \sum_{j=1}^{N} \Psi(t; t_{0,j}, f_{0,j}, \tau_j, A_{+,j}, \phi_{0,+,j}), \tag{5.5}$$

$$H_{\times}(t) = \sum_{j=1}^{N} \Psi(t; t_{0,j}, f_{0,j}, \tau_j, A_{\times,j}, \phi_{0,\times,j}), \tag{5.6}$$

so that the two polarizations have independent amplitude and initial phase, but t_0, f_0, and τ are common parameters. This simplification will not affect the generality of our model, but helps in efficiency by reducing the number of parameters. Then a rotation around the propagation direction is allowed, which introduces the transformation:

$$\bar{H}_{+}(t) = H_{+}(t)\cos(2\psi) - H_{\times}(t)\sin(2\psi), \tag{5.7}$$

$$\bar{H}_{\times}(t) = H_{+}(t)\sin(2\psi) + H_{\times}(t)\cos(2\psi), \tag{5.8}$$

where ψ is the polarization angle. Then the timing residuals are formed with the use of the antenna pattern functions:

$$w_{ki}(\boldsymbol{\theta}_{\mathrm{s}}) = -F_{+}(\Omega_k, \Omega_{\mathrm{GW}})\bar{H}_{+}(t_{ki}; \boldsymbol{\lambda'}) - F_{\times}(\Omega_k, \Omega_{\mathrm{GW}})\bar{H}_{\times}(t_{ki}; \boldsymbol{\lambda'}), \tag{5.9}$$

where $\Omega_k = (\theta_k, \phi_k)$ and $\Omega_{\mathrm{GW}} = (\theta_{\mathrm{GW}}, \phi_{\mathrm{GW}})$ are the sky location of the kth pulsar and the GW source, respectively. $F_{+,\times}(\Omega_k, \Omega_{\mathrm{GW}})$ are the antenna pattern functions, and $\boldsymbol{\lambda'}$ contains the internal parameters of the GW signal.

Note that eq. (5.9) assumes that only the Earth term (see e.g. [157]) contributes to the signal. This simplification can be justified as follows: The time difference between the Earth and pulsar terms is given by

$$\Delta t = L(1 - \cos\beta) \approx 50 \text{ yr} \left(\frac{\beta}{10^\circ}\right)^2 \left(\frac{L}{1 \text{ kpc}}\right), \tag{5.10}$$

where L is the distance to the pulsar, β is the angle between the pulsar and the GW source sky location, and for the second equality we made an approximation valid for small β (which is where we can get the smallest delays). We can see that for typical millisecond pulsar distances of 1 kpc and observing campaigns not longer than a few decades, we will not see the pulsar term signals unless the sky location of the GW source is within a few degrees of a pulsar. Even if that happens, it will only appear in that single pulsar, so it can be modeled as a noise transient.

Note that pulsar term signals from GW bursts with no Earth term observed will appear as noise transients in our dataset. For an array with N_p pulsars, these correspond to an effective observation time N_p times larger than the actual observation time. However, the signal-to-noise ratio (SNR) of a pulsar term signal is a factor of $N_\mathrm{p}^{1/2}$ lower compared to the Earth term signal, since it appears in just a single pulsar. This smaller SNR corresponds to a decrease in the observable volume by a factor of $N_\mathrm{p}^{3/2}$. As a result, we expect that the rate of pulsar term bursts will be a factor of $N_\mathrm{p}^{1/2}$ lower than Earth term bursts, corresponding to a factor of 7 for the 45 pulsars in the NANOGrav 12.5-year dataset. Pulsar term signals can also be coherent between two (or a few) pulsars, but for an array with a few tens of pulsars, the expected rate of such events are even lower [149].

5.4.2 Sampler

As described above, the dimensionality of our model depends on whether the GWB is turned on or not, and also on how many wavelets are used in the signal and the transient noise model. A GWB can carry 1 or 2 parameters depending on whether the spectral slope of the power-law model is fixed or varied. Each noise transient wavelet carries 6 parameters including the 5 internal parameters plus an “external” parameter indexing which pulsar the noise transient belongs to. The signal model has 3 external parameters common to all wavelets (sky location and polarization angle) plus 7 internal parameters for each wavelet.

To sample this variable-dimension parameter space we use an RJMCMC sampler called `BayesHopperBurst`, which not only gives posterior distributions on the model parameters, but also on the number of different components included in the model. `BayesHopperBurst` uses many of the proposals developed for `BayesHopper` such as Fisher matrix proposals, parallel tempering and trans-dimensional proposals. The description of these can be found in [32]. We also need a global proposal to help convergence, however the one used in `BayesHopper` is specific to continuous wave sources, so instead we use a global proposal based on the τ-scan, which we describe below.

The τ-scan is defined as:

$$\mathcal{T}\left(\{x_i\}; t_0', f_0', \tau'\right) = \sum_{i=1}^{N_\mathrm{p}} |(x_i|\Psi_{c,i})_i|^2 + |(x_i|\Psi_{s,i})_i|^2, \tag{5.11}$$

where x_i is the time series in the ith pulsar, N_p is the number of pulsars in the array, $(.|.)_i$ is the noise-weighted inner product in the ith pulsar, $\Psi_{c,i} = \Psi(A = \bar{A}_i, t_0 = t_0', f_0 = f_0', \tau = \tau', \phi_0 = 0)$ is a “cosine wavelet” and $\Psi_{s,i} = \Psi(A = \bar{A}_i', t_0 = t_0', f_0 = f_0', \tau = \tau', \phi_0 = \pi/2)$ is a “sine wavelet”, where $\bar{A}_i$ and $\bar{A}_i'$ are normalization constants chosen for each pulsar so that $(\Psi_{c,i}|\Psi_{c,i}) = (\Psi_{s,i}|\Psi_{s,i}) = 1$.

By calculating $\mathcal{T}$ for various t_0', f_0', and τ' values on a grid, we can get a 3D map showing where the data has some excess power. To determine the necessary spacing between grid points so that they properly cover the parameter space, we calculate the match (M) between two neighboring wavelets and require that it is above some threshold, e.g. $M > 0.9$.

The match between two waveforms (h and h') is defined as:

$$M = \frac{(h|h')}{\sqrt{(h|h)(h'|h')}}, \tag{5.12}$$

where $(.|.) = \sum_{i=1}^{N_{\mathrm{p}}}(.|.)_i$ is the array-wide noise-weighted inner product. M is normalized so that $-1 \leq M \leq 1$, where $M = 1$ indicates that h and h' are exactly the same, while $M = -1$ means there is a perfect anticorrelation between them. We can approximate M for two wavelets close to each other as [112]:

$$M = 1 - \frac{1}{4\bar{\tau}^2}(\Delta\tau)^2 - \frac{1}{2\bar{\tau}^2}(\Delta t_0)^2 - \frac{\pi^2\bar{\tau}^2}{2}(\Delta f_0)^2, \tag{5.13}$$

where Δ indicates the difference in the given parameter between the two wavelets and $\bar{\tau} = (\tau_i + \tau_j)/2$ is the average τ between the two wavelets.

We can then require that the mismatch $(1-M)$ when moving only one of the parameters be less than 0.1. This implies:

$$\Delta \ln \tau \leq \sqrt{\frac{2}{5}}, \tag{5.14}$$

$$\Delta t_0 \leq \frac{\tau}{\sqrt{5}}, \tag{5.15}$$

$$\Delta f_0 \leq \frac{1}{\sqrt{5}\pi\tau}. \tag{5.16}$$

We see that the τ layers should be logarithmically spaced with each layer having a τ value maximum $\exp(\sqrt{2/5}) \approx 1.9$ times larger than the previous one. Time and frequency should be linearly spaced, with the spacing depending on the τ of the given layer. One can always choose to make the spacing finer, but that results in increasing computational costs.

Figure 5.1 shows the τ-scan calculated for a white noise burst signal (see Section 5.5.1) with central time and frequency marked with the white crosses. Here we use 5 layers with different τ values. We can see that as a result of the τ-dependent time and frequency spacing, pixels at higher τ values are more elongated in time but narrower in frequency (resulting in the same time-frequency area). We can see that the τ-scan clearly lights up around the injected signal indicating that, after proper normalization, this provides an excellent proposal. Indeed, we find that including this proposal speeds up convergence by multiple orders of magnitude, even though the remaining parameters of the wavelet are drawn from their priors. This proposal is also used when proposing to add a new signal wavelet, which makes it possible to have a reasonable acceptance rate for these complicated jumps.

A variant of the τ-scan proposal is used for noise transient wavelets. For these we create a τ-scan for each pulsar only using the data for the pulsar in question. These are not only used to draw f_0, t_0 and τ values for the wavelet, but also to draw which pulsar to put in the wavelet based on how much excess power is in each pulsar.

5.5 Tests on simulated data

To test `BayesHopperBurst` we analyze a few different simulated datasets. All the datasets in this section correspond to a simulated PTA with 20 pulsars at random sky locations. Each has been observed for 10 years every 30 days, has the same constant white noise level of 0.5 μs and no red noise. We add different kinds of GW signals and noise transients to this base dataset. In Section 5.5.1 we show results for datasets only containing GW signals and not noise transients; in Section 5.5.2 we look at a scenario with no GW signals but noise transient in many pulsars of the array; while in Section 5.5.3 we look at a few examples of how `BayesHopperBurst` can deal with datasets containing both GW signals and noise transients.

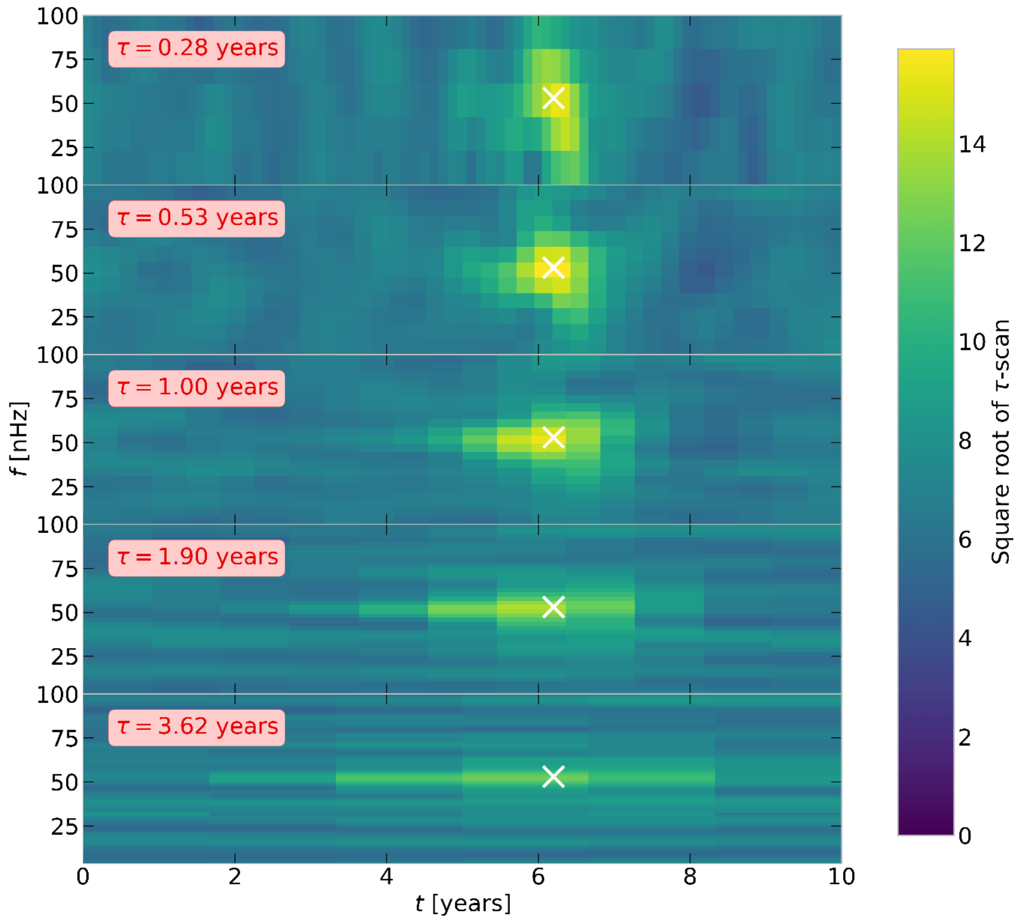


Figure 5.1: Tau-scan map for a simulated white noise burst signal with central time and frequency marked by the white crosses. Note how the map peaks around the true time-frequency location of the signal, suggesting that such a proposal will help finding the signal quickly.

5.5.1 Signal without noise transients

First we look at a dataset with a GW signal that consists of two sine-Gaussian wavelets that are close to each other in the time-frequency plane. This can be considered as one of the easiest signals to reconstruct, since it matches our model, i.e., we can perfectly reconstruct such a signal with two signal wavelets. Figure 5.2 shows the reconstruction of this signal in one of the pulsars in the array. We can see that `BayesHopperBurst` uses no noise transient wavelets and two signal wavelets most of the time, as expected. We can also see that the reconstructed signal matches well the injected GW signal. Note that while Figure 5.2 only shows the reconstruction in a particular pulsar, the projection of the signal in the datastream of other pulsars can be significantly different. The reconstructions for all 20 pulsars in the array are shown in Figure D.1, indicating similarly accurate waveforms in all pulsars. To quantitatively characterize the goodness of fit throughout the whole array, we calculate the match between the median reconstructed signal and the injected signal, which we find to be 0.96. The SNR of the simulated signal is 17.9, which we define as $\mathrm{SNR} = \sqrt{(h_{\mathrm{inj}}|h_{\mathrm{inj}})}$, where h_{inj} is the waveform of the simulated (or injected) signal.

The next dataset contains a white noise burst (WNB), which is produced from a white noise time series by applying a Gaussian window both in time and frequency. The variance of the white noise sets the amplitude of our WNB, while the windowing determines its time-frequency location and spread. While a WNB is similar to a sine-Gaussian wavelet in that it is also well localized on the time-frequency plane, it has a more complicated time dependence. This also means that a WNB cannot be perfectly reproduced by a finite number of sine-Gaussian wavelets, so the number of wavelets used will be determined by the trade-off between getting an increasingly better fit and paying an Occam penalty for each additional wavelet included.

Figure 5.3 shows the reconstruction of a WNB signal. We can see that `BayesHopperBurst` uses a single wavelet to fit the majority of the signal, while low-amplitude parts of the signal are missed. This also leads to a lower match between the median reconstructed signal and the injected signal (0.812) compared to the previous example, even though it has a slightly higher SNR of 18.3. This suggests that `BayesHopperBurst` performs slightly worse for a WNB signal compared to sine-Gaussians. Note however that if we take into account the

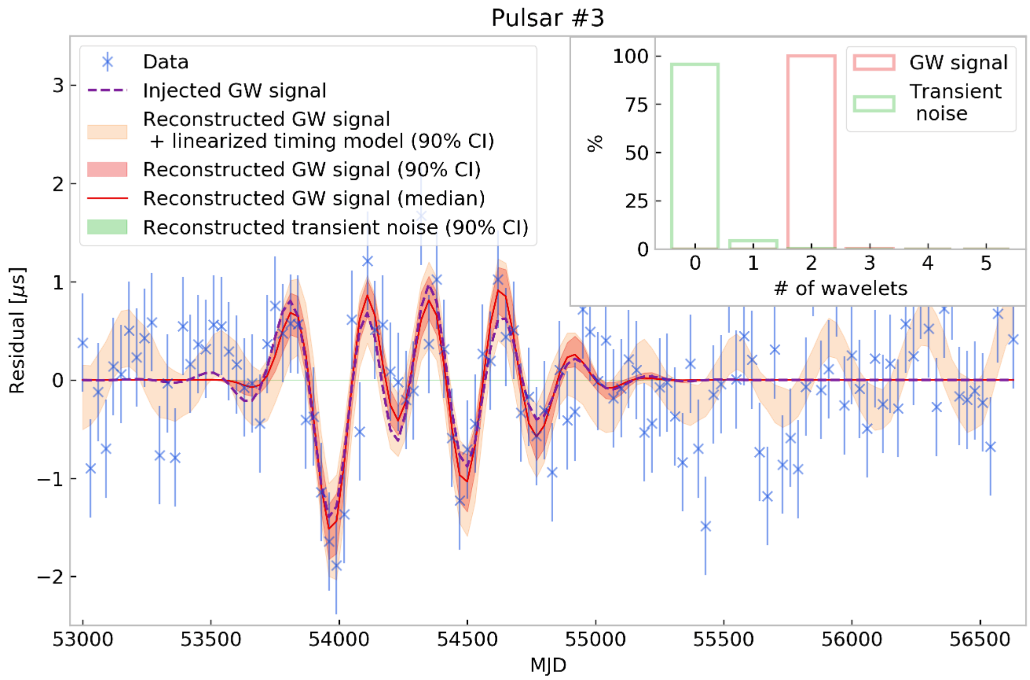


Figure 5.2: Reconstruction of a signal consisting of two sine-Gaussian wavelets (SNR = 17.9) in one of the pulsars in the array. The reconstruction is more precise than one would expect by looking at the spread of data points because the reconstruction is based on data from all 20 pulsars in the array. The match between the median reconstructed waveform and the injected waveform is 0.96. The inlet plot shows the histogram of the number of wavelets used both in the signal and the noise transient model. We can see that `BayesHopperBurst` uses two signal wavelets and no noise transients as expected.

uncertainty in the linearized timing model (orange band on Figure 5.3), the reconstruction is consistent with the injected GW signal at the 90% confidence level.

One of the potential sources of GW bursts in the PTA band are SMBHBs on parabolic or highly eccentric orbits. Here we analyze the simulated signal from a parabolic encounter of two SMBHs. We calculate the waveform from eq. (4-3) of [146]. Beyond being a potential astrophysical source, parabolic SMBHBs are also a good test of the algorithm, because they are not elliptically polarized, so `BayesHopperBurst` must use its flexibility to fit two very different waveforms for the plus and cross polarizations.

Figure 5.4 shows the waveform reconstructed by `BayesHopperBurst` for an equal mass

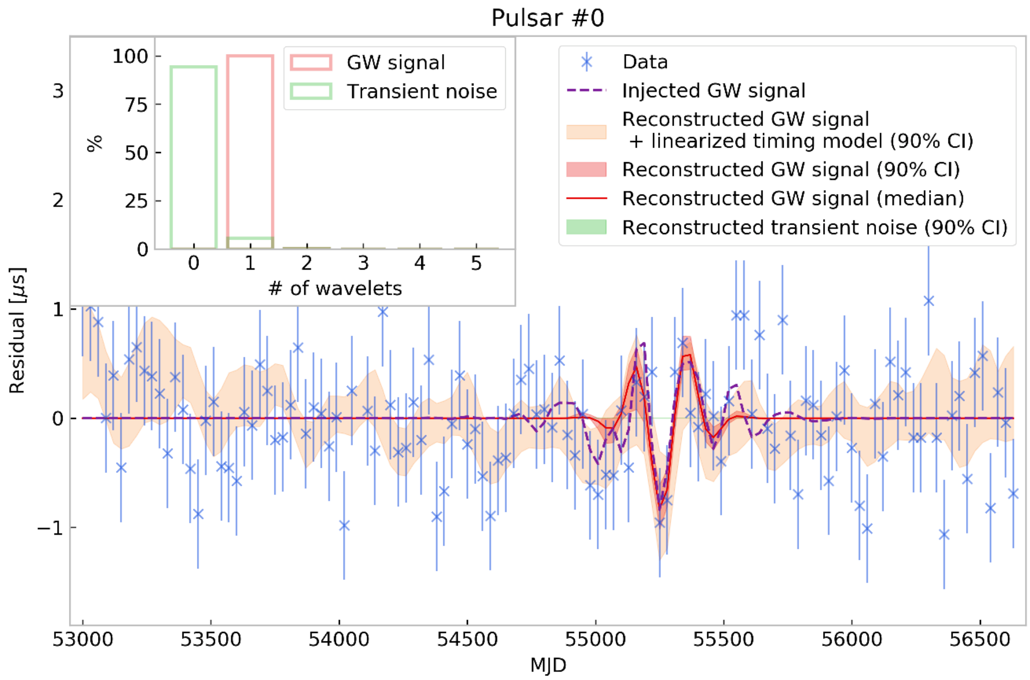


Figure 5.3: Reconstruction of a white noise burst signal (SNR = 18.3) in one of the pulsars of the array. The match between the median reconstructed waveform and the injected waveform is 0.812. We can see that parts of the signal are modeled by the linearized timing model. The inlet plot shows the histogram of the number of wavelets used both in the signal and the noise transient model. We can see that `BayesHopperBurst` uses a single signal wavelet to model the WNB.

SMBHB on a parabolic orbit with a total mass of $M = 2 \times 10^9 M_\odot$, an impact parameter of $b = 60M$, at a distance of 15 Mpc, with an SNR of 16.8. Unlike wavelets or WNBs, these parabolic encounter waveforms have a significant low-frequency component. We can see that the linearized timing model can fit this component with a quadratic function by changing the period and period derivative parameters of the timing model. The high-frequency (and high-amplitude) part of the signal is reconstructed by a single wavelet. The match between the median reconstructed signal and the injected signal is 0.956.

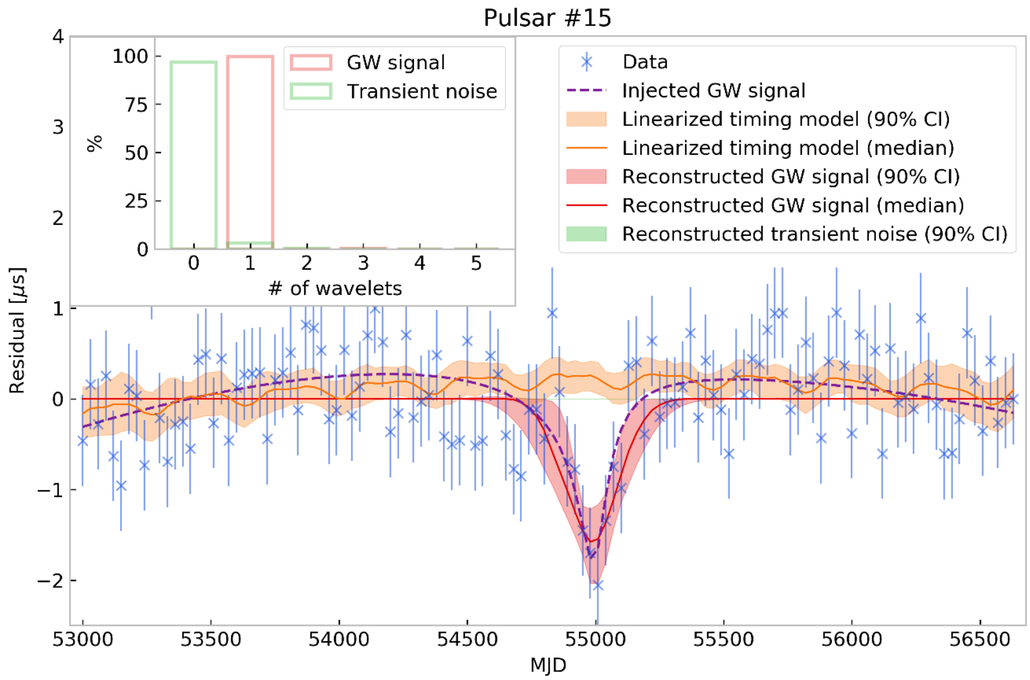


Figure 5.4: Reconstruction of a parabolic BBH signal (SNR = 16.8) in one of the pulsars of the array. The BHs have equal masses of $10^9 M_\odot$, and are on an orbit with an impact parameter of 180 $M_\odot$ at a distance of 15 Mpc. The match between the median reconstructed waveform and the injected waveform is 0.956. We can see that the low-frequency part of the signal is modeled by the linearized timing model, while the high-frequency part is reconstructed using a single signal model wavelet, as can be seen on the inlet plot showing the histogram of the number of wavelets used.

5.5.2 Multiple noise transients

We have seen that `BayesHopperBurst` is able to reconstruct GW bursts in the presence of Gaussian noise. In this section, we test if it can correctly identify incoherent noise transients. We consider four different datasets: i) with a single pulsar containing a sine-Gaussian noise transient; ii) with 3 pulsars containing the same noise transient; iii) with 10 pulsars (half the array) containing the same noise transient; and iv) all 20 pulsars containing the same noise transient. Having the same noise transient in many pulsars can be considered a worst case scenario for discriminating signals and noise transients. It is virtually impossible to have such a situation by chance alignment of incoherent noise transients, but it provides a good test case. Note that even if all the pulsars have the exact same noise transient, in principle, that can be distinguished from a coherent GW burst, because there is no sky location and polarization that would result in the exact same waveform (with the same amplitude) in each pulsar.

Figure 5.5 shows the histogram of the number of wavelets used in the signal model and the noise transient model for each of these datasets. We can see that for the first three datasets we are recovering the correct model of having 1/3/10 noise transient wavelets and no signal wavelets. For the dataset with all the pulsars containing the same noise transient we instead see that `BayesHopperBurst` is using a single signal wavelet and 12 to 15 noise transient wavelets. This can be understood as a result of the inbuilt parsimony of Bayesian statistical analysis. As noted above, there are no such external parameters that could result in the exact same amplitude in the datastream of all the pulsars. However, one can use the signal model to account for noise transients in a few pulsars, and use noise transient wavelets to model the rest. A single-wavelet signal model has 10 parameters, while each noise transient wavelet has 5 parameters. Thus such a solution could be statistically favored if the signal model can account for noise transients in more than two pulsars. Looking at the number of noise transient wavelets used, we can see that in this example the signal model is able to mimic the residuals from up to 8 noise transients.

5.5.3 Signal and noise transients

We have seen how `BayesHopperBurst` can reconstruct signals or noise transients. In this section we test it on datasets containing both signals and noise transients to see how it is able to distinguish between the two. First we analyze a dataset containing a signal and a noise transient in one of the pulsars, both of which have sine-Gaussian waveforms but with significantly different t_0, f_0 and τ parameters. We produce two versions of this dataset with different amplitudes: i) with signal SNR of 26.8, and noise transient SNR of 10; ii) with

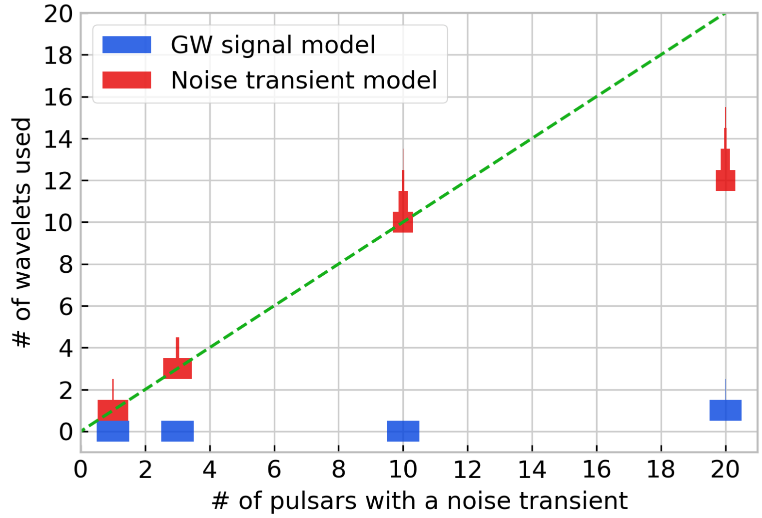
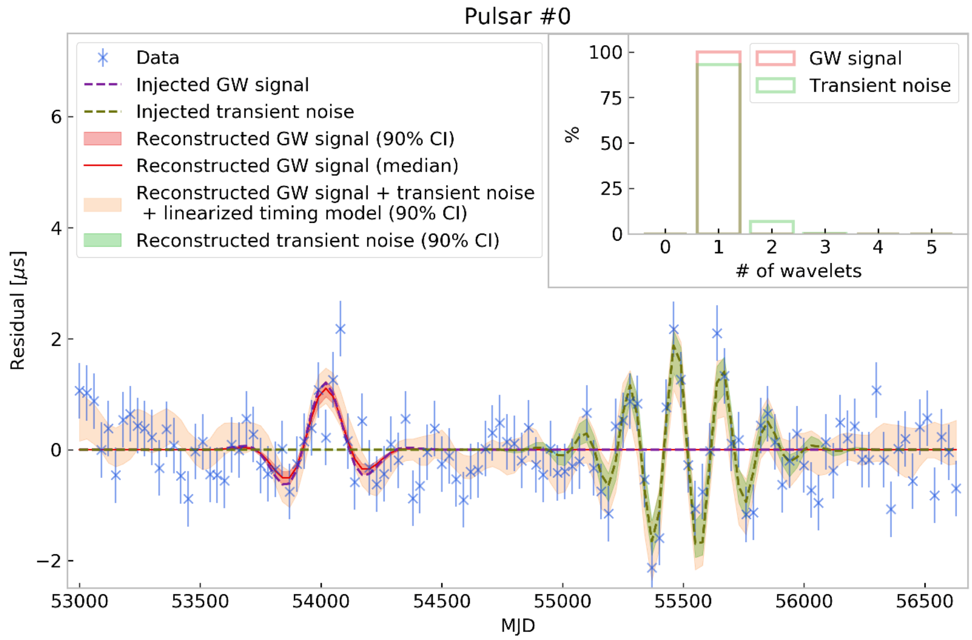


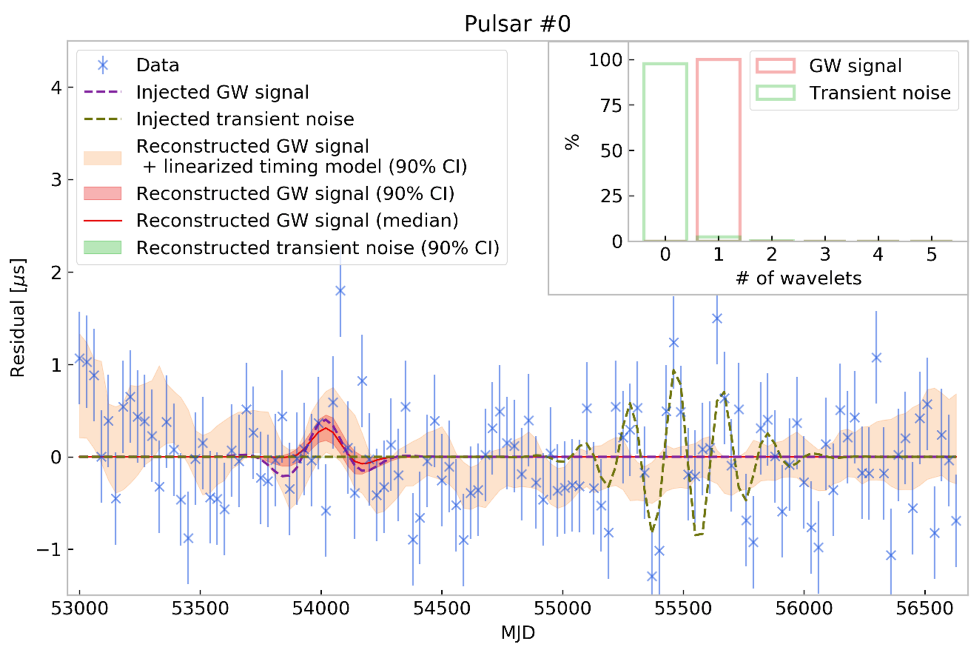
Figure 5.5: "Discrete violin plot" of number of wavelets used in signal and noise transient model for datasets with the exact same noise transient present in a different number of pulsars. We can see that when the noise transient is present in not more than half of the pulsars, the model is correctly identified. However, when the noise transient is present in all the pulsars, the sampler instead turns on a signal model and finds a sky location which takes care of most of the signal and fits the rest with noise transient wavelets.

signal SNR of 8.9, and noise transient SNR of 5. The reconstructions for these datasets are shown in Figure 5.6. We can see that both the signal and the noise transients are correctly identified and reconstructed for the higher amplitude case, and we find $M = 0.994$ between the injected and recovered waveforms. For the lower amplitude case only the signal has been found with a moderate match of 0.87. Note however that both the relatively lower match value and the fact that no noise transient has been found are consistent with the significantly lower SNRs of both signal and noise transient in this dataset.

In the next example, we create a dataset with a sine-Gaussian signal ($\mathrm{SNR} = 16.1$) and a noise transient ($\mathrm{SNR} = 10.8$) in one of the pulsars with the exact same waveform as the signal. This can be considered a worst case scenario, since `BayesHopperBurst` can only rely on the coherence of the signal when trying to distinguish GW signal and noise transient. We make



(a)



(b)

Figure 5.6: Reconstruction of signal and noise transient well separated in time. In the case of high amplitudes (a), both signal (SNR = 26.8) and noise transient (SNR = 10) are accurately reconstructed resulting in a match of 0.994 between the median reconstructed and injected waveforms. In case of low amplitudes (b), only the signal (SNR = 8.9) is reconstructed by `BayesHopperBurst` with $M = 0.87$, while the noise transient (SNR = 5) is found to be insignificant. This is not surprising given the low SNR of the noise transient.

two versions of this dataset with the noise transient being in two different pulsars. Figure 5.7a shows the reconstruction for the case where the noise transient is in a pulsar far away from the GW source on the sky as shown by the sky location posterior map (see Figure 5.7b). We can see that both the waveforms and the sky location is accurately reconstructed in this case ($M = 0.98$). However, having a noise transient in a pulsar that lies close to the true GW source sky location results in no significant noise transient reconstructed (see Figure 5.7c). Instead, `BayesHopperBurst` fits the data by changing the sky location slightly to account for the effect of the noise transient (see corresponding skymap in Figure 5.7d). Since the noise transient is in one of the pulsars closest to the GW source on the sky, the location of the GW source can be changed slightly in a way that it can account for the noise transient in that pulsar but not change significantly the projection of the signal in other pulsars. This reconstruction results in a match of 0.96, which is only slightly lower than the previous example, even though it is not using the true model.

5.6 Upper limit predictions for the NANOGrav 12.5-year dataset

In a follow-up study, we plan to use `BayesHopperBurst` to search for nHz GW bursts in the NANOGrav 12.5-year dataset [3]. To assess the sensitivity we might expect from such an analysis we run `BayesHopperBurst` on a simulated dataset made to resemble the NANOGrav 12.5-year dataset and used in [12]. The dataset has been made using the properties (timing parameters, white and red noise parameters) of pulsars and observational time stamps from the NANOGrav 12.5-year dataset. Note that the red noise parameters are derived by filtering out a common red noise process, so that they truly characterize pulsar intrinsic red noise. To reduce computational costs the residuals are epoch averaged. More details on this simulated dataset can be found in [12].

This dataset does not contain any GW bursts, so by running `BayesHopperBurst` on it we can set upper limits on the amplitude of generic GW transients. We consider these

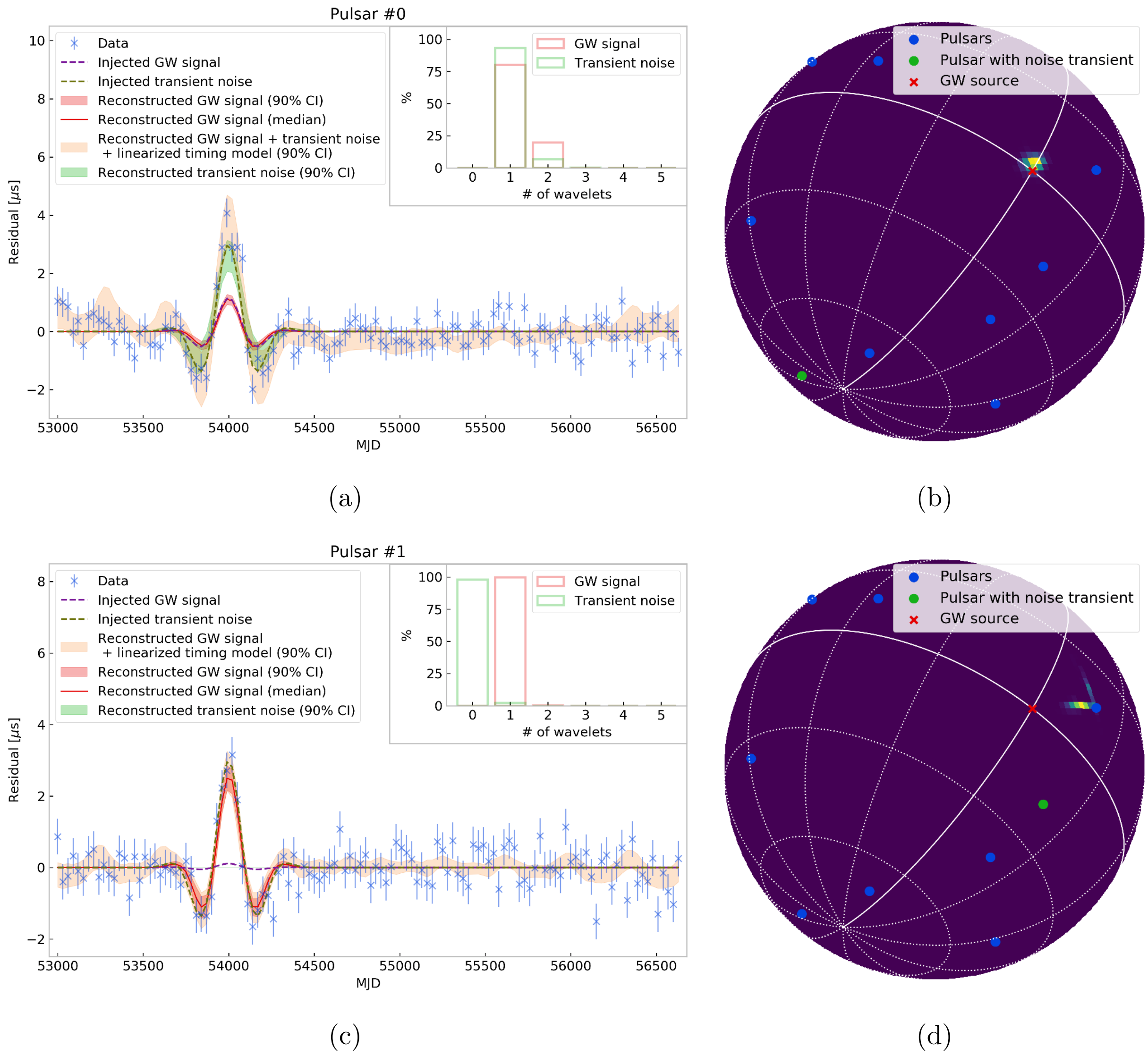


Figure 5.7: Reconstruction of a dataset with a sine-Gaussian GW signal (SNR = 16.1) and a noise transient (SNR = 10.8) with the exact same waveform. Both noise transient and signal are correctly reconstructed when the noise transient is in a pulsar far away from the GW source on the sky (a-b). However, when the noise transient is in a pulsar close to the GW signal location, the effect of the noise transient can be accounted for by slightly changing the sky location of the GW without affecting the signal in other pulsars (c-d).

upper limits to be indicative of the upper limits we might expect to get from running on the real NANOGrav 12.5-year dataset assuming we do not find any significant GW candidates. Figure 5.8 shows the uppper limits we get from this simulated dataset as a function of central time (t_0) and central frequency (f_0) defined e.g. by equations (8a) and (8b) in [158], which coincide with the parameters used to describe the wavelets if only 1 wavelet is used. We quote these upper limits in terms of the root-sum-squared strain amplitude defined as:

$$h_{rss} = \sqrt{\int_{-\infty}^{\infty} [h_+^2 + h_\times^2] \, \mathrm{dt}} = \sqrt{\int_{-\infty}^{\infty} \left[\left(\frac{\mathrm{dH}_+}{\mathrm{dt}} \right)^2 + \left(\frac{\mathrm{dH}_\times}{\mathrm{dt}} \right)^2 \right] \mathrm{dt}}. \tag{5.17}$$

To produce this plot we calculate t_0, f_0 and h_{rss} for each sample from our MCMC. Then we bin the samples based on a grid on the t_0–f_0 plane, and in each bin we calculate the 95th percentile of the h_{rss} samples from that bin, which we quote as the 95% upper limit.

A clear trend from Figure 5.8 is that the upper limit decreases as we move to later times. This is because the array contains more and more pulsars as time goes on, which greatly improves the array's sensitivity. In addition, we see drastic improvements in the second half of the dataset, which can be explained by the improved backends used after ∼7.5 year and a high-cadence observing campaign, which started at ∼8.5 year (see e.g. [3]). We can also see that the sensitivity decreases at the temporal edges of the dataset, which is due to the fact that waveforms with a central time near the edges will have a significant portion that lies outside of the observational time span of the dataset. The worst sensitivity in the analyzed time-frequency region is achieved at very high frequencies and very early times. This is due to the fact that there are less than 10 pulsars observed at those times. There is also a slight increase in the upper limit around the frequency corresponding to a period of 1 year ($\log_{10} f_0 \simeq -7.5$), where PTAs are known to be less sensitive due to the sky location fitting in the timing model.

The best sensitivity is achieved around $t_0 \simeq 10$ years and $f_0 \simeq 8$ nHz at an amplitude

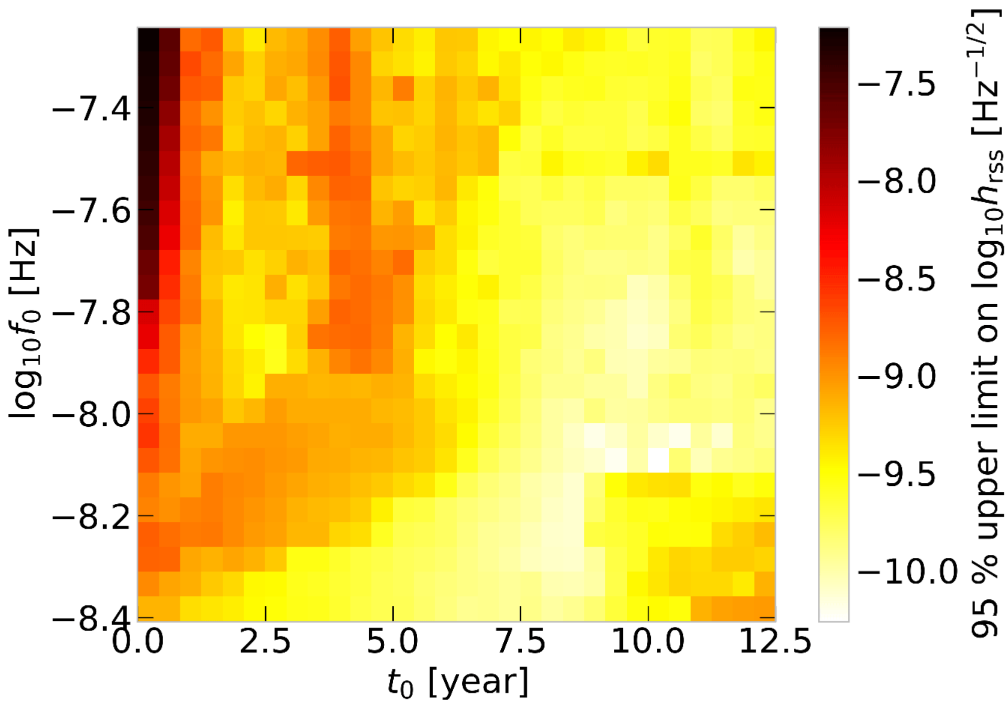


Figure 5.8: Upper limit on the root-sum-squared strain amplitude (h_{rss}) as a function of central time (t_0) and central frequency (f_0) based on a simulated dataset made to resemble the NANOGrav 12.5-year dataset [12]. The most sensitive time-frequency location is at around 10 years – 8 nHz where we get an upper limit of $h_{rss} \simeq 5 \times 10^{-11}$ Hz$^{-1/2}$, which corresponds to $\sim 40 M_{\odot} c^2$ emitted in GWs at a fiducial distance of 100 Mpc.

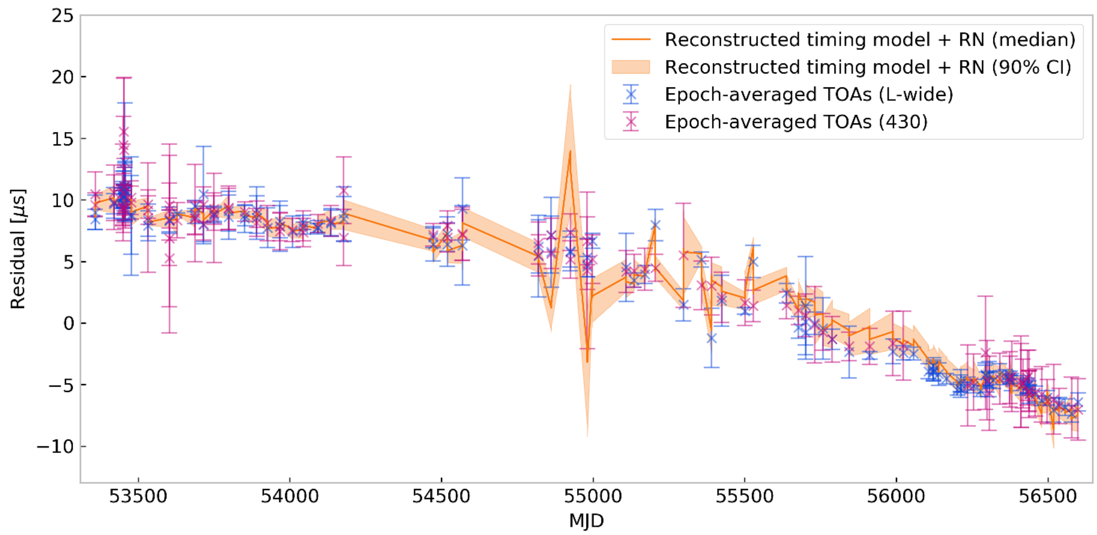
of $h_{rss} \simeq 5 \times 10^{-11}$ Hz$^{-1/2}$. This can be converted into the minimum GW energy that can be detected at a fiducial distance using e.g. equation (4.2) of [159]. We get that at a fiducial distance of 100 Mpc we are sensitive to sources emitting an energy of more than $\sim 40 M_{\odot} c^2$ in GWs.

Note that we use a uniform prior on the wavelet amplitude which pushes up the amplitude values resulting in a conservative upper limit. However, as we also allow the sky location to change, and the distribution of the pulsars is highly anisotropic, this also pushes the sampler to the least sensitive sky locations where the highest amplitude signals can remain hidden. This means that the sensitivity at the best sky location could be significantly better than the value quoted above. On the other hand, in this run we fix the red noise parameters in each pulsar to the true values due to computational constraints. This in principle could make our upper limit estimates overly optimistic as it does not take into account any covariances between the red noise and GW burst signal models. However, since the red noise only dominates at the lowest frequencies, we do not expect this to have a drastic effect at the most sensitive time-frequency location reported above. In addition, even while fixing the red noise parameters to their true values, we still allow the realization of the red noise to vary, which can accommodate some covariance between the signal and red noise models.

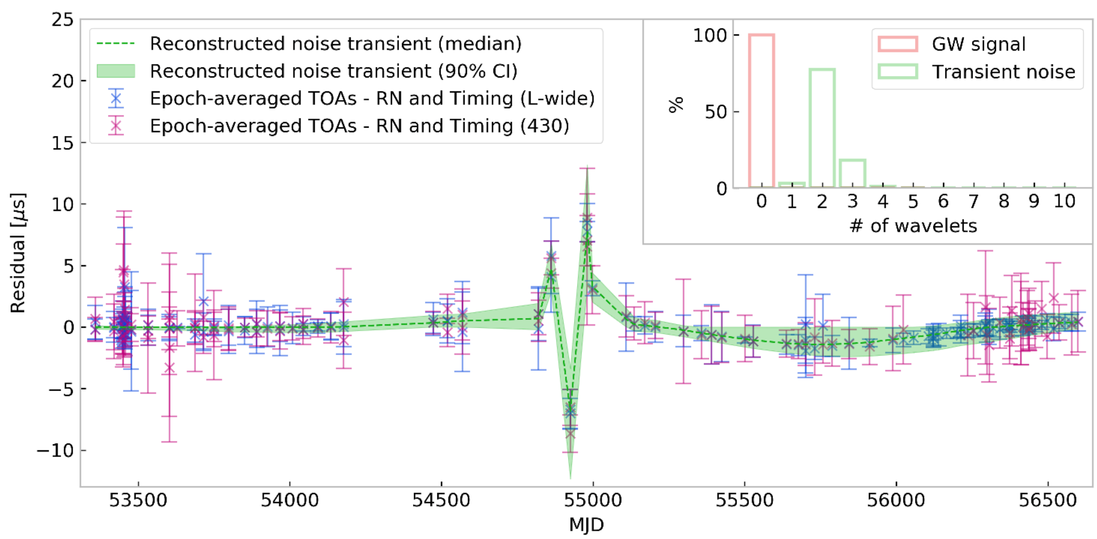
5.7 Single-pulsar test on real data

Real data can bring complications compared to simulated data like uneven sampling, temporal gaps in the datastream, chromatic effects, etc. To test how `BayesHopperBurst` can handle these complications we run it on a single pulsar (B1855+09) from the NANOGrav 9-year dataset [153]. We show the epoch-averaged residuals from B1855+09 in Figure 5.9a along with the reconstruction of the linearized timing model and the red noise model. Figure 5.9b shows the epoch averaged residuals corrected with the median timing and red noise model. These indicate a transient feature around 55000 Modified Julian Date (MJD), which is reconstructed by the noise transient model also shown on this figure. Note that although we show the epoch-averaged residuals for plot clarity, our analysis used all individual residuals.

We can see in Figure 5.9 that `BayesHopperBurst` adjusts the timing parameters such that some features of the data were smoothed out while others were amplified. Then it used a few wavelets to correct those amplified features. In particular, we can see that the observations made at different frequencies at the same epoch are sometimes inconsistent in the original dataset. One can correct that by changing the value of the dispersion measure



(a)



(b)

Figure 5.9: Waveform reconstruction for B1855+09 from the NANOGrav 9-year dataset. (a) shows the original epoch-averaged residuals along with the reconstructed timing model and red noise model realizations. (b) shows the epoch-averaged residuals corrected by the median red noise and timing model realizations, along with the noise transient reconstruction from `BayesHopperBurst`. The latter suggests that there is a noise transient in this dataset around MJD 55000. Note that there is a strong covariance between the presence of the noise transients and DMX parameters.

parameter (DMX, see e.g. [153]) for that epoch, but that will result in a net residual in both observations. The legacy analysis thus cannot do that, since it cannot account for transient noise features. On the other hand, `BayesHopperBurst` can change the dispersion measure values and model the resulting noise transient with a collection of wavelets. Thus, our analysis suggests that there is a noise transient in this dataset. Such a noise transient was not found before, since the canonical analysis does not model noise transients and their presence can be masked by inflating the white noise parameters. In fact, `BayesHopperBurst` only finds the noise transient when we allow the white noise parameters to vary, so that it can lower their values compared to the canonical analysis.

Based on Figure 5.9, it might look like that the feature around 55000 MJD fitted by the noise transient model was introduced by the similar feature with opposite sign in the timing and red noise model. Note however that the linearized timing model introduces perturbations around timing model parameters previously determined by an analysis which does not allow for unmodeled noise transients. Such an analysis will try to adjust its free parameters to model out any noise transients present in the dataset. Our results indicate that if we introduce a model for noise transients, the Bayesian analysis prefers to undo that timing model adjustment, and instead model the feature with the dedicated noise transient model.

5.8 Conclusion

We have presented `BayesHopperBurst`, a Bayesian algorithm to search for nHz GW burst using PTA data, by modeling them as a collection of sine-Gaussian wavelets, where the number of wavelets is found by a trans-dimensional RJMCMC sampler. We demonstrated how it can reconstruct signals with a wide range of morphologies, and also how it can distinguish between coherent GW burst and incoherent noise transients that may occur in the datastream of some pulsars.

In the near future, we plan to use `BayesHoperBurst` to carry out the first ever search for generic GW bursts in the nHz frequency regime using the NANOGrav 12.5-year dataset. In preparation for that we analyzed a simulated dataset similar to the NANOGrav 12.5-year dataset and made predictions on the sensitivity we might expect from such a search. To demonstrate that `BayesHopperBurst` is capable of dealing with the additional complications arising when analyzing real data, we analyzed a single pulsar (B1855+09) from the NANOGrav 9-year dataset.

In addition to searching for generic GW bursts, an algorithm like `BayesHopperBurst` could also be a useful tool for analyzing transient noise features in the datastream of single pulsars as illustrated in [150]. We are planning to further explore the potential of such an approach in a follow-up study. Systematically analyzing a large collection of pulsars will also help us better understand the covariances we have seen in Section 5.7 between transient noise, DMX, and red noise models.

Important areas to investigate in future studies are further development and optimization of `BayesHopperBurst`. One particularly interesting direction would be experimenting with different types of functions in our signal and noise transient models (e.g. shapelets [156]), which might be more suited for certain signals or noise transients.

ACKNOWLEDGEMENTS

The authors thank David Nice and Justin Ellis for informative discussions about B1855+09, and Sarah Vigeland for feedback on the manuscript. We appreciate the support of the NSF Physics Frontiers Center Award PFC-1430284. Some computations were performed on the Hyalite High Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure at Montana State University.

CHAPTER SIX

CONCLUSION

Pulsar timing arrays offer a new way of observing the Universe through nanohertz gravitational waves. They will provide astrophysical information complementary both to electromagnetic observations and to high-frequency gravitational waves observed by ground-based interferometric detectors. In particular, this frequency band is expected to contain signals from binaries of the largest supermassive black holes residing at the center of essentially all galaxies. Thus pulsar timing arrays have the potential to help us understand how these black holes form and evolve through cosmic time. To make best use of this new discovery space, we need a good understanding of how different GW signals appear in our data and how we can optimize our algorithms to robustly find and characterize these sources.

In Chapter 2 we explored how the collection of hundreds of millions of supermassive balck hole binaries affect pulsar timing array datasets. The vast majority of these sources will be too faint to be detected individually, but they will combine to form a stochastic gravitational-wave background. In addition, the few brightest binaries might have a strong enough signal to be detected individually. We studied realistic datasets based on synthetic catalogs of supermassive balck hole binaries that holistically model the overall signal and include both of these contributions. We showed that such datasets exhibit several differences compared to the simple model of an isotropic Gaussian stochastic gravitational-wave background with a power-law spectrum. These are due to the finite number of binaries dominating the signal, which results in the breakdown of the central limit theorem. These effects can be used to discern the origin of the gravitational-wave background, since other astrophysical processes, which produce the background in the early universe, would result in a perfectly isotropic and Gaussian background. Thus detecting an individual binary or the anisotropy of the background would be evidence for the background originating from black hole binaries. We also make predictions about the detectability of individual binaries in the presence of the confusion noise from the stochastic background for a simulated pulsar timing array based on the NANOGrav 12.5-year dataset with a timespan extended to 15 years. We find that about 8% of the realizations produce a source with signal-to-noise ratio higher than 5, which might correspond to a detectable signal depending on the details of the data analysis used. These can be taken as pessimistic estimates for the upcoming NANOGrav 15-year dataset, since they do not take into account any potential improvements of the timing model solutions or the addition of new pulsars relative to the 12.5-year dataset. This shows that while the chances are relatively low for the upcoming NANOGrav 15-year dataset, we might not be that far away from detecting gravitational waves from an individual supermassive black hole binary (the probability of a source with a signal-to-noise ratio larger than 5 increases to about 32% at 20 years).

To make such a detection possible we presented a new Bayesian search for individual binaries, which results in a significant speedup compared to the methods currently in use (see Chapter 3). Our formulation does not apply any new approximations. Instead, the increased efficiency comes from 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. These methods will drastically reduce the computational cost of searches for individual sources in upcoming pulsar timing array datasets, and improve the tractability of achieving a well-converged analysis as the size of datasets increase. They will also make more sophisticated analyses computationally feasible, like searching for eccentric binaries, or simultaneously searching for multiple binaries.

As our datasets improve, we can also expect to be able to detect multiple individual binaries. In preparation for that, we described the first method to jointly search for multiple individual supermassive black hole binaries and an isotropic stochastic background in Chapter 4. We use a trans-dimensional Reversible Jump Markov Chain Monte Carlo sampler to explore models with different numbers of binaries and with or without a stochastic background. We found that the significance of individual sources can decrease in presence of a stochastic background both because it increases the noise levels at low frequencies, and because the stochastic background model can partially fit the signal of the individual sources. This highlights the importance of such joint analyses, which will be important in the future to avoid biasing the parameter estimation of the stochastic background by the presence of an individual binary, or vice versa.

In addition to signals from anticipated sources, such as those from supermassive black hole binaries, some previously unimagined sources may emit gravitational wave bursts with unknown morphology in the nanohertz frequency band. To find these potential sources we introduced a Bayesian algorithm to search for them using pulsar timing array data (see Chapter 5). We model signals as a collection of sine-Gaussian wavelets, where the number of wavelets is found by a trans-dimensional Reversible Jump Markov Chain Monte Carlo sampler. We demonstrated how it can reconstruct signals with a wide range of morphologies, and also how it can distinguish between coherent gravitational-wave signals and incoherent noise transients that may occur in the datastream of some pulsars. This algorithm will be an invaluable tool not only to carry out the first search for unmodeled signals in the nanohertz frequency band, but also to search for and mitigate potential noise transients present in pulsar timing array datasets.

The next decade will bring a confident detection of nHz GWs, opening a new window onto the Universe. In this thesis, we presented several new methodologies which will help us realize the full detection potential of the emerging field of low-frequency GW astronomy.

REFERENCES CITED

[1] Abbott, B.P. et al. Observation of Gravitational Waves from a Binary Black Hole Merger. Phys. Rev. Lett., 116(6):061102, February 2016. doi: 10.1103/PhysRevLett.116.061102.

[2] The LIGO-Virgo-KAGRA collaborations. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv:2111.03606*, art. arXiv:2111.03606, November 2021.

[3] Alam, M.F. et al. The NANOGrav 12.5 yr Data Set: Observations and Narrowband Timing of 47 Millisecond Pulsars. ApJS, 252(1):4, January 2021. doi: 10.3847/1538-4365/abc6a0.

[4] Desvignes, G. et al. High-precision timing of 42 millisecond pulsars with the European Pulsar Timing Array. MNRAS, 458(3):3341–3380, May 2016. doi: 10.1093/mnras/stw483.

[5] Manchester, R.N. et al. The Parkes Pulsar Timing Array Project. PASA, 30:e017, January 2013. doi: 10.1017/pasa.2012.017.

[6] Hobbs, G. et al. The International Pulsar Timing Array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27(8):084013, Apr 2010. doi: 10.1088/0264-9381/27/8/084013.

[7] Arzoumanian, Z. et al. The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. ApJ, 905(2):L34, December 2020. doi: 10.3847/2041-8213/abd401.

[8] Chen, S. et al. Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: Inferences in the stochastic gravitational-wave background search. *arXiv:2110.13184*, art. arXiv:2110.13184, October 2021.

[9] Goncharov, B. et al. On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array. ApJ, 917(2):L19, August 2021. doi: 10.3847/2041-8213/ac17f4.

[10] Antoniadis, J. et al. The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background. MNRAS, 510(4):4873–4887, March 2022. doi: 10.1093/mnras/stab3418.

[11] Hellings, R.W. and Downs, G.S. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. ApJ, 265:L39–L42, February 1983. doi: 10.1086/183954.

[12] Pol, N.S. et al. Astrophysics Milestones for Pulsar Timing Array Gravitational-wave Detection. ApJ, 911(2):L34, April 2021. doi: 10.3847/2041-8213/abf2c9.

[13] Kelley, L.Z. et al. Single sources in the low-frequency gravitational wave sky: properties and time to detection by pulsar timing arrays. MNRAS, 477(1):964–976, June 2018. doi: 10.1093/mnras/sty689.

[14] Burke-Spolaor, S. et al. The astrophysics of nanohertz gravitational waves. A&A Rev., 27(1):5, Jun 2019. doi: 10.1007/s00159-019-0115-7.

[15] Taylor, S.R. The Nanohertz Gravitational Wave Astronomer. *arXiv e-prints*, art. arXiv:2105.13270, May 2021.

[16] Arzoumanian, Z. et al. The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. ApJ, 905(2):L34, December 2020. doi: 10.3847/2041-8213/abd401.

[17] Rajagopal, M. and Romani, R.W. Ultra–Low-Frequency Gravitational Radiation from Massive Black Hole Binaries. ApJ, 446:543, June 1995. doi: 10.1086/175813.

[18] Sesana, A. et al. Low-Frequency Gravitational Radiation from Coalescing Massive Black Hole Binaries in Hierarchical Cosmologies. ApJ, 611(2):623–632, August 2004. doi: 10.1086/422185.

[19] Enoki, M. et al. Gravitational Waves from Supermassive Black Hole Coalescence in a Hierarchical Galaxy Formation Model. ApJ, 615(1):19–28, November 2004. doi: 10.1086/424475.

[20] Kelley, L.Z. et al. The gravitational wave background from massive black hole binaries in Illustris: spectral features and time to detection with pulsar timing arrays. MNRAS, 471(4):4508–4526, November 2017. doi: 10.1093/mnras/stx1638.

[21] Arzoumanian, Z. et al. Searching for Gravitational Waves from Cosmological Phase Transitions with the NANOGrav 12.5-Year Dataset. Phys. Rev. Lett., 127(25):251302, December 2021. doi: 10.1103/PhysRevLett.127.251302.

[22] van Haasteren, R. et al. Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data. MNRAS, 414(4):3117–3128, July 2011. doi: 10.1111/j.1365-2966.2011.18613.x.

[23] Blasi, S., Brdar, V. and Schmitz, K. Has NANOGrav Found First Evidence for Cosmic Strings? Phys. Rev. Lett., 126(4):041305, January 2021. doi: 10.1103/PhysRevLett.126.041305.

[24] De Luca, V., Franciolini, G. and Riotto, A. NANOGrav Data Hints at Primordial Black Holes as Dark Matter. Phys. Rev. Lett., 126(4):041303, January 2021. doi: 10.1103/PhysRevLett.126.041303.

[25] Vagnozzi, S. Implications of the NANOGrav results for inflation. MNRAS, 502(1):L11–L15, March 2021. doi: 10.1093/mnrasl/slaa203.

[26] Cornish, N.J. and Romano, J.D. When is a gravitational-wave signal stochastic? *Phys. Rev. D*, 92:042001, Aug 2015. doi: 10.1103/PhysRevD.92.042001. URL https://link.aps.org/doi/10.1103/PhysRevD.92.042001.

[27] Cornish, N.J. and Sampson, L. Towards robust gravitational wave detection with pulsar timing arrays. Phys. Rev. D, 93(10):104047, May 2016. doi: 10.1103/PhysRevD.93.104047.

[28] Kelley, L.Z. et al. Single sources in the low-frequency gravitational wave sky: properties and time to detection by pulsar timing arrays. MNRAS, 477(1):964–976, June 2018. doi: 10.1093/mnras/sty689.

[29] Vogelsberger, M. et al. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. MNRAS, 444(2):1518–1547, October 2014. doi: 10.1093/mnras/stu1536.

[30] Kelley, L.Z., Blecha, L. and Hernquist, L. Massive black hole binary mergers in dynamical galactic environments. MNRAS, 464(3):3131–3157, January 2017. doi: 10.1093/mnras/stw2452.

[31] Aggarwal, K. et al. The NANOGrav 11 yr Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. ApJ, 880(2):116, August 2019. doi: 10.3847/1538-4357/ab2236.

[32] Bécsy, B. and Cornish, N.J. Joint search for isolated sources and an unresolved confusion background in pulsar timing array data. *Classical and Quantum Gravity*, 37(13):135011, July 2020. doi: 10.1088/1361-6382/ab8bbd.

[33] Phinney, E.S. A Practical Theorem on Gravitational Wave Backgrounds. *arXiv e-prints*, art. astro-ph/0108028, August 2001.

[34] Sesana, A., Vecchio, A. and Colacino, C.N. The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. MNRAS, 390(1):192–209, October 2008. doi: 10.1111/j.1365-2966.2008.13682.x.

[35] Mingarelli, C.M.F. et al. Characterizing gravitational wave stochastic background anisotropy with pulsar timing arrays. *Phys. Rev. D*, 88:062005, Sep 2013. doi: 10.1103/PhysRevD.88.062005. URL https://link.aps.org/doi/10.1103/PhysRevD.88.062005.

[36] Cornish, N.J. and van Haasteren, R. Mapping the nano-Hertz gravitational wave sky. *arXiv e-prints*, art. arXiv:1406.4511, June 2014.

[37] Taylor, S.R. and Gair, J.R. Searching for anisotropic gravitational-wave backgrounds using pulsar timing arrays. *Phys. Rev. D*, 88:084001, Oct 2013. doi: 10.1103/

PhysRevD.88.084001. URL `https://link.aps.org/doi/10.1103/PhysRevD.88.084001`.

[38] Hotinli, S.C., Kamionkowski, M. and Jaffe, A.H. The search for statistical anisotropy in the gravitational-wave background with pulsar timing arrays. *The Open Journal of Astrophysics*, 2(1):8, September 2019. doi: 10.21105/astro.1904.05348.

[39] Taylor, S.R., van Haasteren, R. and Sesana, A. From bright binaries to bumpy backgrounds: Mapping realistic gravitational wave skies with pulsar-timing arrays. *Phys. Rev. D*, 102:084039, Oct 2020. doi: 10.1103/PhysRevD.102.084039. URL `https://link.aps.org/doi/10.1103/PhysRevD.102.084039`.

[40] Ali-Haïmoud, Y., Smith, T.L. and Mingarelli, C.M.F. Fisher formalism for anisotropic gravitational-wave background searches with pulsar timing arrays. *Phys. Rev. D*, 102: 122005, Dec 2020. doi: 10.1103/PhysRevD.102.122005. URL `https://link.aps.org/doi/10.1103/PhysRevD.102.122005`.

[41] Pol, N., Taylor, S.R. and Romano, J.D. Forecasting pulsar timing array sensitivity to anisotropy in the stochastic gravitational wave background. *arXiv e-prints*, art. arXiv:2206.09936, June 2022.

[42] Górski, K.M. et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, 622:759–771, April 2005. doi: 10.1086/427976.

[43] Mingarelli, C.M.F. et al. The local nanohertz gravitational-wave landscape from supermassive black hole binaries. *Nature Astronomy*, 1:886–892, November 2017. doi: 10.1038/s41550-017-0299-6.

[44] Hajian, A. and Souradeep, T. Measuring the statistical isotropy of the cosmic microwave background anisotropy. *The Astrophysical Journal*, 597(1):L5–L8, oct 2003. doi: 10.1086/379757. URL `https://doi.org/10.1086/379757`.

[45] Hajian, A. and Souradeep, T. The Cosmic Microwave Background Bipolar Power Spectrum: Basic Formalism and Applications. *arXiv e-prints*, art. astro-ph/0501001, December 2004.

[46] Hajian, A., Souradeep, T. and Cornish, N. Statistical Isotropy of the Wilkinson Microwave Anisotropy Probe Data: A Bipolar Power Spectrum Analysis. ApJ, 618(2): L63–L66, January 2005. doi: 10.1086/427652.

[47] Cornish, N.J. and Sesana, A. Pulsar timing array analysis for black hole backgrounds. *Classical and Quantum Gravity*, 30(22):224005, November 2013. doi: 10.1088/0264-9381/30/22/224005.

[48] Allen, B. Variance of the Hellings-Downs Correlation. *arXiv e-prints*, art. arXiv:2205.05637, May 2022.

[49] Di Matteo, T., King, A. and Cornish, N.J. *Black hole formation and growth*. Springer, 2019.

[50] Ellis, J.A. et al. Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite. Zenodo, September 2020. URL `https://doi.org/10.5281/zenodo.4059815`.

[51] Alam, M.F. et al. The NANOGrav 12.5 yr Data Set: Observations and Narrowband Timing of 47 Millisecond Pulsars. ApJS, 252(1):4, January 2021. doi: 10.3847/1538-4365/abc6a0.

[52] Bécsy, B., Cornish, N.J. and Digman, M.C. Fast Bayesian analysis of individual binaries in pulsar timing array data. *arXiv e-prints*, art. arXiv:2204.07160, April 2022.

[53] Zonca, A. et al. healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python. *Journal of Open Source Software*, 4(35):1298, March 2019. doi: 10.21105/joss.01298. URL `https://doi.org/10.21105/joss.01298`.

[54] Foreman-Mackey, D. corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, 1(2):24, jun 2016. doi: 10.21105/joss.00024. URL `https://doi.org/10.21105/joss.00024`.

[55] Lam, S.K., Pitrou, A. and Seibert, S. Numba: A llvm-based python jit compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450340052. doi: 10.1145/2833157.2833162. URL `https://doi.org/10.1145/2833157.2833162`.

[56] Lam, S.K. et al. numba/numba: Version 0.55.0, January 2022. URL `https://doi.org/10.5281/zenodo.5847553`.

[57] Romano, J.D. et al. Common-spectrum process versus cross-correlation for gravitational-wave searches using pulsar timing arrays. Phys. Rev. D, 103(6):063027, March 2021. doi: 10.1103/PhysRevD.103.063027.

[58] Rosado, P.A., Sesana, A. and Gair, J. Expected properties of the first gravitational wave signal detected with pulsar timing arrays. MNRAS, 451(3):2417–2433, Aug 2015. doi: 10.1093/mnras/stv1098.

[59] Kelley, L.Z. et al. Single sources in the low-frequency gravitational wave sky: properties and time to detection by pulsar timing arrays. MNRAS, 477(1):964–976, June 2018. doi: 10.1093/mnras/sty689.

[60] Arzoumanian, Z. et al. Searching for Gravitational Waves from Cosmological Phase Transitions with the NANOGrav 12.5-Year Dataset. Phys. Rev. Lett., 127(25):251302, December 2021. doi: 10.1103/PhysRevLett.127.251302.

[61] Charisi, M. et al. Multimessenger time-domain signatures of supermassive black hole binaries. MNRAS, 510(4):5929–5944, March 2022. doi: 10.1093/mnras/stab3713.

[62] Aggarwal, K. et al. The NANOGrav 11 yr Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. ApJ, 880(2):116, August 2019. doi: 10.3847/1538-4357/ab2236.

[63] Babak, S. et al. European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive black hole binaries. MNRAS, 455(2):1665–1679, January 2016. doi: 10.1093/mnras/stv2092.

[64] Zhu, X.J. et al. An all-sky search for continuous gravitational waves in the Parkes Pulsar Timing Array data set. MNRAS, 444(4):3709–3720, November 2014. doi: 10.1093/mnras/stu1717.

[65] Corbin, V. and Cornish, N.J. Pulsar Timing Array Observations of Massive Black Hole Binaries. *arXiv e-prints*, art. arXiv:1008.1782, August 2010.

[66] Lee, K.J. et al. Gravitational wave astronomy of single sources with a pulsar timing array. MNRAS, 414(4):3251–3264, July 2011. doi: 10.1111/j.1365-2966.2011.18622.x.

[67] Ellis, J.A. A Bayesian analysis pipeline for continuous GW sources in the PTA band. *Classical and Quantum Gravity*, 30(22):224004, November 2013. doi: 10.1088/0264-9381/30/22/224004.

[68] Taylor, S., Ellis, J. and Gair, J. Accelerated Bayesian model-selection and parameter-estimation in continuous gravitational-wave searches with pulsar-timing arrays. Phys. Rev. D, 90(10):104028, November 2014. doi: 10.1103/PhysRevD.90.104028.

[69] Abbott, B.P. et al. A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals. *Classical and Quantum Gravity*, 37(5):055002, March 2020. doi: 10.1088/1361-6382/ab685e.

[70] Taylor, S.R. et al. Detecting Eccentric Supermassive Black Hole Binaries with Pulsar Timing Arrays: Resolvable Source Strategies. ApJ, 817(1):70, January 2016. doi: 10.3847/0004-637X/817/1/70.

[71] Susobhanan, A. et al. Pulsar timing array signals induced by black hole binaries in relativistic eccentric orbits. Phys. Rev. D, 101(4):043022, February 2020. doi: 10.1103/PhysRevD.101.043022.

[72] Ellis, J.A., Siemens, X. and Creighton, J.D.E. Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays. ApJ, 756(2):175, September 2012. doi: 10.1088/0004-637X/756/2/175.

[73] Arzoumanian, Z. et al. The NANOGrav 11-year Data Set: High-precision Timing of 45 Millisecond Pulsars. ApJS, 235(2):37, April 2018. doi: 10.3847/1538-4365/aab5b0.

[74] Jaranowski, P., Królak, A. and Schutz, B.F. Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection. Phys. Rev. D, 58(6):063001, September 1998. doi: 10.1103/PhysRevD.58.063001.

[75] Cutler, C. and Schutz, B.F. Generalized F-statistic: Multiple detectors and multiple gravitational wave pulsars. Phys. Rev. D, 72(6):063006, September 2005. doi: 10.1103/PhysRevD.72.063006.

[76] Cornish, N.J. and Porter, E.K. The search for massive black hole binaries with LISA. *Classical and Quantum Gravity*, 24(23):5729–5755, December 2007. doi: 10.1088/0264-9381/24/23/001.

[77] Babak, S. and Sesana, A. Resolving multiple supermassive black hole binaries with pulsar timing arrays. Phys. Rev. D, 85(4):044034, February 2012. doi: 10.1103/PhysRevD.85.044034.

[78] Bai, Y. An adaptive directional metropolis-within-gibbs algorithm. *Preprint*, 2009.

[79] Liu, J.S., Liang, F. and Wong, W.H. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, 2000. ISSN 01621459. URL `http://www.jstor.org/stable/2669532`.

[80] Martino, L. A Review of Multiple Try MCMC algorithms for Signal Processing. *arXiv e-prints*, art. arXiv:1801.09065, January 2018.

[81] Craiu, R. and Lemieux, C. Acceleration of the multiple-try metropolis algorithm using antithetic and stratified sampling. *Statistics and Computing*, 17:109–120, 06 2007. doi: 10.1007/s11222-006-9009-4.

[82] Casarin, R., Craiu, R.V. and Leisen, F. Interacting Multiple Try Algorithms with Different Proposal Distributions. *arXiv e-prints*, art. arXiv:1011.1170, November 2010.

[83] Goldstein, J.M. et al. Associating host galaxy candidates to massive black hole binaries resolved by pulsar timing arrays. MNRAS, 485(1):248–259, May 2019. doi: 10.1093/mnras/stz420.

[84] Arzoumanian, Z. et al. Multimessenger Gravitational-wave Searches with Pulsar Timing Arrays: Application to 3C 66B Using the NANOGrav 11-year Data Set. ApJ, 900(2):102, September 2020. doi: 10.3847/1538-4357/ababa1.

[85] Liu, T. and Vigeland, S.J. Multi-messenger Approaches to Supermassive Black Hole Binary Detection and Parameter Estimation: Implications for Nanohertz Gravitational Wave Searches with Pulsar Timing Arrays. ApJ, 921(2):178, November 2021. doi: 10.3847/1538-4357/ac1da9.

[86] Cornish, N.J. and Sampson, L. Towards robust gravitational wave detection with pulsar timing arrays. Phys. Rev. D, 93(10):104047, May 2016. doi: 10.1103/PhysRevD.93.104047.

[87] Taylor, S.R. et al. All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar timing arrays. Phys. Rev. D, 95(4):042002, February 2017. doi: 10.1103/PhysRevD.95.042002.

[88] Matthews, C. et al. Umbrella sampling: a powerful method to sample tails of distributions. MNRAS, 480(3):4069–4079, November 2018. doi: 10.1093/mnras/sty2140.

[89] Carlin, B.P. and Chib, S. Bayesian model choice via markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(3):473–484, 1995. doi: https://doi.org/10.1111/j.2517-6161.1995.tb02042.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02042.x`.

[90] O'Beirne, L. et al. Constraining alternative polarization states of gravitational waves from individual black hole binaries using pulsar timing arrays. Phys. Rev. D, 99(12):124039, June 2019. doi: 10.1103/PhysRevD.99.124039.

[91] Bécsy, B. and Cornish, N.J. Bayesian search for gravitational wave bursts in pulsar timing array data. *Classical and Quantum Gravity*, 38(9):095012, May 2021. doi: 10.1088/1361-6382/abf1c6.

[92] Arzoumanian, Z. et al. The NANOGrav 11-year Data Set: High-precision Timing of 45 Millisecond Pulsars. ApJS, 235(2):37, Apr 2018. doi: 10.3847/1538-4365/aab5b0.

[93] Verbiest, J.P.W. et al. The International Pulsar Timing Array: First data release. MNRAS, 458(2):1267–1288, May 2016. doi: 10.1093/mnras/stw347.

[94] Arzoumanian, Z. et al. The NANOGrav 11 Year Data Set: Pulsar-timing Constraints on the Stochastic Gravitational-wave Background. ApJ, 859(1):47, May 2018. doi: 10.3847/1538-4357/aabd3b.

[95] Lentati, L. et al. European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. MNRAS, 453(3):2576–2598, Nov 2015. doi: 10.1093/mnras/stv1538.

[96] Sesana, A., Vecchio, A. and Colacino, C.N. The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. *Mon. Not. Roy. Astron. Soc.*, 390:192, 2008. doi: 10.1111/j.1365-2966.2008.13682.x.

[97] Kocsis, B. and Sesana, A. Gas-driven massive black hole binaries: signatures in the nHz gravitational wave background. MNRAS, 411(3):1467–1479, Mar 2011. doi: 10.1111/j.1365-2966.2010.17782.x.

[98] Ravi, V. et al. Does a 'stochastic' background of gravitational waves exist in the pulsar timing band? *Astrophys. J.*, 761:84, 2012. doi: 10.1088/0004-637X/761/2/84.

[99] Mingarelli, C.M.F. et al. The local nanohertz gravitational-wave landscape from supermassive black hole binaries. *Nature Astronomy*, 1:886–892, Nov 2017. doi: 10.1038/s41550-017-0299-6.

[100] Cornish, N.J. and Sesana, A. Pulsar Timing Array Analysis for Black Hole Backgrounds. *Class. Quant. Grav.*, 30:224005, 2013. doi: 10.1088/0264-9381/30/22/224005.

[101] Mingarelli, C.M.F. et al. Characterizing gravitational wave stochastic background anisotropy with pulsar timing arrays. Phys. Rev. D, 88(6):062005, Sep 2013. doi: 10.1103/PhysRevD.88.062005.

[102] Taylor, S.R. et al. Limits on Anisotropy in the Nanohertz Stochastic Gravitational Wave Background. Phys. Rev. Lett., 115(4):041101, Jul 2015. doi: 10.1103/PhysRevLett.115.041101.

[103] Aggarwal, K. et al. The NANOGrav 11 yr Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. ApJ, 880(2):116, Aug 2019. doi: 10.3847/1538-4357/ab2236.

[104] Babak, S. et al. European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive black hole binaries. MNRAS, 455(2):1665–1679, Jan 2016. doi: 10.1093/mnras/stv2092.

[105] Amaro-Seoane, P. et al. Laser Interferometer Space Antenna. *arXiv e-prints*, art. arXiv:1702.00786, Feb 2017.

[106] Timpano, S.E., Rubbo, L.J. and Cornish, N.J. Characterizing the galactic gravitational wave background with LISA. *Phys. Rev.*, D73:122001, 2006. doi: 10.1103/PhysRevD.73.122001.

[107] LIGO Scientific Collaboration et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, Apr 2015. doi: 10.1088/0264-9381/32/7/074001.

[108] Cornish, N.J. and Romano, J.D. When is a gravitational-wave signal stochastic? Phys. Rev. D, 92(4):042001, Aug 2015. doi: 10.1103/PhysRevD.92.042001.

[109] Babak, S. and Sesana, A. Resolving multiple supermassive black hole binaries with pulsar timing arrays. Phys. Rev. D, 85(4):044034, Feb 2012. doi: 10.1103/PhysRevD.85.044034.

[110] Petiteau, A. et al. Resolving multiple supermassive black hole binaries with pulsar timing arrays. II. Genetic algorithm implementation. Phys. Rev. D, 87(6):064036, Mar 2013. doi: 10.1103/PhysRevD.87.064036.

[111] Green, P.J. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[112] Cornish, N.J. and Littenberg, T.B. Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, 32(13):135012, 2015.

[113] Umstätter, R. et al. Bayesian modeling of source confusion in LISA data. Phys. Rev. D, 72(2):022001, July 2005. doi: 10.1103/PhysRevD.72.022001.

[114] Crowder, J. and Cornish, N. A Solution to the Galactic Foreground Problem for LISA. *Phys. Rev.*, D75:043008, 2007. doi: 10.1103/PhysRevD.75.043008.

[115] Littenberg, T.B. A detection pipeline for galactic binaries in LISA data. *Phys. Rev.*, D84:063009, 2011. doi: 10.1103/PhysRevD.84.063009.

[116] Littenberg, T. et al. Global Analysis of the Gravitational Wave Signal from Galactic Binaries. *arXiv e-prints*, art. arXiv:2004.08464, April 2020.

[117] van Haasteren, R. et al. On measuring the gravitational-wave background using Pulsar Timing Arrays. *Mon. Not. Roy. Astron. Soc.*, 395:1005, 2009. doi: 10.1111/j.1365-2966.2009.14590.x.

[118] Arzoumanian, Z. et al. The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background. ApJ, 821(1):13, Apr 2016. doi: 10.3847/0004-637X/821/1/13.

[119] Hellings, R.W. and Downs, G.S. Upper limits on the isotopic gravitational radiation background frompulsar timing analysis. ApJ, 265:L39–L42, February 1983. doi: 10.1086/183954.

[120] Phinney, E.S. A Practical Theorem on Gravitational Wave Backgrounds. *arXiv e-prints*, art. astro-ph/0108028, Aug 2001.

[121] Sesana, A., Vecchio, A. and Colacino, C.N. The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. MNRAS, 390(1):192–209, October 2008. doi: 10.1111/j.1365-2966.2008.13682.x.

[122] Hastie, D.I. and Green, P.J. Model choice using reversible jump markov chain monte carlo. *Statistica Neerlandica*, 66(3):309–338, 2012. doi: 10.1111/j.1467-9574.2012.00516.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9574.2012.00516.x`.

[123] MacKay, D.J. *Bayesian methods for adaptive models*. PhD thesis, California Institute of Technology, 1992.

[124] Sambridge, M. et al. Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, 167(2):528–542, 2006.

[125] Ellis, J.A., Siemens, X. and Creighton, J.D. Optimal strategies for continuous gravitational wave detection in pulsar timing arrays. *The Astrophysical Journal*, 756 (2):175, 2012.

[126] Gramacy, R., Samworth, R. and King, R. Importance tempering. *Statistics and Computing*, 20(1):1–7, 2010.

[127] Chatziioannou, K. et al. Constructing gravitational waves from generic spin-precessing compact binary inspirals. *Physical Review D*, 95(10):104004, 2017.

[128] Chatziioannou, K. et al. Detection and parameter estimation of gravitational waves from compact binary inspirals with analytical double-precessing templates. *Physical Review D*, 89(10):104023, 2014.

[129] Anholm, M. et al. Optimal strategies for gravitational wave stochastic background searches in pulsar timing data. Phys. Rev. D, 79(8):084030, Apr 2009. doi: 10.1103/PhysRevD.79.084030.

[130] Siemens, X. et al. The stochastic background: scaling laws and time to detection for pulsar timing arrays. *Classical and Quantum Gravity*, 30(22):224015, Nov 2013. doi: 10.1088/0264-9381/30/22/224015.

[131] Joshi, B.C. et al. Precision pulsar timing with the ORT and the GMRT and its applications in pulsar astrophysics. *Journal of Astrophysics and Astronomy*, 39(4):51, August 2018. doi: 10.1007/s12036-018-9549-y.

[132] Lee, K.J. Prospects of Gravitational Wave Detection Using Pulsar Timing Array for Chinese Future Telescopes. In Qain, L. and Li, D., editors, *Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015*, volume 502 of *Astronomical Society of the Pacific Conference Series*, page 19, February 2016.

[133] Bailes, M. et al. MeerTime - the MeerKAT Key Science Program on Pulsar Timing. In *MeerKAT Science: On the Pathway to the SKA*, page 11, January 2016.

[134] Perera, B.B.P. et al. The International Pulsar Timing Array: second data release. MNRAS, 490(4):4666–4687, December 2019. doi: 10.1093/mnras/stz2857.

[135] Aggarwal, K. et al. The NANOGrav 11 yr Data Set: Limits on Gravitational Wave Memory. ApJ, 889(1):38, January 2020. doi: 10.3847/1538-4357/ab6083.

[136] Littenberg, T.B. and Cornish, N.J. Bayesian inference for spectral estimation of gravitational wave detector noise. *Phys. Rev. D*, 91:084034, Apr 2015. doi: 10.1103/PhysRevD.91.084034. URL `https://link.aps.org/doi/10.1103/PhysRevD.91.084034`.

[137] Cornish, N.J. et al. BayesWave analysis pipeline in the era of gravitational wave observations. Phys. Rev. D, 103(4):044006, February 2021. doi: 10.1103/PhysRevD.103.044006.

[138] Klimenko, S. et al. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. Phys. Rev. D, 93(4):042004, February 2016. doi: 10.1103/PhysRevD.93.042004.

[139] Thrane, E. and Coughlin, M. Detecting Gravitation-Wave Transients at 5 σ : A Hierarchical Approach. Phys. Rev. Lett., 115(18):181102, October 2015. doi: 10.1103/PhysRevLett.115.181102.

[140] Lynch, R. et al. Information-theoretic approach to the gravitational-wave burst detection problem. Phys. Rev. D, 95(10):104046, May 2017. doi: 10.1103/PhysRevD.95.104046.

[141] Aasi, J. et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, April 2015. doi: 10.1088/0264-9381/32/7/074001.

[142] Acernese, F. et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, January 2015. doi: 10.1088/0264-9381/32/2/024001.

[143] The LIGO Scientific Collaboration and The Virgo Collaboration. All-sky search for short gravitational-wave bursts in the second Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 100:024017, Jul 2019. doi: 10.1103/PhysRevD.100.024017. URL `https://link.aps.org/doi/10.1103/PhysRevD.100.024017`.

[144] LIGO Scientific Collaboration and Virgo Collaboration. All-sky search for long-duration gravitational-wave transients in the second Advanced LIGO observing run. Phys. Rev. D, 99(10):104033, May 2019. doi: 10.1103/PhysRevD.99.104033.

[145] Robson, T. and Cornish, N.J. Detecting gravitational wave bursts with LISA in the presence of instrumental glitches. Phys. Rev. D, 99(2):024019, January 2019. doi: 10.1103/PhysRevD.99.024019.

[146] Finn, L.S. and Lommen, A.N. Detection, Localization, and Characterization of Gravitational Wave Bursts in a Pulsar Timing Array. ApJ, 718(2):1400–1415, August 2010. doi: 10.1088/0004-637X/718/2/1400.

[147] Deng, X. Searching for gravitational wave bursts via Bayesian nonparametric data analysis with pulsar timing arrays. Phys. Rev. D, 90(2):024020, July 2014. doi: 10.1103/PhysRevD.90.024020.

[148] Zhu, X.J. et al. Detection and localization of single-source gravitational waves with pulsar timing arrays. MNRAS, 449(2):1650–1663, May 2015. doi: 10.1093/mnras/stv381.

[149] Pitkin, M. Extending gravitational wave burst searches with pulsar timing arrays. MNRAS, 425(4):2688–2697, October 2012. doi: 10.1111/j.1365-2966.2012.21655.x.

[150] Ellis, J.A. and Cornish, N.J. Transdimensional Bayesian approach to pulsar timing noise analysis. Phys. Rev. D, 93(8):084048, April 2016. doi: 10.1103/PhysRevD.93.084048.

[151] Amaro-Seoane, P. et al. Triplets of supermassive black holes: astrophysics, gravitational waves and detection. MNRAS, 402(4):2308–2320, March 2010. doi: 10.1111/j.1365-2966.2009.16104.x.

[152] Damour, T. and Vilenkin, A. Gravitational wave bursts from cusps and kinks on cosmic strings. Phys. Rev. D, 64(6):064008, September 2001. doi: 10.1103/PhysRevD.64.064008.

[153] NANOGrav Collaboration et al. The NANOGrav Nine-year Data Set: Observations, Arrival Time Measurements, and Analysis of 37 Millisecond Pulsars. ApJ, 813(1):65, November 2015. doi: 10.1088/0004-637X/813/1/65.

[154] Hazboun, J.S. La forge. Zenodo, October 2020. URL `https://doi.org/10.5281/zenodo.4152550`.

[155] Gabor, D. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[156] Lentati, L., Alexander, P. and Hobson, M.P. Generative pulsar timing analysis. *Monthly Notices of the Royal Astronomical Society*, 447(3):2159–2168, Jan 2015. ISSN 1365-2966. doi: 10.1093/mnras/stu2611. URL `http://dx.doi.org/10.1093/mnras/stu2611`.

[157] Corbin, V. and Cornish, N.J. Pulsar Timing Array Observations of Massive Black Hole Binaries. *arXiv e-prints*, art. arXiv:1008.1782, August 2010.

[158] Bécsy, B. et al. Parameter Estimation for Gravitational-wave Bursts with the BayesWave Pipeline. ApJ, 839(1):15, April 2017. doi: 10.3847/1538-4357/aa63ef.

[159] Abadie, J. et al. All-sky search for gravitational-wave bursts in the second joint LIGO-Virgo run. Phys. Rev. D, 85(12):122007, June 2012. doi: 10.1103/PhysRevD.85.122007.

[160] Lentati, L. et al. Hyper-efficient model-independent Bayesian method for the analysis of pulsar timing data. *Phys. Rev. D*, 87(10):104021, 2013. doi: 10.1103/PhysRevD.87.104021.

[161] Taylor, S.R. et al. A Parallelized Bayesian Approach To Accelerated Gravitational-Wave Background Characterization. *arXiv e-prints*, art. arXiv:2202.08293, February 2022.

[162] Cornish, N.J. and Romano, J.D. Towards a unified treatment of gravitational-wave data analysis. *Phys. Rev. D*, 87(12):122003, 2013. doi: 10.1103/PhysRevD.87.122003.

[163] Siemens, X. et al. The likelihood evaluation in pta searches part 2: Factorizing model 3a. NANOGrav Fall Meeting, 2021.

APPENDICES

APPENDIX A

VARIANCE OF SPATIAL CORRELATIONS WITH DIFFERENT NORMALIZATIONS

In this appendix we explore how the variance of the HD correlations from a single binary depend on the way we normalize the correlation values. We reproduce some results of Ref. [48] and derive formulae for the case with a different normalization used in Section 2.5.3. We compare these results with simulations for a CW signal in a simulated PTA with an isotropic distribution of pulsars.

A.0.1 Without pulsar terms

From Eq. (A31) of Ref. [48] the mean HD correlation for a binary with a given inclination (ι) and GW amplitude ($\mathcal{A}$) is:

$$\mu = \frac{\mathcal{A}^2}{16}\left[1 + 6\cos^2\iota + \cos^4\iota\right]\mu_u(\gamma), \tag{A.1}$$

and neglecting pulsar terms the variance is:

$$\sigma^2 = \frac{\mathcal{A}^4}{256}\left[1 + 6\cos^2\iota + \cos^4\iota\right]^2\sigma_u^2(\gamma) + \frac{\mathcal{A}^4}{512}\left[\sin^8\iota\right]\sigma_c^2(\gamma). \tag{A.2}$$

We can average these over $\cos\iota \in [-1, 1]$ to get (see Eq. (A32) of Ref. [48]):

$$\mu_{\text{average}} = \frac{\mathcal{A}^2}{5}\mu_u(\gamma), \tag{A.3}$$

and:

$$\sigma^2_{\text{average}} = \frac{71\mathcal{A}^4}{1260}\sigma_u^2(\gamma) + \frac{\mathcal{A}^4}{1260}\sigma_c^2(\gamma). \tag{A.4}$$

We normalize these so that $\mu_{\text{average}} = \mu_u$, which implies $\mathcal{A}^2 = 5$ and:

$$\sigma_{\text{average}} = \sqrt{\frac{355}{252}\sigma_u^2(\gamma) + \frac{5}{252}\sigma_c^2(\gamma)}. \tag{A.5}$$

This is appropriate if we normalize the correlation values globally over all realizations. However, if we normalize the correlations realization by realization, then we need to normalize Eq. (A.1) before averaging. We thus require $\mathcal{A}^2/16\left[1 + 6\cos^2\iota + \cos^4\iota\right] = 1$, which implies:

$$\sigma^2 = \sigma_u^2(\gamma) + \frac{\sin^8\iota}{2\left[1 + 6\cos^2\iota + \cos^4\iota\right]^2}\sigma_c^2(\gamma). \tag{A.6}$$

If we average this over $\cos\iota \in [-1, 1]$, we get:

$$\sigma_{\text{average}} = \sqrt{\sigma_u^2(\gamma) + \frac{1}{4}\alpha\sigma_c^2(\gamma)}, \tag{A.7}$$

where $\alpha = 4 + \pi - 3\pi/\sqrt{2}$.

Figure A.1 shows the comparison of simulation results with Eqs. (A.5) and (A.7).

The simulation normalized per realization seems to match Eq. (A.7) really well. The one normalized globally generally follows Eq. (A.5), but deviates at small and large angles.

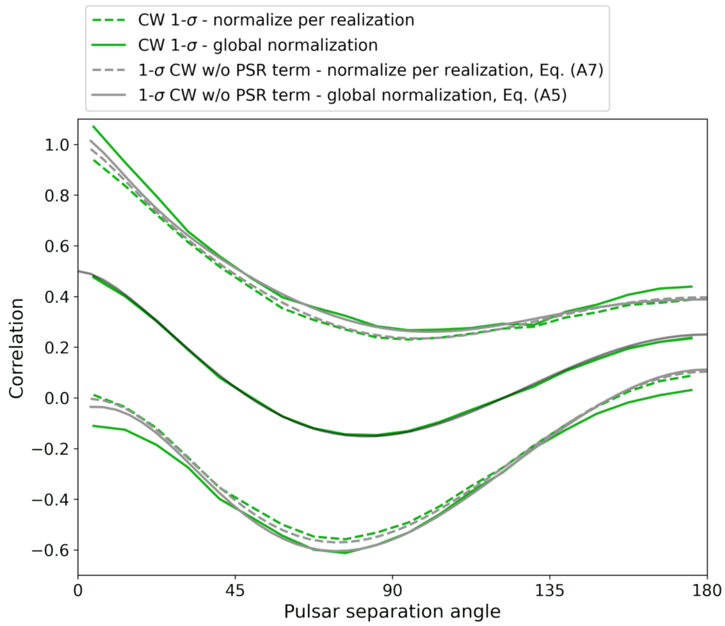


Figure A.1: Variance of HD correlations from a CW simulation without pulsar terms and a simulated PTA with isotropically distributed pulsars. We also show the theoretical variance with two different averaging.

A.0.2 With pulsar terms

If we include pulsar terms, the mean correlation is unchanged, but the variance becomes (cf. Eq. (B10) of Ref. [48]):

$$\sigma^2 = \frac{\mathcal{A}^4}{256}\left[1 + 6\cos^2\iota + \cos^4\iota\right]^2\sigma_u^2(\gamma) - \frac{3\mathcal{A}^4}{512}\left[\sin^8\iota\right]\sigma_p^2(\gamma) + \\ \mathcal{A}^4\left(\frac{3}{512}\left[1 + 6\cos^2\iota + \cos^4\iota\right]^2 + \frac{5}{1024}\left[\sin^8\iota\right]\right)\sigma_c^2(\gamma). \quad \text{(A.8)}$$

Averaging this over $\cos\iota \in [-1, 1]$ gives (see Eq. (B11) of Ref. [48]):

$$\sigma^2 = \frac{71\mathcal{A}^4}{1260}\sigma_u^2(\gamma) - \frac{3\mathcal{A}^4}{1260}\sigma_p^2(\gamma) + \frac{109\mathcal{A}^4}{1260}\sigma_c^2(\gamma). \quad \text{(A.9)}$$

Plugging in $\mathcal{A}^4 = 25$ gives:

$$\sigma_{\text{average}} = \sqrt{\frac{355}{252}\sigma_u^2(\gamma) - \frac{5}{84}\sigma_p^2(\gamma) + \frac{545}{252}\sigma_c^2(\gamma)}. \tag{A.10}$$

This is the average standard deviation if we normalize correlations globally, over many realizations. Alternatively, we can express $\mathcal{A}$ from Eq. (A.1) and plug into Eq. (A.8) before averaging. This gives:

$$\sigma^2 = \sigma_u^2(\gamma) - \frac{3\sin^8\iota}{2\left[1+6\cos^2\iota+\cos^4\iota\right]^2}\sigma_p^2(\gamma) + \left(\frac{3}{2} + \frac{5\sin^8\iota}{4\left[1+6\cos^2\iota+\cos^4\iota\right]^2}\right)\sigma_c^2(\gamma). \tag{A.11}$$

Averaging over $\cos\iota \in [-1,1]$ gives:

$$\sigma_{\text{average}} = \sqrt{\sigma_u^2(\gamma) - \frac{3}{4}\alpha\sigma_p^2(\gamma) + \left(\frac{3}{2} + \frac{5}{8}\alpha\right)\sigma_c^2(\gamma)}. \tag{A.12}$$

Figure A.2 shows simulation results (both with global normalization and per realization normalization) and Eqs. (A.10) and (A.12). We can see that both match well with the appropriate simulation.

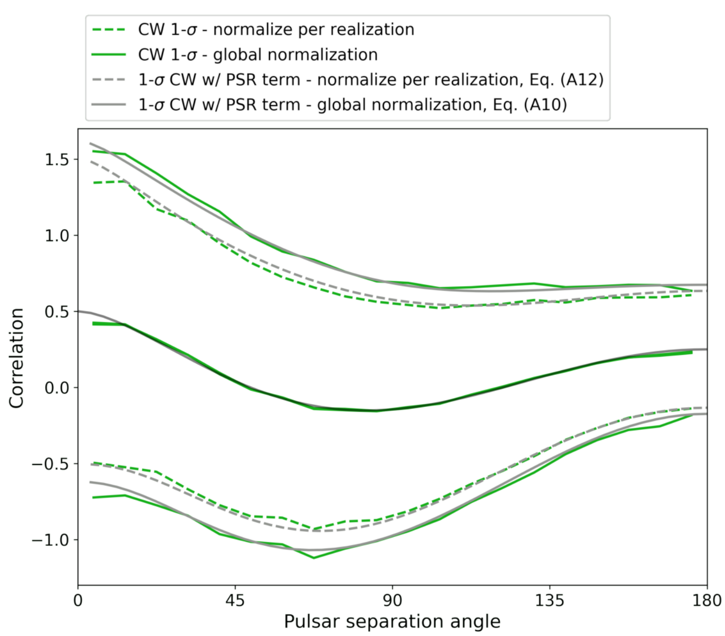


Figure A.2: Variance of HD correlations from a CW simulation with pulsar terms and a simulated PTA with isotropically distributed pulsars. We also show the theoretical variance with two different averaging.

APPENDIX B

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 $\mathbf{a}$ [160]. The log likelihood then becomes:

$$\log L = -\frac{1}{2}(\delta t - s - \mathbf{F}\cdot\mathbf{a}|\delta t - s - \mathbf{F}\cdot\mathbf{a})\,. \tag{B.1}$$

The usual approach is to marginalize over the $\mathbf{a}$'s analytically. If we assume the signal is not correlated between pulsars, eq. (B.1) can still be evaluated pulsar-by-pulsar [161]. However, the per-pulsar factorization breaks down if we intend to model a GWB with Hellings-Downs correlations between the pulsars [162]. 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. [163]. For simplicity let us ignore the individual sources for now and focus on the GWB. Expanding the log likelihood we have:

$$\log L = -\frac{1}{2}\left[(\delta t|\delta t) - 2\mathbf{a}\cdot\mathbf{P} + \mathbf{a}\cdot\mathbf{Q}\cdot\mathbf{a}\right]\,, \tag{B.2}$$

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 hyper-prior for the $\mathbf{a}$ is such that each a_k is drawn from a Gaussian distribution with variance $S(f)$ and correlations given by:

$$\begin{aligned} \mathrm{E}[a_{(2i-1)}a_{(2j-1)}] &= H_{ij}S(f) \\ \mathrm{E}[a_{(2i-1)}a_{(2j)}] &= 0 \\ \mathrm{E}[a_{(2i)}a_{(2j)}] &= H_{ij}S(f)\,, \end{aligned} \tag{B.3}$$

where H_{ij} is the Hellings-Downs correlation between pulsars i, j. More schematically we can write $\mathrm{E}[a_k a_l] = \varphi_{kl}$ where the entries of φ_{kl} are given by (B.3). The posterior probability distribution for $\mathbf{a}$ can then be written as:

$$p(\mathbf{a}|\delta t, S) = \frac{e^{-\frac{1}{2}(\delta t|\delta t)}}{\sqrt{\det(2\pi C)\det(2\pi\varphi)}} \times e^{a_k P^k - \frac{1}{2}a_k a_l Q^{kl} - \frac{1}{2}a_k a_l (\varphi^{-1})^{kl}}\,. \tag{B.4}$$

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 multi-variate 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 C

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 [50], 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 C.1 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.

Table C.1: Runtimes of old likelihood and different components of the new likelihood for various representative datasets made by multiplicating a dataset made to resemble the NANOGrav 12.5-year dataset (NG12.5)

	NG12.5	2 × NG12.5	5 × NG12.5
# of pulsars	45	90	225
# 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[1]	0.016 ms	0.020 ms	0.029 ms
GW shape parameter update[2]	90 ms	220 ms	770 ms
Pulsar red noise update	180 ms	400 ms	1300 ms

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 C.1 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 ∼20,000 times faster than the old likelihood. We also timed the recalculation of the

[1]With fixed shape parameters

[2]Common parameters or pulsar distances

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 C.1, 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 sub-optimal.

APPENDIX D

WAVEFORM RECONSTRUCTION IN ALL PULSARS

In Section 5.5, we only show the waveforms reconstructed in a single pulsar (see Figures 5.2, 5.3, 5.4, 5.6, and 5.7), but it is also interesting to look at how the signal appears in different pulsars. Figure D.1 shows the simulated data and the reconstructed signal in all 20 pulsars of the array for the injection with two sine-Gaussian wavelets (cf. Figure 5.2). While the amplitude and waveform of the signal appearing in datastreams of different pulsars are significantly different, `BayesHopperBurst` consistently reconstructs the signal in each pulsar. This is not surprising given that we employ a coherent signal model which uses information from all the pulsars to constrain the waveform of the signal.

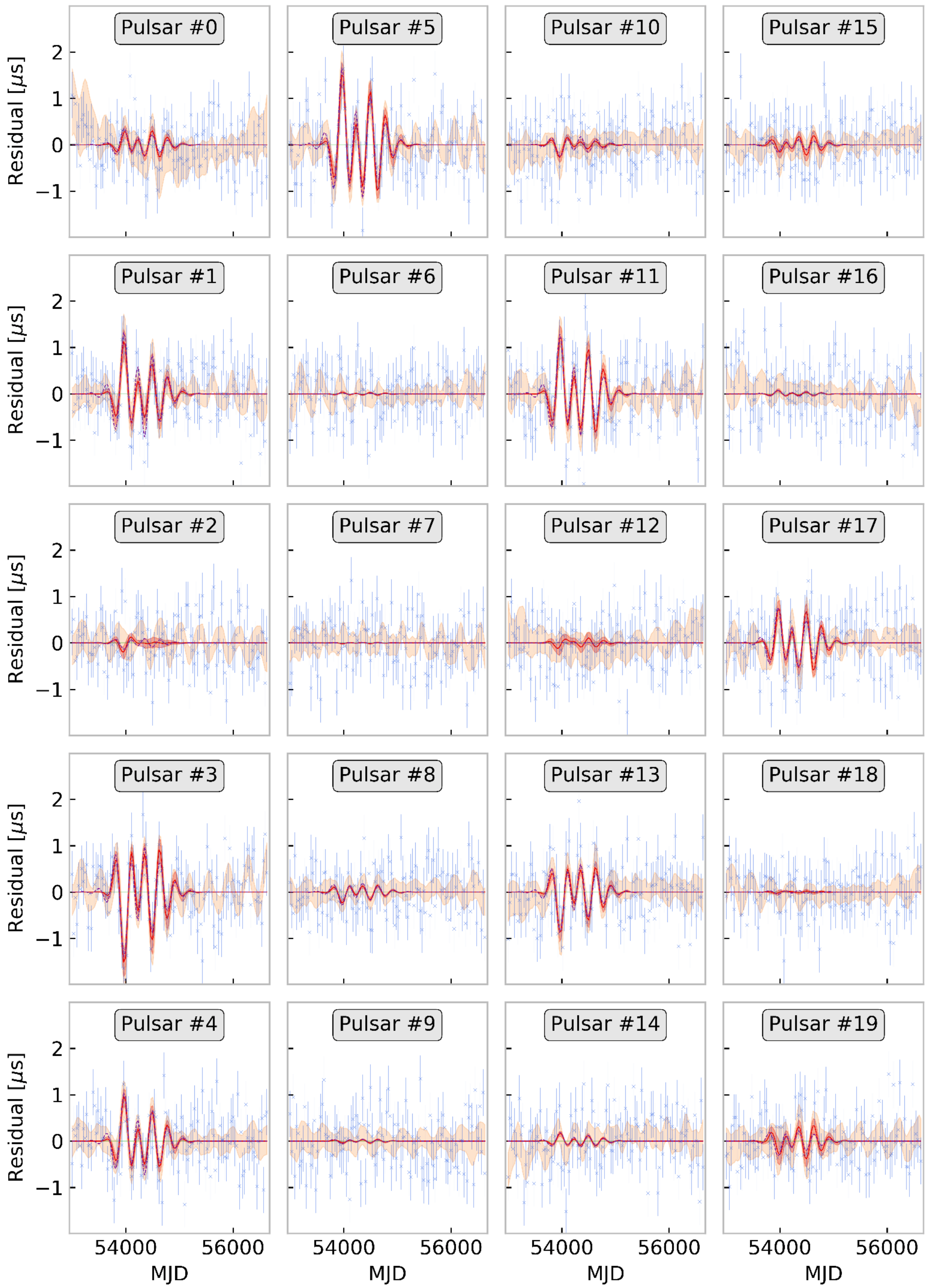


Figure D.1: Reconstruction of a signal consisting of two sine-Gaussian wavelets in all 20 pulsars in the array. This result is from the same dataset for which the reconstruction in pulsar #3 is shown in Figure 5.2. For more details see the caption of Figure 5.2.