Searching for Gravitational Waves Using Pulsar Timing Arrays

by

Justin A. Ellis

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the degree of

Doctor of Philosophy
in Physics

at
The University of Wisconsin–Milwaukee
August 2014

ABSTRACT

SEARCHING FOR GRAVITATIONAL WAVES USING PULSAR TIMING ARRAYS

The University of Wisconsin–Milwaukee, August 2014

Under the Supervision of Professor Xavier Siemens

Gravitational Waves (GWs) are tiny ripples in the fabric of spacetime predicted by Einstein's theory of General Relativity. Pulsar timing arrays (PTAs) offer a unique opportunity to detect low frequency GWs in the near future. Such a detection would be complementary to both LISA and LIGO GW efforts. In this frequency band, the expected source of GWs are Supermassive Black Hole Binaries (SMBHBs) that will most likely form an ensemble creating a stochastic GW background with possibly a few nearby/massive sources that will be individually resolvable. A direct detection of GWs will open a new window into the fields of astronomy and astrophysics by allowing us to constrain the coalescence rate of SMBHBs, providing us with further tests on the theory of General Relativity, and giving us access to properties of black holes not accessible by current astronomical techniques.

This dissertation work focuses primarily on the development of several robust data analysis pipelines for the detection and characterization of continuous GWs and a stochastic GW background. The data analysis problem for PTAs is quite difficult as one must fully take into account the timing model that must be fit in order to obtain the residuals, uneven sampling (including large gaps), and potential red noise processes. The data analysis techniques presented here handle all of these effects completely while allowing additional freedom in parameterizing the noise present in the data. The accumulation of work from this dissertation has resulted in a fully functional, robust, and efficient data analysis pipeline that has been successfully applied to the 5- and 9-year NANOGrav data releases.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost I would like to thank Dr. Jason Best. I likely would have never joined the physics community if it were not for his charisma and supervision during the early days. After one semester in Honors general science at Shepherd University, I knew that I wanted to do astrophysics. Next of course I would like to thank my Graduate Advisor Dr. Xavier Siemens whose mentorship has gone above and beyond. Xavi would know when I was getting burned out or something was wrong and would actually tell me to stop working. He has called me during the weekend to tell me to take a break and come have a beer, pulled me aside at meetings to make sure everything was ok. This relationship led to a very natural communication between us and I think it led to us both being more productive during my dissertation work. Next I would like to thank Dr. Jolien Creighton whose data analysis expertise was always welcome after Xavi and I had been staring at his board perplexed for hours.

There are several post-docs along the way who have been extremely helpful, especially Chris Pankow. Any time I had any type of programming or computational question, Chris was my go to guy. I would also like to thank Richard O'Shaughnessy whose frantic conversations were very helpful when putting together my MCMC codes. During the early days at WVU, Joris Verbiest was extremely helpful in introducing me to the world of pulsar timing, the C programming language, and shell scripting. I still think everything I know about Tempo2 came from Joris. I'd also like to thank Evan Oschner and Brennan Hughey for their help and advice along the way. I'd also like to thank Rutger van Haasteren who has been my sounding board for a few years now and was my contact in the EPTA during his time there.

Next come the graduate students. As I transferred into UWM during my third year, I didn't meet many of the other graduate students but Maddie and Les Wade as well as Sydney Chamberlin made me feel at home right away. I'd really like to thank my office mate, Alex Urban, for putting up with all of my rants everyday and just being a pretty cool guy in general. Even though we didn't meet on the best of terms (apparently we both hated each other) his friendship has meant a lot to me of the past few years.

Lastly and most importantly, I'd like to thank my family for supporting me through

it all. I'd like to thank my parents who instilled in me at a young age a competitive desire to succeed and to make them proud. I thank my mother for listening to me ramble every day on my walk home from the office and my father for having an unwavering faith in me. I know he was there to calm her down when she thought I was in over my head. I'd like to thank my wife, Ilene, who has also been *incredibly* patient with me over these past few years as I spent late nights and many weekends working. I thank her for always being there to calm me down when I get panicked and reassuring me when I think I'm in over my head. Finally I'd like to thank my dog River and cat Erwin who have been my main companions during these last weeks of writing late into the night and throughout the day when Ilene is off at work.

Chapter 1

Introduction

> "*Through fire and water. From the lowest dungeon to the highest peak, I fought him, the Balrog of Morgoth. Until at last, I threw down my enemy and smote his ruin upon the mountainside. Darkness took me. And I strayed out of thought and time. Stars wheeled overhead and every day was as long as a life-age of the earth. But it was not the end. I felt life in me again. I've been sent back until my task is done.*"
>
> — Gandalf, *The Lord of the Rings: The Two Towers*

Einstein's theory of general relativity is a theory of space, time and gravitation that was completed in 1915. It constitutes the foundation of our understanding of large-scale phenomena, from planetary motion to the evolution of the Universe itself. In general relativity, the gravitational force is no longer a "spooky action at a distance" but is instead explained as a manifestation of spacetime curvature. More precisely, the intrinsic, observer-independent properties of spacetime are described by a spacetime metric whose deviation from flatness (i.e., curvature) accounts for the physical effects of a gravitational field. Furthermore, this curvature of spacetime is related to the energy and momentum of the matter in spacetime and this relationship is embodied in Einstein's equation. Although the differences between general relativity and Newtonian physics are negligible in laboratory physics, they are crucial for many areas of astrophysics and cosmology. Furthermore, general relativity has made several important predictions that have been observed and confirmed such as gravitational time dilation, gravitational lensing, and the existence of black holes; while other predictions, such as gravitational waves, have not yet been confirmed conclusively.

One of the most important predictions of general relativity is gravitational waves (GWs) which are tiny ripples in the fabric of spacetime that have thus far eluded direct detection. Using precise timing observations of the Hulse-Taylor double neutron star (DNS) system, astronomers were able to prove (indirectly) that gravitational radiation exists by measuring the orbital rate of decay due to gravitational wave emission, which was found to be in excellent agreement with general relativity (Hulse & Taylor 1975; Taylor & Weisberg 1982). This work was awarded the 1993 Nobel prize in physics. Recently the BICEP2 team has announced the discovery of a significant detection of GWs in the B-mode power spectrum of Cosmic Microwave Background (CMB) (BICEP2 Collaboration et al. 2014) lending further credence to the definite existence of GWs on very different frequency scales (see section 1.1.1 for more details on the GW spectrum). Although both of these experiments have given strong evidence[1] for the existence of GWs and are in very good agreement with the predictions of general relativity, they are both, to varying degrees, *indirect* detections. The elusive *direct* detection has still yet to be made.

Over the last few decades, scientists have constructed new experiments to detect and characterize GWs from astrophysical sources. These experiments fall into two broad categories; interferometric detectors such as the Laser Interferometer Gravitational wave Observatory (LIGO; Waldman 2011) and the Laser Interferometer Space Antenna (LISA; Danzmann & Rdiger 2003), and Pulsar Timing Arrays (PTAs; Sazhin 1978; Detweiler 1979; Hellings & Downs 1983; Romani 1989; Foster & Backer 1990). The remainder of this dissertation will focus specifically on the detection and characterization of GWs using PTAs.

One of the most promising means of detecting GWs is through the precise timing of an array of millisecond pulsars (MSPs). Twenty years after their conception, three main PTAs are in full operation around the world: the North American Nanohertz Observatory for Gravitational waves (NANOGrav; Jenet et al. 2009), the Parkes Pulsar Timing Array (PPTA; Manchester 2008), and the European Pulsar Timing Array (EPTA; Janssen et al. 2008). The three PTAs collaborate to form the International Pulsar Timing Array (IPTA;

[1]At the time of writing the BICEP2 results are under scrutiny and the statistical significance may not be as large as first reported. Furthermore the results need to be confirmed by other experiments.

Hobbs et al. 2010) which will result in increased sensitivity to GWs through more data and longer time-spans than any single PTA. PTAs are most sensitive to GWs with frequencies in the nanohertz regime (i.e., 10^{-9} Hz – 10^{-7} Hz). Potential sources of GWs in this frequency range include supermassive black hole binary systems (SMBHBs) (Sesana et al. 2008), cosmic (super)strings (Olmez et al. 2010), inflation (Starobinsky 1979), and a first order phase transition at the QCD scale (Caprini et al. 2010).

While several potential sources of GWs may exist in the PTA frequency band, SMBHBs are the most likely and most studied. In the following section we will review the basic concepts of general relativity and then move on to the linearized theory of gravity and the production of GWs. We will then discuss GWs from SMBHBs by deriving their general waveform and then deriving the expected stochastic background from a superposition of single sources.

1.1 General Relativity

In this section, we will only give a brief overview of these concepts of general relativity that will be necessary for the derivation and understanding of GWs. For a much more formal description of general relativity, see a standard textbook such as Wald (1984). General relativity is a metric theory of gravity governed by Einstein's equations which describe the relationship between the four-dimensional manifold, representing spacetime, and the energy-momentum contained in that spacetime. Furthermore, general relativity is based on the notion that there are *no* inertial observers to measure the gravitational force as is done in special relativity. In order to accomplish this, general relativity states that the spacetime metric is not flat and that "background observers" simply follow geodesics of the curved spacetime metric and coincide with what was previously viewed as free-fall motion in a gravitational field. As a result, there is no meaningful way to describe gravity as a force field; but instead, gravity is an aspect of spacetime structure.

We begin by introducing the notion of the spacetime metric through the spacetime

interval[2]

$$ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta, \tag{1.1.1}$$

where $g_{\alpha\beta}$ is the spacetime metric and $\{x^\alpha\}$ is a set of coordinates. For example, in flat spacetime in cartesian coordinates we have

$$ds^2 = \eta_{\alpha\beta}dx^\alpha dx^\beta = -(dt)^2 + (dx)^2 + (dy)^2 + (dz)^2, \tag{1.1.2}$$

where $\eta_{\alpha\beta} = \text{diag}\{-1, 1, 1, 1\}$ is the flat spacetime metric and t, x, y and z are time plus the usual cartesian coordinates.[3] In general, the spacetime metric can be represented as a non-singular symmetric 4×4 matrix that is not necessarily diagonal. The spacetime metric is key to all of general relativity as specified by the principle of general covariance which states that the metric g_{ab} and quantities derivable from it are the only spacetime quantities that can appear in the equations of physics. Next we define the geodesic equation, which is the equation of motion for freely-falling (i.e., background) observers. A geodesic is a curve whose tangent vector, T^a satisfies the equation

$$T^a \nabla_a T^b = 0, \tag{1.1.3}$$

where ∇_a is the covariant derivative and

$$\nabla_a T^b = \partial_a T^b + \Gamma^b_{ac} T^c, \tag{1.1.4}$$

with Γ^b_{ac} the Christoffel symbol defined by

$$\Gamma^c_{ab} = \frac{1}{2} g^{cd} \left(\partial_a g_{bd} + \partial_b g_{ad} - \partial_d g_{ab} \right), \tag{1.1.5}$$

and ∂_a is the ordinary derivative operator[4] defined as $\partial_\alpha = \partial/\partial x^\alpha$ in a coordinate basis. Furthermore, the geodesic equation for T^α in the coordinate basis is

$$\frac{dT^\alpha}{dt} + \Gamma^\alpha_{\beta\gamma} T^\beta T^\gamma = 0, \tag{1.1.6}$$

[2]In this dissertation, we will use the Einstein index notation where repeated upper and lower indices denotes a sum (e.g., $x^\alpha y_\alpha = \sum^\alpha x_\alpha y_\alpha$). Furthermore, we use the abstract index notation where greek indices denote the components of a tensor in some coordinate system and latin indices simply represent the rank of the general tensor.

[3]We use geometerized units where $G = c = 1$.

[4]Note that the ordinary derivative obeys $\partial_a \eta_{bc} = 0$ and the covariant derivative obeys $\nabla_a g_{bc} = 0$.

where t is the affine parameter along the curve. Furthermore, in a coordinate basis the tangent vector is

$$T^\alpha = \frac{dx^\alpha}{dt}, \tag{1.1.7}$$

and the geodesic equation becomes

$$\frac{d^2x^\alpha}{dt^2} + \Gamma^\alpha_{\beta\gamma}\frac{dx^\beta}{dt}\frac{dx^\gamma}{dt} = 0. \tag{1.1.8}$$

Note that in flat space and our usual cartesian coordinate system, we recover our expression for an inertial observer (i.e., constant velocity).

The Riemann curvature tensor can be derived in many ways, none of which we will pursue here; however, we note that the curvature is directly related to a path dependent nature of parallel transport, i.e., the failure of a vector to return to its original value when parallel transported in a closed loop. Most important for our discussion of gravitational waves is the Riemann curvature tensor's role in geodesic deviation, that is, the failure of initially parallel geodesics to remain parallel. Let X^a be a deviation vector representing the displacement of a set of geodesics, then $v^a = T^b\nabla_b X^a$ is the rate of change of the deviation vector and correspondingly, the relative acceleration of the deviation vector is given by

$$\begin{aligned} a^a &= T^b\nabla_b v^a = T^c\nabla_c(T^b\nabla_b X^a) \\ &= -R_{cbd}{}^a X^b T^c T^d, \end{aligned} \tag{1.1.9}$$

where $R_{cbd}{}^a$ is the Riemann tensor defined by

$$R_{abc}{}^d = \partial_b\Gamma^d_{ac} - \partial_a\Gamma^d_{bc} + \Gamma^e_{ca}\Gamma^d_{be} - \Gamma^e_{cb}\Gamma^d_{ae}. \tag{1.1.10}$$

The geodesic deviation equation above shows that initially parallel geodesics will fail to remain parallel unless the curvature is zero. We will see in the next section how this will allow us to detect GWs. Lastly, we move on to Einstein's equation itself. The motivation for the equation comes from the Newtonian tidal acceleration of two nearby particles in a gravitational field and from Poisson's equation relating the gravitational potential to the energy density of matter. In general relativity continuous matter distributions and fields are described by a stress energy tensor T_{ab} satisfying the continuity equation $\nabla^a T_{ab} = 0$.

This covariant form of the stress energy tensor along with the non-relativistic motivations leads to the Einstein equation

$$G_{ab} = R_{ab} - \frac{1}{2}Rg_{ab} = 8\pi T_{ab}, \tag{1.1.11}$$

where G_{ab} is the Einstein tensor, $R_{ab} = R_{acb}{}^{c}$ is the Ricci tensor and $R = R_{a}{}^{a}$ is the scalar curvature. Recall that the Riemann tensor, and thus the Ricci tensor, is composed of first and second derivatives of the spacetime metric making Einstein's equations a coupled set of nonlinear second order partial differential equations for the metric components g_{ab}. However, if we are only interested in a small perturbation to the flat spacetime metric, then we can greatly simplify this set of equations.

1.1.1 Linearized Gravity

In the case of all current space and ground based GW detectors, GWs can be treated as a small *linear* perturbation to the flat spacetime metric, η_{ab}. In this section we will introduce the theory of linearized gravity and will introduce the notion of GWs. In the next section we will look at GWs specifically from supermassive black hole binary systems. We begin by writing the spacetime metric as a "small" perturbation on flat spacetime

$$g_{ab} = \eta_{ab} + h_{ab}, \tag{1.1.12}$$

where h_{ab} is our small metric perturbation. Essentially linearized gravity consists of substituting the above metric into Einstein's equations and keeping only terms linear in h_{ab}. Also note that the inverse metric is

$$g^{ab} = \eta^{ab} - h^{ab}. \tag{1.1.13}$$

Furthermore, to linear order in h_{ab} the Christoffel symbol is

$$\Gamma^{c}_{ab} = \frac{1}{2}\eta^{cd}\left(\partial_a h_{bd} + \partial_b h_{ad} - \partial_d h_{ab}\right), \tag{1.1.14}$$

and the Ricci tensor, to linear order, is

$$\begin{aligned} R_{ab} &= \partial_c \Gamma^{c}_{ab} - \partial_a \Gamma^{c}_{cb} \\ &= \frac{1}{2}\left(\partial^c \partial_b h_{ac} + \partial^c \partial_a h_{bc} - \partial^c \partial_c h_{ab} - \partial_a \partial_b h\right), \end{aligned} \tag{1.1.15}$$

where $h = h_a{}^a$ is the trace of h_{ab}. Therefore, the Einstein tensor to linear order is

$$\begin{aligned} G_{ab} &= R_{ab} - \frac{1}{2}\eta_{ab}R \\ &= \frac{1}{2}\left(\partial^c\partial_b h_{ac} + \partial^c\partial_a h_{bc} - \partial^c\partial_c h_{ab} - \partial_a\partial_b h - \eta_{ab}\left(\partial^c\partial^d h_{cd} - \partial^c\partial_c h\right)\right), \end{aligned} \tag{1.1.16}$$

which can be simplified by defining the trace reversed metric

$$\bar{h}_{ab} = h_{ab} - \frac{1}{2}\eta_{ab}h. \tag{1.1.17}$$

Now, the linearized Einstein equation is

$$-\partial^c\partial_c\bar{h}_{ab} + \partial^c\partial_b\bar{h}_{ac} + \partial^c\partial_a\bar{h}_{bc} - \eta_{ab}\partial^c\partial^d\bar{h}_{cd} = 16\pi T_{ab}. \tag{1.1.18}$$

Note that *if* the left hand side contained only the first term, then the linearized Einstein equation would simply be a flat spacetime wave equation. We would like all terms containing the divergence of the trace reversed metric to vanish (i.e., $\partial^b\bar{h}_{ab} = 0$). This can be accomplished by making an appropriate coordinate transform or gauge choice. It can be shown that linearized gravity has gauge freedom given by

$$h_{ab} \to h_{ab} + \partial_b\xi_a + \partial_a\xi_b, \tag{1.1.19}$$

which is analogous to the electromagnetic gauge freedom that allows us to add the gradient of a scalar field to the vector potential (i.e., $A_a \to A_a + \partial_a\phi$) and still obtain the same physics. In terms of the trace reversed metric, this gauge freedom is given by

$$\bar{h}_{ab} \to \bar{h}_{ab} + \partial_b\xi_a + \partial_a\xi_b - \frac{1}{2}\eta_{ab}\partial_c\xi^c. \tag{1.1.20}$$

Therefore, by solving

$$\partial^b\partial_b\xi_a = -\partial^b\bar{h}_{ab}, \tag{1.1.21}$$

for ξ_a, we have defined the Lorenz gauge condition

$$\partial^b\bar{h}_{ab} = 0 \tag{1.1.22}$$

and the linearized Einstein equation reduces to

$$\partial^c\partial_c\bar{h}_{ab} = -16\pi T_{ab}. \tag{1.1.23}$$

1.1.2 Gravitational Waves

We now move on to the description of GWs. We have seen that the solution to the linearized Einstein equation for perturbations on flat spacetime in the Lorenz gauge is a wave equation with the stress-energy tensor as a source term. In vacuum, the solution to these equations will be traveling waves. For the vacuum linearized Einstein equation we have

$$\partial^c \partial_c \bar{h}_{ab} = 0, \tag{1.1.24}$$

where we have used the Lorenz gauge to simplify things. However, there is extra gauge freedom that will simplify matters further. By solving

$$\partial^b \partial_b \xi^a = 0 \tag{1.1.25}$$

for ξ^a we obtain the transverse traceless gauge that is standard in the GW literature. While we will not show the derivation here, it is possible to use this extra gauge freedom to specify the following conditions on the metric perturbation

$$h_a{}^a = h = 0 \tag{1.1.26}$$

$$h_{a0} = 0 \tag{1.1.27}$$

$$\partial^j h_{ij} = 0, \tag{1.1.28}$$

denoting a traceless, spatial, and transverse solution to the vacuum linearized Einstein equation and the latin indices i and j denote only the spatial components of the tensor. Note that the traceless condition implies $\bar{h}_{ab} = h_{ab}$. In general the metric perturbation, h_{ab} has 10 independent components; however, this gauge choice, along with the Lorenz condition, reduce the total number of degrees of freedom form 10 to 2. First, the spatial condition removes 4 degrees of freedom. Next the traceless condition removes one more degree of freedom and finally, the transverse condition removes a final 3 degrees of freedom. Thus only two independent degrees of freedom remain and the spatial metric

perturbation (all non-spatial components are 0) can be written as[5]

$$h_{ij} = \begin{bmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{1.1.29}$$

where the naming convention of h_+ and $h_\times$ are standard and will become clear momentarily.

Thus far we have derived a form of the metric perturbation by specifying the transverse-traceless gauge, but how would one go about detecting such a perturbation? As mentioned before the physical manifestations of gravity can be measured by studying the relative acceleration of a set of "test masses". If these point masses are at rest in some global inertial coordinate system in *nearly* flat spacetime (i.e, the tangent vector $T^\alpha = \text{diag}\{1, 0, 0, 0\}$) then the relative acceleration between the test masses is

$$\frac{d^2 X^\alpha}{dt^2} = R_{\beta 00}{}^\alpha X^\beta, \tag{1.1.30}$$

where, again, X^α is the deviation vector. In the transverse traceless gauge to linear order in h, we can show that

$$R_{\alpha 00\beta} = \frac{1}{2}\ddot{h}_{\alpha\beta}. \tag{1.1.31}$$

Since we have shown that only two components of $h_{\alpha\beta}$ are independent, this means that there are only two independent components of the Riemann tensor and thus two *physical* polarization states of GWs in general relativity. These two polarization states are manifest in certain polarization patterns when a GW passes a ring of test masses as depicted in Figure 1. As we see from the figure, the plus and cross polarization states (the names are derived from the shape of the pattern) differ by an angle of $\pi/4$ rather than $\pi/2$ for electromagnetic waves. This is due to the quadrupolar nature of the wave as opposed to a dipolar nature of electromagnetic waves. Furthermore, GWs can be detected by tracking the positions of a set of "test masses" and measuring their relative separation. This is exactly what PTAs (and other ground and space based experiments) aim to do.

[5]Note that this expression applies to a GW traveling in the z-direction. This tensor can be rotated to denote GWs traveling in any direction.

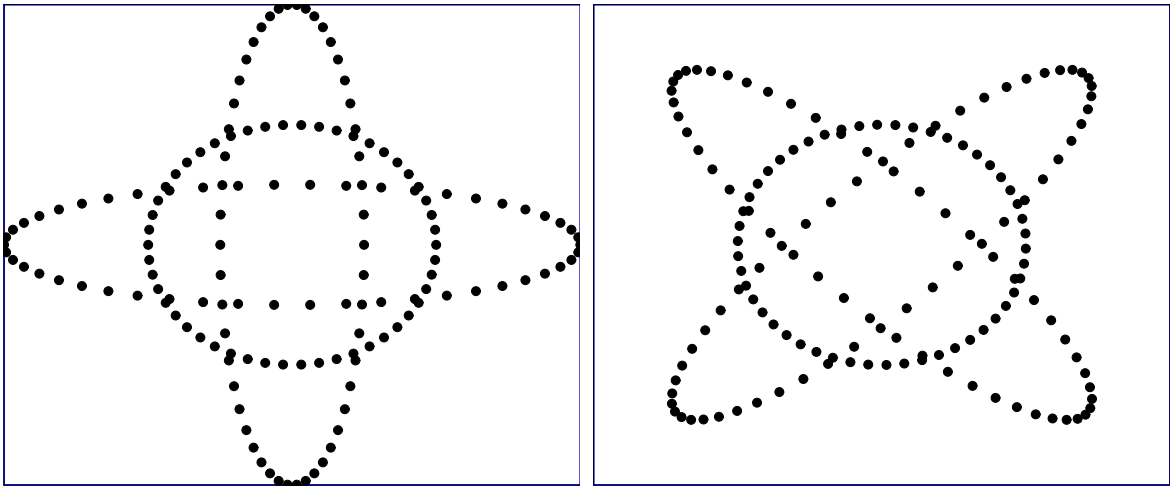


Figure 1 : Two independent polarization modes for GWs allowed by general relativity. The effect on a ring of test masses in the x-y plane due to a plus-polarized (left) and cross-polarized (right) GW travelling in the z-direction. The ring gets deformed into one of the ellipses and back during one half of the GW period and gets deformed into the other in the other half. (Image credit: Sathyaprakash & Schutz 2009)

1.1.3 Production of Gravitational Waves

In the previous section we considered the vacuum linearized Einstein equation. In this far-zone, the metric perturbation is radiative. Now we connect these far-field solutions to the near field solutions where the GWs are generated. Since the non-vacuum Einstein equation of Eq. (1.1.23) is just the flat space wave equation, the solution, in the coordinate basis is given by the retarded Green's function

$$\bar{h}^{\alpha\beta}(t,\vec{x}) = 4\int \frac{T^{\alpha\beta}(t-|\vec{x}-\vec{x}'|,\vec{x}')}{|\vec{x}-\vec{x}'|}d^3x'. \tag{1.1.32}$$

We will seek a solution in the far-zone, that is, we assume that the distance from the source to the field point r is much greater than the GW wavelength, which is much greater than the size of the source. Therefore, the approximation can be made such that $|\vec{x}-\vec{x}'| \simeq r$ is approximately constant over the source. We also assume slow motion of the source such that $t-|\vec{x}-\vec{x}'| \simeq t-r$. The metric perturbation is then

$$\bar{h}^{\alpha\beta}(t,\vec{x}) = \frac{4}{r}\int T^{\alpha\beta}(t-r,\vec{x}')d^3x'. \tag{1.1.33}$$

Making use of the conservation law $\partial_a T^{ab} = 0$, it can be shown that the spatial components of the trace-reverse metric perturbation is

$$\bar{h}^{ij}(t, \vec{x}) = \frac{2}{r}\frac{\partial^2}{\partial t^2}\int x'^i x'^j T^{00}(t - r, \vec{x}')d^3x', \tag{1.1.34}$$

where we are only interested in the spatial components because we seek an eventual solution in the transverse traceless gauge. The quadrupolar tensor is then defined as

$$I^{ij}(t) = \int x'^i x'^j T^{00}(t - r, \vec{x}')d^3x', \tag{1.1.35}$$

giving us the solution

$$\bar{h}^{ij}(t, \vec{x}) = \frac{2}{r}\ddot{I}^{ij}(t - r). \tag{1.1.36}$$

As mentioned above, we would like to give a solution in the transverse-traceless gauge. Defining the transverse projection operator

$$P_{ij} = \delta_{ij} - \hat{n}_i\hat{n}_j, \tag{1.1.37}$$

with $\hat{n}_i = x^i/r$ the unit vector in the propagation direction, the solution is

$$h^{ij}(t, \vec{x}) = \frac{2}{r}\ddot{\mathcal{I}}^{ij}(t - r), \tag{1.1.38}$$

where

$$\mathcal{I}_{ij} = P_{ik}I^{kl}P_{lj} - \frac{1}{2}P_{ij}P_{kl}I^{kl}. \tag{1.1.39}$$

So in essence, GWs are produced by any object that has a quadrupole function that is a non-linear function time (i.e, accelerating masses). Furthermore, in the following sections we will refer to the components of the metric perturbation, h_{ab} as the GW strain and can be thought of as the ratio of the change in length between two test masses and the total distance between the test masses. The main source of GWs in the pulsar timing frequency (10^{-9}Hz–10^{-6}Hz) band are supermassive black hole binaries and we will discuss them in the following sections; however, we expect detectable GWs over a very large frequency range of 10^{-18} Hz – 10^4 Hz. In Figure 2 we plot the expected gravitational wave strain vs. GW frequency along with the expected sources in different frequency ranges. We have also included the different detectors that will be sensitive to these GWs. At the lowest frequencies we expect GWs due to small fluctuations in the early universe that

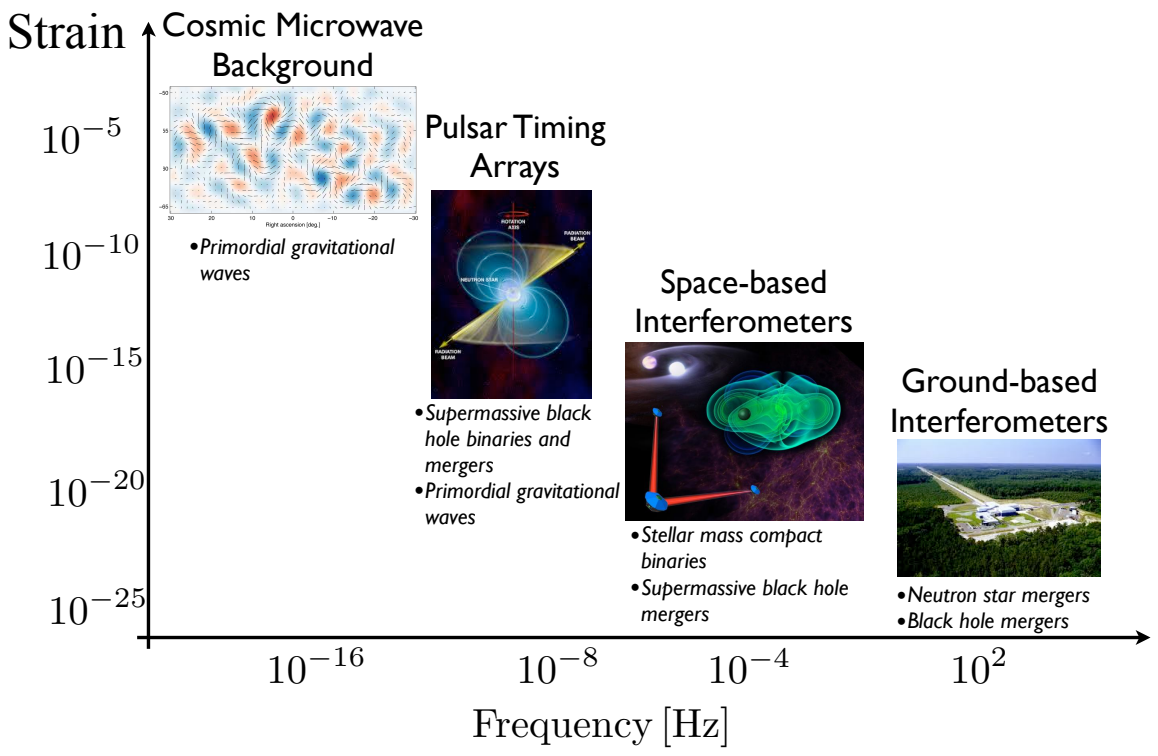


Figure 2 : Schematic spectrum of GWs and corresponding detectors along with possible sources. The current range of operable or proposed GW detectors covers nearly the entire range of expected sources of detectable GWs. (Image credit: Fredrick Jenet, Xavier Siemens)

were imprinted on the cosmic microwave background after inflation. At higher frequencies $\sim 10^{-9}$ Hz we expect a large population of supermassive black hole binaries with masses $\sim 10^9\ M_\odot$ that may be detectable with PTAs. In the millihertz frequency range, we expect a large population of stellar mass compact binaries such as white dwarf binaries and also supermassive black hole binaries with masses $\sim 10^6\ M_\odot$ potentialy detectable with a space based GW detector such as LISA. Finally at the kilohertz scale we expect neutron star binary mergers, neutron star black hole mergers, and black hole binary mergers potentially detectable with LIGO. As is evident from the figure, there is a large coordinated effort to detect and characterize GWs across the entire frequency spectrum of which PTAs are complementary.

1.2 Supermassive Black Hole Binaries

Supermassive black hole binaries (SMBHBs) are the most promising and most studied sources for the detection of GWs (either from individual sources or from a stochastic background) in the PTA frequency band. SMBHBs are ubiquitous in the low-redshift universe and are the building blocks in currently favored theories of hierarchical structure formation; however, their origin remains mostly unknown. It is well known, however, that the masses of nuclear SMBHs correlate with the velocity dispersion and luminosity of the host galaxy, an indication that the galaxies and SMBHs evolve together (Gültekin et al. 2009a). In theories of hierarchical structure formation, initial black hole seeds evolve over cosmic time driven by accretion and mergers of massive galaxies (Sesana 2013a). PTAs are sensitive to GWs from these sources; though, there is a large uncertainty in the efficiency at which galactic mergers result in SMBH mergers (radiating in the PTA frequency band). This uncertainty translates into a poorly constrained rate of SMBHB coalescences. Another potential problem is known as the "final parsec problem" in which the exact physical processes (if any!) that drive the evolution to merger after orbital separations less than one parsec is not well understood. The eventual observations of a stochastic GW background and individually resolvable sources with PTAs will provide insights into the conversion efficiency of galactic mergers into black hole mergers and possibly give some hints into the nature of the first black hole seeds. Furthermore, the detection of GWs from SMBHBs will prove conclusively that nature does indeed solve the final parsec problem.

The single SMBHB mergers that may stand out above the background can be combined with electromagnetic (EM) observations to obtain further information about galaxy mergers and galaxy formation (Tanaka & Haiman 2013). Fortunately, there are several EM SMBH tracers such as peculiar AGN emission lines, double AGN, circumbinary disk emission, nuclear periodicities, and tidal disruption events (Burke-Spolaor 2013), all of which result in prominent X-ray emission events that may be detectable with X-ray telescopes such as Chandra and NuStar.

1.2.1 Continuous Gravitational Waves from SMBMBs

In the above section, we mention that the physical processes that drive SMBHBs to small orbital separations is still somewhat unknown; however, here we assume that SMBHBs do reach an orbital separation where the orbital evolution is dominated by GW emission. We also note that throughout this dissertation, we only consider *non-spinning* black holes in *circular* orbits. Spin effects are not likely to play any measurable role in the orbital dynamics (Sesana & Vecchio 2010) and eccentric systems are possible and maybe ubiquitous (Roedig & Sesana 2012; Ravi et al. 2012; Sesana 2013b; Ravi et al. 2014) but will not be considered here.

Consider a SMBHB composed of black holes of masses m_1 and m_2 and a coordinate system such that the orbit of the binary lies in the x–y plane. We will work in center of mass coordinates where the orbital separation is a and the reduced mass and total mass are $\mu = m_1 m_2/(m_1 + m_2)$ and $M = m_1 + m_2$, respectively. We also consider uniform circular motion initially and will then make the necessary corrections to include GW emission and frequency evolution. Therefore, the orbit is given by

$$\begin{aligned} x(t) &= a\cos(\omega_s t) \\ y(t) &= a\sin(\omega_s t) \\ z(t) &= 0, \end{aligned} \tag{1.2.1}$$

where ω_s is the orbital frequency of the binary in the *source* frame. From Eq. (1.1.35) and choosing a reference frame where the center of mass is at the origin we have $I^{ij} = \mu x^i x^j$ leading to

$$\begin{aligned} I^{11} &= \mu a^2 \frac{1-\cos(2\omega_s t)}{2} \\ I^{22} &= \mu a^2 \frac{1+\cos(2\omega_s t)}{2} \\ I^{12} &= -\frac{1}{2}\mu a^2 \sin(2\omega_s t). \end{aligned} \tag{1.2.2}$$

Taking two time derivative we are left with

$$\begin{aligned} \ddot{I}^{11} &= -\ddot{I}^{22} = 2\mu a^2 \omega_s^2 \cos(2\omega_s t) \\ \ddot{I}^{12} &= \ddot{I}^{21} = 2\mu a^2 \omega_s^2 \sin(2\omega_s t). \end{aligned} \tag{1.2.3}$$

Using Kepler's third law we can write the semi-major axis a in terms of the orbital frequency as

$$a = M^{1/3}\omega_s^{-2/3}. \tag{1.2.4}$$

Introducing the chirp mass $M_c = \mu^{3/5}M^{2/5}$ and substituting into Eq. (1.1.38) then we have

$$h_+(t) = \frac{4M_c^{5/3}\omega_s^{2/3}}{r}\cos(2\omega_s t_{\rm ret}) \tag{1.2.5}$$

$$h_\times(t) = \frac{4M_c^{5/3}\omega_s^{2/3}}{r}\sin(2\omega_s t_{\rm ret}), \tag{1.2.6}$$

where $t_{\rm ret} = t - r$ is the retarded time.

GWs will radiate power away from a SMBHB source and to compensate for this loss of energy the orbital separation must decrease with time. Equivalently, though Kepler's third law, GW radiation will cause the orbital frequency to *increase* with time. By setting the power radiated in GWs equal to the change of orbital energy due to increasing orbital frequency, $-dE_{\rm orbit}/dt$, and assuming that we have quasi-circular motion (i.e., $\dot{\omega}_s \ll \omega_s^2$) we obtain

$$\dot{\omega}_s = \frac{96}{5}M_c^{5/3}\omega_s^{11/3}. \tag{1.2.7}$$

We can now use this expression to analytically solve for the orbital frequency as a function of time

$$\begin{aligned}
\int_{t_0}^{t} dt &= \frac{5}{96}M_c^{-5/3}\int_{\omega_s(t=t_0)}^{\omega_s(t)} d\omega_s\, \omega_s^{-11/3} \\
t - t_0 &= \frac{5}{256}M_c^{-5/3}\left(\omega_0^{-8/3} - \omega_s(t)^{-8/3}\right) \\
\therefore\ \omega_s(t) &= \omega_0\left(1 - \frac{256}{5}M_c^{5/3}\omega_0^{8/3}(t-t_0)\right)^{-3/8},
\end{aligned} \tag{1.2.8}$$

where t_0 is some fiducial reference time and $\omega_0 = \omega(t = t_0)$ is the initial orbital frequency. For a circular orbit, we define the phase to be

$$\frac{d\Phi}{dt} = \omega_s. \tag{1.2.9}$$

We can solve this equation similarly

$$\begin{aligned}
\int_{\Phi(t=t_0)}^{\Phi(t)} d\Phi &= \int_{t=t_0}^{t} dt' \omega_s(t') \\
\Phi(t) - \Phi_0 &= \int_{\omega(t=t_0)}^{\omega(t)} d\omega_s \, \frac{\omega_s}{\dot{\omega}_s} \\
&= \frac{5}{96} M_c^{-5/3} \int_{\omega_s(t=t_0)}^{\omega_s(t)} d\omega_s \, \omega_s^{-8/3} \\
\therefore \Phi(t) &= \Phi_0 + \frac{1}{32 M_c^{5/3}} \left(\omega_0^{-5/3} - \omega_s(t)^{-5/3} \right),
\end{aligned} \tag{1.2.10}$$

where, again, $\Phi_0 = \Phi(t = t_0)$. In order to take this frequency evolution into account in our above derivation of the strain amplitudes we replace $\omega_s t_{\rm ret}$ with $\Phi(t_{\rm ret})$ and replace the ω_s in the pre-factors with $\omega_s(t)$. In principle, we should also have terms that depend on $\dot{a}$ and $\dot{\omega}_s$ when taking the time derivates of the quadrupole tensor; however, since we are working in the quasi-circular regime, we can safely ignore both terms and write

$$h_+(t) = \frac{4 M_c^{5/3} \omega_s(t)^{2/3}}{r} \cos(2\Phi(t)) \tag{1.2.11}$$

$$h_\times(t) = \frac{4 M_c^{5/3} \omega_s(t)^{2/3}}{r} \sin(2\Phi(t)), \tag{1.2.12}$$

where we now express the result in terms of t as opposed to $t_{\rm ret}$ since we can absorb the extra term in $t_{\rm ret}$ in to our integration constants t_0 and Φ_0.

Since the sources of these GWs are SMBHBs with non-negligible redshifts we must also take cosmological effects into account when computing the induces GW strain. We will not go through the entire derivation here but instead will summarize the changes that need to be made to our above expressions. First, in an expanding universe there is a time dilation associated with the time measured by an observer relative to the time at the source

$$dt_{\rm obs} = (1 + z) dt_s. \tag{1.2.13}$$

Therefore, the frequency measured by an observer related to the source frequency via

$$f_{\rm obs} = \frac{f_s}{1 + z}. \tag{1.2.14}$$

Furthermore, in an expanding universe distances can be measured in a so-called luminosity distance

$$d_L = (1 + z) a(t_0) r, \tag{1.2.15}$$

where $a(t_0)$ is the scale factor at the present time t_0. The luminosity distance is formally defined by

$$d_L(z) = (1+z)\frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_\Lambda + \Omega_m(1+z)^3}} \tag{1.2.16}$$

where c is the speed of light (note that we explicitly include c here for clarity), H_0 is the Hubble constant, and Ω_Λ and Ω_m are the dimensionless density parameters for dark energy and matter, respectively. Finally, it is possible to show that the form of our resulting strain amplitudes will be unchanged but with the following replacements, $r \to d_L$, $M_c \to \mathcal{M} = (1+z)M_c$, and $\omega_s \to (1+z)\omega_{\rm obs}$ giving

$$h_+(t) = \frac{4\mathcal{M}^{5/3}\omega(t)^{2/3}}{d_L}\cos(2\Phi(t)) \tag{1.2.17}$$

$$h_\times(t) = \frac{4\mathcal{M}^{5/3}\omega(t)^{2/3}}{d_L}\sin(2\Phi(t)), \tag{1.2.18}$$

where we will hereafter refer to ω as the observed frequency.

Recall, that all of this was derived for a binary system in the x-y plane and an observer on the z-axis. For an observer at inclination ι to the binary, we need to rotate the quadrupole tensor by an angle ι about the x-axis and then take the transverse traceless projection. This can be done by using the rotation operator

$$\mathcal{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\iota & \sin\iota \\ 0 & -\sin\iota & \cos\iota \end{bmatrix} \tag{1.2.19}$$

and computing the matrix product $\mathcal{R}\ddot{I}\mathcal{R}^{-1}$. When this is done, we arrive at our final expression for the plus and cross GW strain functions

$$h_+(t) = \frac{2\mathcal{M}^{5/3}\omega(t)^{-2/3}}{d_L}(1+\cos^2\iota)\cos(2\Phi(t)) \tag{1.2.20}$$

$$h_\times(t) = \frac{4\mathcal{M}^{5/3}\omega(t)^{-2/3}}{d_L}\cos\iota\sin(2\Phi(t)), \tag{1.2.21}$$

where $\iota = 0$ corresponds to seeing the system "face-on" where we have circular polarization (i.e, an equal mix of plus and cross) and $\iota = \pi/2$ corresponds to linear polarization in which we only see the plus polarization mode. We have now derived the GW strain from a single SMBHB in a circular orbit. Next we compute the characteristic strain of a background of many such events.

1.2.2 Stochastic Gravitational Wave Background from SMBMBs

We now will briefly derive the expected spectral shape of the stochastic GW produced by the superposition of individual SMBHBs in circular orbits as described in the previous section. This is important, as all of the work for this dissertation and most of the stochastic background data analysis literature to date has used this simple framework. For a more complete and detailed analysis of this derivation see Phinney (2001) and Sesana et al. (2008). A stochastic background can be described in terms of the present day GW energy density per logarithmic frequency, normalized to the critical density

$$\Omega_{\rm gw}(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm gw}(f)}{d\log f}, \tag{1.2.22}$$

where ρ_c is the critical density required to close the universe, $\rho_{\rm gw}$ is the GW energy density and f is the frequency of the GW in the observed frame (i.e., $f = \omega/\pi$). The energy density for a population of GW events is simply an integral over the cosmic history of the (comoving) number density of sources, multiplied by the energy emitted by each source in the corresponding frequency range. This can then be related to characteristic GW strain amplitude as follows

$$\frac{d\rho_{\rm gw}(f)}{d\log f} = \frac{\pi}{4}f^2 h_c(f) = \int_0^\infty dz\,\frac{dn}{dz}\,\frac{1}{1+z}\,\frac{dE_{\rm gw}}{d\log f_s}\bigg|_{f_s=(1+z)f}, \tag{1.2.23}$$

where dn/dz is the comoving number density per unit redshift, $E_{\rm gw}$ is the energy output of GWs, and again f_s is the GW frequency in the *source* frame. This quite general result can be applied to a population of SMBHBs emitting GW radiation far from their last stable circular orbit. The comoving number density of sources depends on quite a few variables so let us write

$$\frac{dn}{dz} = \int_0^\infty dM_c\,\frac{d^2n}{dzdM_c}, \tag{1.2.24}$$

where, again M_c is the chirp mass and

$$\frac{d^2n}{dzdM_c} = \frac{d^3N}{dzdM_c d\log f_s}\,\frac{d\log f_s}{dt_s}\,\frac{dt_s}{dz}\,\frac{dz}{dV_c}, \tag{1.2.25}$$

where dV_c is the comoving volume shell between z and $z+dz$. Finally, it can be shown that the characteristic strain is

$$h_c^2(f) = \int_0^\infty dz \int_0^\infty dM_c \frac{d^3N}{dzdM_c d\log f_s}\bar{h}^2(f_s), \tag{1.2.26}$$

where $\bar{h}(f_s)$ is the polarization-average strain spectrum of an individual SMBMB. Again, this has the interpretation that the characteristic strain is an integral over all sources emitting in a given frequency band multiplied by the strain of each source. It is possible to ascertain the spectral shape for the stochastic GW background from this expression. Recall from Eqs. (1.2.25) and (1.2.7) the the first term in the integral scales as $f \times f^{-11/3}$ and from Eq. (1.2.20) that the second term scales as $f^{4/3}$. Therefore the overall scaling of $h_c(f)$ goes like $f^{-2/3}$. Furthermore, it is customary in the PTA literature to characterize the strain amplitude by

$$h_c(f) = A \left(\frac{f}{f_{1\mathrm{yr}}} \right)^{-2/3}, \tag{1.2.27}$$

where $f_{1\mathrm{yr}}$ is the frequency corresponding to 1 yr^{-1} and A depends on the SMBHB population and merger rate. The best estimates for the amplitude factor A range from $\sim 5 \times 10^{-16}$ – $\sim 2 \times 10^{-15}$ (Sesana 2013b; Ravi et al. 2014) and the most constraining published upper limits are just above this range at $A \leq 2.4 \times 10^{-15}$ (Shannon et al. 2013). The above derivation and standard assumption that this power-law spectral shape continues at large orbital separations (i.e., low frequencies) is present in nearly all data analysis techniques. However, other effects such as gas dynamics (Kocsis & Sesana 2011) or stellar hardening (Roedig & Sesana 2012; Ravi et al. 2014) could dominate the evolution of the SMBHBs at large orbital separations. The effect of the environment on the evolution and subsequent GW signal is still an open and active area of research and will not be discussed further in this dissertation.

1.3 Pulsars and Pulsar Timing

Pulsars are rapidly rotating, highly magnetized neutron stars first discovered in 1967 (Hewish et al. 1968). Since then, mainly due to our ability to time the radio pulses that pulsars emit to extremely high precision, pulsars have been used as physical tools to study basic physics and astrophysics and general relativity. For a review of basic astronomy and practical pulsar tools see Lorimer & Kramer (2005); for a review of GR tests with pulsars see Stairs (2003); and for a review of millisecond pulsars see Lorimer (2008). Since the work presented in this dissertation is based on pulsar timing data, we will focus primarily

on pulsar timing techniques and the use of pulsars as GW detectors; however, we will begin with a short introduction to pulsars.

Pulsars are normally detected by their radio pulses that are produced by the so-called "lighthouse effect". As the neutron star spins, charged particles are accelerated along the magnetic field lines. These accelerating particles emit electromagnetic radiation (most easily detected at radio frequencies) observed as a sequence of pulses produced as the magnetic axis crosses the line of sight of an observer. The rate of repetition of the radio pulses is simply the spin period of the neutron star. As the pulsar rotates, the outgoing radiation carries away rotational kinetic energy causing the pulse period to gradually decrease over time. Therefore, the two most basic properties that one can measure when observing a pulsar are the spin period P and the rate of spin-down $\dot{P}$. In fact, several fundamental properties of the pulsar can be determined from these two numbers including the characteristic age $\tau \propto P/\dot{P}$ and the magnetic field strength $B \propto \sqrt{P\dot{P}}$. Furthermore, studying the values of P and $\dot{P}$ tells us something about the evolutionary history and population of pulsars. Figure 3 plots the spin-down vs. the period for a nearly up-to-date population of pulsars. This plot is the radio pulsar analog to the Hertzsprung-Russel diagram and is commonly referred to as the P–$\dot{P}$ diagram. Pulsars in binary systems are highlighted by green circles. It is clear from the figure that there are two distinct populations of pulsars. The canonical pulsars (upper right) have typical spin periods on the order of seconds, large magnetic fields, and are relatively young. Conversely, the millisecond pulsars (MSPs) have typical spin periods on the order of milliseconds, small magnetic fields, and are much older. We also note that the MSPs have much lower spin-down rates, meaning that they are incredibly stable over long periods of time. As we will see in the next section, due to their rotational stability, MSPs are the best timers and therefore, all current PTAs are comprised entirely of MSPs.

Before moving on to pulsar timing, we briefly review the currently favored models (see e.g., Lorimer 2008, section 2.6 and references therin) to explain the evolutionary paths of pulsars. In Figure 4 we show a cartoon outline of possible evolutionary tracks of pulsars. Beginning with a binary system, a neutron star is formed following the supernova explosion of the initially more massive star. If the binary disrupts after the supernova

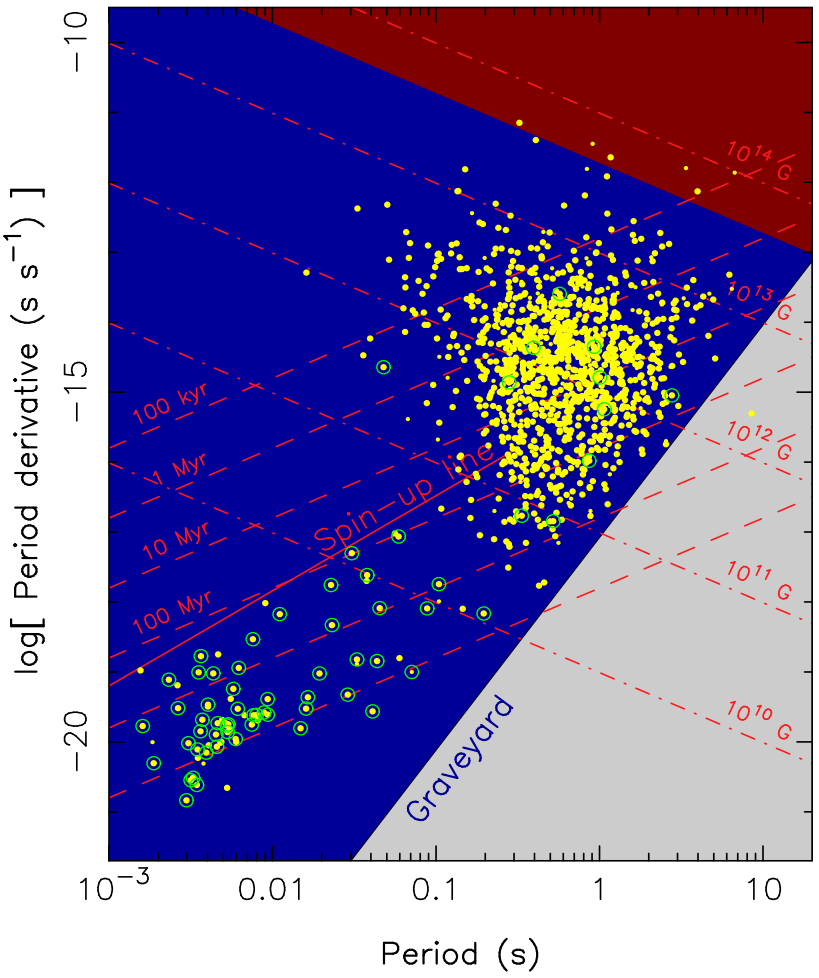


Figure 3 : P–$\dot{P}$ diagram. The binary pulsars are shown with green circles. Lines of constant characteristic age and constant magnetic field strength are also shown. (Image credit: Lorimer 2008)

explosion, we are left with a high velocity isolated neutron star and an OB runaway star. There is a high probability of disruption which qualitatively explains why so few canonical pulsars have binary companions. If the binary survives the supernova it may be observable as a radio pulsar with a massive main sequence companion. Eventually ($\sim 10^8$ yr) the pulsar will spin down so much that the energy output is no longer sufficient to produce significant amounts of radio emission and will end its life in the pulsar graveyard of the P–$\dot{P}$ diagram. For binaries that remain bound with a companion that is sufficiently massive to overflow its Roche lobe, the old spun-down pulsar can become *undead* by accreting matter and angular momentum from its massive companion. These objects are known as "recycled pulsars" and have potentially two further evolutionary tracks depending

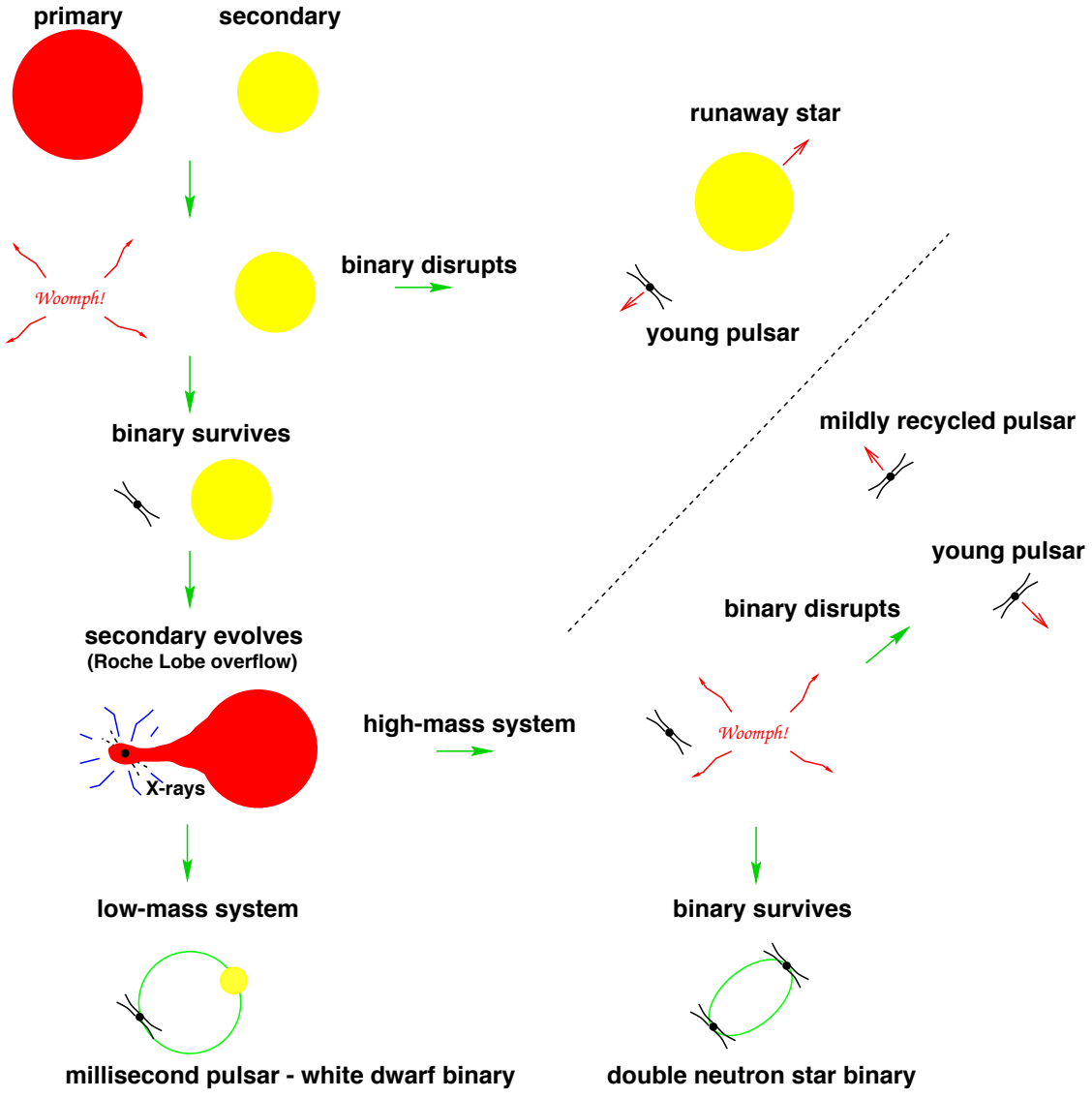


Figure 4 : Various evolutionary scenarios. (Image credit: Lorimer 2008)

on whether or not the companion is high-mass or low-mass. In the high mass case, the companion may also explode in a supernova producing a second neutron star and a double neutron star system if the binary survives the explosion. These systems lie in between the very rapidly rotating MSPs and the canonical pulsars, typically having periods of tens to hundreds of milliseconds. In the low-mass case, the binary evolves and transfers matter onto the neutron star much more slowly resulting in an MSP with periods on the order of a few milliseconds. At the end of the evolution the binary companion sheds its outer layers and becomes a white dwarf. These systems lie in the lower left of the P–$\dot{P}$

diagram and are the tools used in GW detection efforts.

1.3.1 Pulsar Timing

In the next chapter we discuss the data analysis techniques used in pulsar timing. Here we will briefly give an overview of the observation aspect of pulsar timing. The basic premise of pulsar timing is to measure the pulse times-of-arrival (TOAs) and compare them with a theoretical timing model. In practice this can be done but single pulses are generally not timed individually. The first reason is that the pulse emission is not perfectly stable, that is, the pulse-to-pulse variation in the pulse shape is quite large (see e.g. Cordes & Shannon 2012, and references therein). However, averaging a series of consecutive pulses does result in a stable average profile which can be timed with high precision. Figure 5 shows this pulse-to-pulse variation in single pulses combining to form a stable pulse profile when summed. Another reason for using averaged pulse profiles is to increase the SNR by reducing the radiometer noise. The uncertainty on the TOA is

$$\sigma_{\mathrm{TOA}} \simeq \frac{W}{SNR}, \tag{1.3.1}$$

where W is the pulse width. As we see from Figure 5, the single pulses jitter within the pulse window so that timing a single pulse would result in uncertainties $\sigma_{\mathrm{TOA}} \approx W$; however, if we use many single pulses in the average pulse profile the uncertainty scales like $\sigma_{\mathrm{TOA}} \propto 1/\sqrt{N_{\mathrm{pulses}}}$ (Lorimer & Kramer 2005). For millisecond pulsars, thousands of pulses can be averaged in a matter of minutes resulting in stable pulse profiles. Practically, determining the precise TOA works as follows: during observations, the pulse period is derived from a timing model used in previous observations (or the initial search). The data are then folded using that period (i.e., data samples with the same phase are averaged). The folded profile is then recorded with a timestamp from the observatory atomic clock. To obtain the precise TOA, this average profile is then cross-correlated with a template profile, usually constructed though the addition of many bright observations. From the cross-correlation, one can derive the phase offset between the two profiles and add that to the timestamp to create the site-arrival-time (SAT). The SATs must then be converted to barycentric-arrival-times (BATs) in order to operate in an inertial reference frame. This conversion is beyond the scope of this dissertation, see Hobbs et al. (2006) and

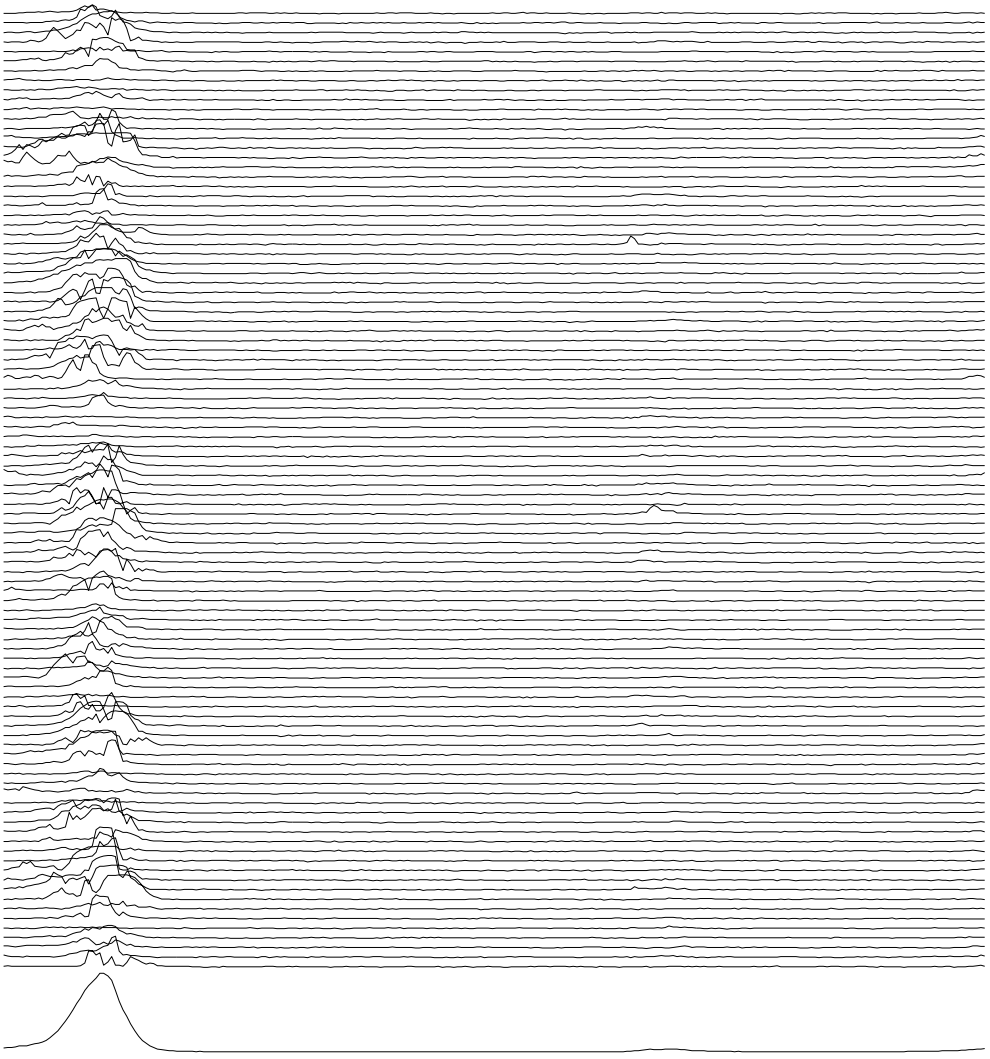


Figure 5 : Single pulses from PSR B0959+08 demonstrating the pulse-to-pulse variability. At the bottom is plotted the averaged pulse profile from summing 1200 pulses. (Image credit: Stairs 2003)

Edwards et al. (2006) for details of this process. Finally, the BATs are subtracted from arrival times predicted by a timing model for the pulsar to form the *residuals*. It is the goal of pulsar timing to minimize these residuals by finding the best model that fits the data, the details of which will be discussed in the next section. Current PTAs are able to correctly predict the TOAs to nanosecond precision with stability over several years while still attaining sub-100 nanosecond root-mean-squared (RMS) residuals for several pulsars (van Haasteren et al. 2011; Demorest et al. 2013; Manchester et al. 2013).

1.3.2 Pulsars as GW Detectors

Above, we described the method for constructing pulsar timing residuals by constructing a timing model and subtracting it from measured TOAs; however, in general, the timing model will not include perturbations due to GWs. The idea of using pulsars to detect low-frequency GWs was simultaneously developed by Sazhin (1978) and Detweiler (1979) in which they compute the expected induced residuals from GWs. Here we will formally derive the response function of PTAs to incident GWs.

The metric perturbation can be written as follows

$$h_{ab}(t,\hat{\Omega}) = \sum_{A=\{+,\times\}} e^{A}_{ab}(\hat{\Omega}) h_A(t) = e^{+}_{ab}(\hat{\Omega}) h_+(t) + e^{\times}_{ab}(\hat{\Omega}) h_\times(t), \tag{1.3.2}$$

where $\hat{\Omega}$ is the unit vector pointing from the GW source to the Solar System Barycenter (SSB), h_+, $h_\times$ and e^A_{ab} ($A = +, \times$) are the polarization amplitudes and polarization basis tensors, respectively. The polarization tensors can be converted to the SSB by the following transformation. Following Wahlquist (1987) we write

$$\begin{aligned} e^{+}_{ab}(\hat{\Omega}) &= \hat{m}_a \hat{m}_b - \hat{n}_a \hat{n}_b, \\ e^{\times}_{ab}(\hat{\Omega}) &= \hat{m}_a \hat{n}_b + \hat{n}_a \hat{m}_b, \end{aligned} \tag{1.3.3}$$

where

$$\begin{aligned} \hat{\Omega} &= -(\sin\theta\cos\varphi)\hat{x} - (\sin\theta\sin\varphi)\hat{y} - (\cos\theta)\hat{z}, \\ \hat{m} &= -(\sin\varphi)\hat{x} + (\cos\varphi)\hat{y}, \\ \hat{n} &= -(\cos\theta\cos\varphi)\hat{x} - (\cos\theta\sin\varphi)\hat{y} + (\sin\theta)\hat{z}. \end{aligned} \tag{1.3.4}$$

In this coordinate system, $\theta = \pi/2 - \delta$ and $\varphi = \alpha$ are the polar and azimuthal angles of the source, respectively, where δ and α are declination and right ascension in usual equatorial coordinates, where the North Celestial Pole is in the $\hat{z}$ direction and the Vernal Equinox is in the $\hat{x}$ direction. We now would like to determine the effect of a passing GW on the redshift of the radio pulse. Let $\hat{p}$ be the unit that points from the Earth to the pulsar and $\hat{\Omega}$ is the unit vector that points from the GW source to earth. The redshift of the radio signal is then

$$z = \frac{\delta\nu}{\nu} = \frac{\nu_p - \nu_e}{\nu_e}, \tag{1.3.5}$$

where the "e" and "p" subscripts denote the receiver (the earth) and the emitter (the pulsar). Furthermore, the radio signal follows a null geodesic k^α with components $k^\alpha = \nu(1, -\hat{p})$. Since k^α is a null vector it is parameterized by an affine parameter λ. The geodesic equation of (1.1.6) requires

$$\frac{dk^\alpha}{d\lambda} = -\Gamma^\alpha_{\beta\gamma} k^\beta k^\gamma. \tag{1.3.6}$$

We see that the time component of the above equation is

$$\frac{dk^0}{d\lambda} = \frac{d\nu}{d\lambda} = -\Gamma^0_{\beta\gamma} k^\beta k^\gamma. \tag{1.3.7}$$

To determine the redshift, we must compute $d\nu/d\lambda$ and integrate over the geodesic from the pulsar to the earth. Using Eq. (1.1.5) and the fact that the metric perturbation is purely spatial, the only relevant surviving components of the Chirstoffel symbols are

$$\Gamma^0_{ij} = \frac{1}{2}\frac{\partial h_{ij}}{\partial t} \tag{1.3.8}$$

where the latin subscripts represent the spatial components. From the above expressions, the quantity of interest can be written as

$$\frac{d\nu}{d\lambda} = -\nu^2 \Gamma^0_{ij} \hat{p}^i \hat{p}^j. \tag{1.3.9}$$

If we denote $\vec{x} = x\hat{p}$ as a vector along the earth-pulsar line of sight and recall that h_{ij} is a plane wave and thus a function of $t - \hat{\Omega} \cdot \vec{x}$ then the total derivative of the metric perturbation with respect to λ is then

$$\frac{dh_{ij}(t - \hat{\Omega} \cdot \vec{x})}{d\lambda} = \frac{dt}{d\lambda}\frac{\partial h_{ij}}{\partial t} + \frac{d(\hat{\Omega} \cdot \vec{x})}{d\lambda}\frac{\partial h_{ij}}{\partial(\hat{\Omega} \cdot \vec{x})}. \tag{1.3.10}$$

Since $k^\alpha = \nu(1, \hat{p})$, then $k^0 = dt/d\lambda = \nu$ and $k^i = d\vec{x}/d\lambda = -\nu\hat{p}$. Furthermore, $\partial h_{ij}/\partial(\hat{\Omega} \cdot \vec{x}) = -\partial h_{ij}/\partial t$ and the above derivative becomes

$$\frac{dh_{ij}(t - \hat{\Omega} \cdot \vec{x})}{d\lambda} = \nu(1 + \hat{\Omega} \cdot \hat{p})\frac{\partial h_{ij}}{\partial t}. \tag{1.3.11}$$

Substituting this back in to Eq. (1.3.9)

$$\frac{d\nu}{d\lambda} = -\frac{\nu}{2}\frac{\hat{p}^i \hat{p}^j}{1 + \hat{\Omega} \cdot \hat{p}}\frac{dh_{ij}}{d\lambda}. \tag{1.3.12}$$

Integrating both sides we have

$$\begin{aligned} -\int_{\lambda_p}^{\lambda_e} \frac{d\log\nu}{d\lambda} d\lambda = \log(\nu_p/\nu_e) &= \frac{1}{2}\frac{\hat{p}^i\hat{p}^j}{1+\hat{\Omega}\cdot\hat{p}} \int_{\lambda_p}^{\lambda_e} \frac{dh_{ij}}{d\lambda} d\lambda \\ &= \frac{1}{2}\sum_A \frac{\hat{p}^i\, e_{ij}^A(\hat{\Omega})\,\hat{p}^j}{1+\hat{\Omega}\cdot\hat{p}} \int_{t_p}^{t_e} \frac{dh_A(t)}{dt} dt, \end{aligned} \tag{1.3.13}$$

and since $\log(\nu_p/\nu_e) = \log(1+z) \approx z$, the redshift of the radio pulse induced by the GW is then

$$\begin{aligned} z(t,\Omega) = \frac{\delta\nu}{\nu} &= \frac{1}{2}\sum_A \frac{\hat{p}^i\, e_{ij}^A(\hat{\Omega})\,\hat{p}^j}{1+\hat{\Omega}\cdot\hat{p}} \Delta h_A(t) \\ &= \sum_A F^A(\hat{\Omega}) \Delta h_A(t) \end{aligned} \tag{1.3.14}$$

where we have used Eq. (1.3.4) to define the antenna pattern functions

$$\begin{aligned} F^+(\hat{\Omega}) &= \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}} \\ F^\times(\hat{\Omega}) &= \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})}{1+\hat{\Omega}\cdot\hat{p}}, \end{aligned} \tag{1.3.15}$$

which are geometric functions that describe a pulsar's sensitivity to GWs as a function of their sky location and

$$\Delta h_A(t) = h_A(t_e,\psi) - h_A(t_p,\psi), \tag{1.3.16}$$

where t_e and $t_p = t_e - L(1+\hat{\Omega}\cdot\hat{p})$ are the times at which the GW wavefront passes the earth and pulsar, respectively, and L is the distance to the pulsar. Henceforth we will drop the subscript on t_e and simply refer to t as the time measured on earth. ψ is the polarization angle that comes about from a rotation by an angle ψ in the transverse plane, similar to our previous rotation that leads to the inclination angle dependence. With the inclusion of this final rotation angle, we have

$$\begin{aligned} h_+(t,\psi) &= h_+(t)\cos(2\psi) - h_\times(t)\sin(2\psi) \\ h_\times(t,\psi) &= h_\times(t)\cos(2\psi) + h_+(t)\sin(2\psi), \end{aligned} \tag{1.3.17}$$

where $h_+(t)$ and $h_\times(t)$ are those defined in Eq. (1.2.20). The GW induced residuals are defined as the integral of the redshift over time

$$s(t,\hat{\Omega}) = \int_0^t \frac{\delta\nu}{\nu} dt = \sum_A F^A(\hat{\Omega}) \int_0^t \Delta h_A(t)\, dt. \tag{1.3.18}$$

In general, the right hand side of this equation cannot be calculated analytically; however, recall that when we derived the strain amplitude in Eq. (1.2.20) we have assumed quasi-circular orbits stating that the orbital frequency ω is a slowly varying function of time. Further recall that the strain amplitudes only had time dependence in the orbital frequency and phase, then we can carry out the above integrals nearly exactly by using the following approximation

$$\begin{aligned} I &= \mathscr{A}\int_0^t \omega(t)^{2/3}\sin[2\Phi(t)]dt \\ &\simeq \mathscr{A}\omega(t)^{2/3}\int_0^{\Phi(t)} \sin[2\Phi(t)]\left(\frac{d\Phi(t)}{dt}\right)^{-1} d\Phi \\ &\approx -\mathscr{A}\omega(t)^{-1/3}\cos[2\Phi(t)]. \end{aligned} \tag{1.3.19}$$

Therefore, in our case, when converting from strain to residuals one only needs to make the following substitutions in Eq. (1.2.20)

$$\begin{aligned} \sin[2\Phi(t)] &\to -\omega(t)^{-1}\cos[2\Phi(t)] \\ \cos[2\Phi(t)] &\to \omega(t)^{-1}\sin[2\Phi(t)]. \end{aligned} \tag{1.3.20}$$

With these definitions we can now write the plus and cross components of the GW induced timing residuals

$$s_+(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}\Big[-\sin[2\Phi(t)](1+\cos^2\iota)\cos 2\psi - 2\cos[2\Phi(t)]\cos\iota\sin 2\psi\Big] \tag{1.3.21}$$

$$s_\times(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}\Big[-\sin[2\Phi(t)](1+\cos^2\iota)\sin 2\psi + 2\cos[2\Phi(t)]\cos\iota\cos 2\psi\Big], \tag{1.3.22}$$

and the total GW induced residuals are

$$s(t,\hat{\Omega}) = \sum_A F^A(\hat{\Omega})\left[s_A(t) - s_A(t_p)\right]. \tag{1.3.23}$$

Make note that with the inclusion of the pulsar term, our residuals measured today contain information about the binary system thousands of years in the past (since the time delay is proportional to the pulsar distance which is typically on the order of 1 kpc). This unique feature of pulsar timing can allow us to study the orbital dynamics of the SMBHB system even if the system evolves little over our $\sim$ 10 year observation time. This waveform is a function of several unknown parameters which must be mapped out in order to detect such a signal. Furthermore, detection of a sinusoid of this type in a

single pulsar is ambiguous as it cannot be disentangled from any other intrinsic noise signal. To confidently detect such a GW signal we will need to be sure that the same signal with the correct weighting given by the antenna pattern functions is present in a large set of timing residuals from many pulsars. A large part of this dissertation work has dealt with robust and efficient ways of completing this task.

Above, we have derived the characteristic signal that GWs from an individual SMBHB would induce in the measured pulsar timing residuals. Now we turn to a different kind of characteristic signal that is induced by a background of SMBHBs. Here we strive to give a fairly complete picture of the effects of a stochastic GWB on pulsar timing residuals; however, for a complete derivation and more details see Allen & Romano (1999) and Anholm et al. (2009). We note the metric perturbation written in Eq. (1.3.2) can be integrated over the sky to obtain

$$h_{ab}(t) = \sum_A \int_{-\infty}^{\infty} df \int_{S^2} d\Omega\, h_A(f,\hat{\Omega})\, e^{i2\pi f(t-\hat{\Omega}\cdot\vec{x})}\, e^A_{ab}(\hat{\Omega}), \tag{1.3.24}$$

where S^2 denotes the unit two-sphere, $\vec{x}$ is a vector pointing along the earth-pulsar line of sight, and $h_A(f,\hat{\Omega})$ are complex functions satisfying $h_A(-f,\hat{\Omega}) = h^*_A(f,\hat{\Omega})$, where $*$ denotes the complex conjugate. As we assumed above in Section 1.2.2 the stochastic background is isotropic, unpolarized, and stationary. These assumptions imply that the expectation value of the Fourier amplitudes satisfies

$$\langle h^*_A(f,\hat{\Omega}) h_{A'}(f',\hat{\Omega}')\rangle = \delta^2(\hat{\Omega},\hat{\Omega}')\delta_{AA'}\delta(f-f')H(f), \tag{1.3.25}$$

where $\delta^2(\hat{\Omega},\hat{\Omega}')$ denotes isotropy, $\delta_{AA'}$ denotes unpolarized, $\delta(f-f')$ denotes stationarity, and $H(f)$ is a real, non-negative function related to the GW spectrum. Lastly, we also assume that the stochastic background has zero mean (i.e, $\langle h_A(f,\hat{\Omega})\rangle = 0$). Since the stochastic background is assumed to be gaussian, the above expectation values completely specify its statistical properties. It can be shown that the function $H(f)$ can be written in terms of the GW energy density of Eq. (1.2.22) as

$$H(f) = \frac{3H_0^2}{32\pi^3} f^{-3}\Omega_{\rm gw}(f). \tag{1.3.26}$$

Above, we have shown the GW induced residuals are simply the induced redshift integrated over time, which itself is an integral of the GW strain along a null geodesic

from the pulsar to the earth. In the frequency domain, the GW induced residuals are

$$s(f,\hat{\Omega}) = \frac{1}{2\pi i f}\left(e^{-2\pi i f L(1+\hat{\Omega}\cdot\hat{p})} - 1\right)\sum_A F^A(\hat{\Omega})h_A(f,\hat{\Omega}), \tag{1.3.27}$$

where the term in parendissertation comes from the $\Delta h_A(t)$ term in the time domain, denoting the difference in the strain at the pulsar and earth. Now, since the GW background is a stochastic process and does not have a deterministic signal model which we can extract from the detector noise we must instead base our detection scheme on its *statistical* properties, in particular the power spectrum and cross-power spectrum of the residuals. The quantity of interest is then

$$\langle s_I(f,\hat{\Omega})s_J^*(f',\hat{\Omega}')\rangle = \frac{1}{4\pi^2 f^2}\langle h_{A,I}^*(f,\hat{\Omega})h_{A',J}(f',\hat{\Omega}')\rangle\Gamma_{IJ}(f), \tag{1.3.28}$$

where the subscripts I and J denote separate pulsars and

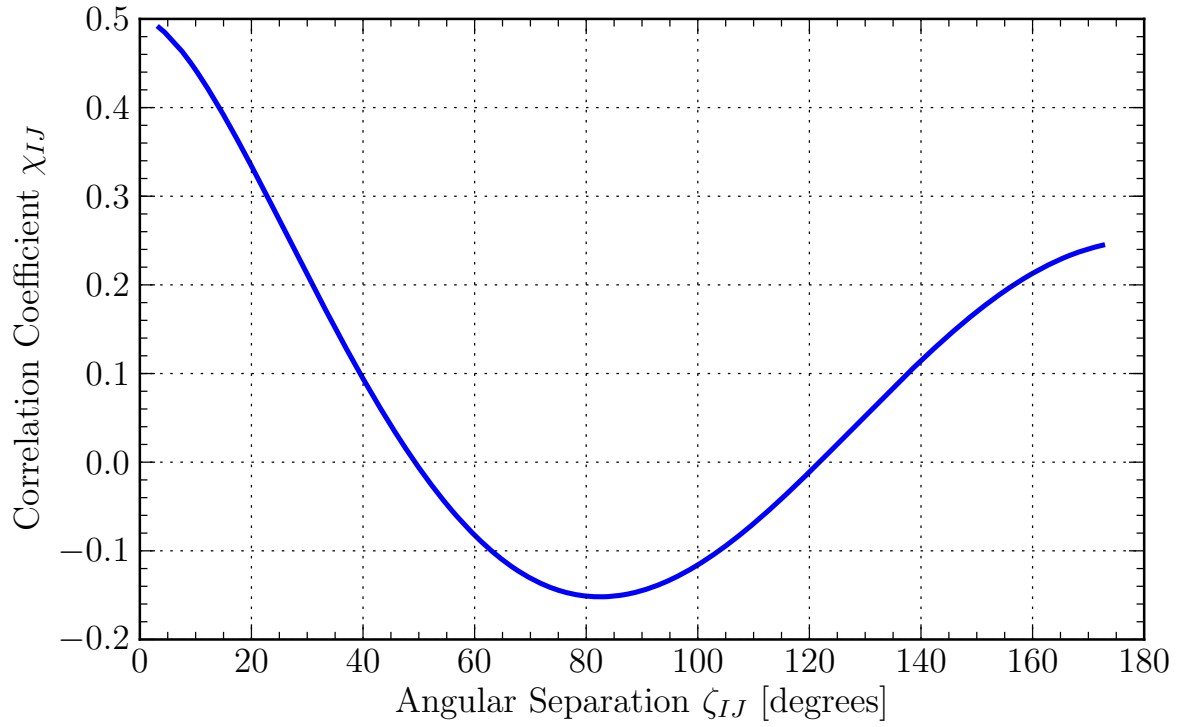


Figure 6 : Hellings and Downs correlation coefficients, χ_{IJ} as a function of the angular separation of pulsar pairs, ζ_{IJ}.

$$\Gamma_{IJ}(f) = \sum_A \int_{S^2} d\Omega \left(e^{-2\pi i f L_I(1+\hat{\Omega}\cdot\hat{p}_i)} - 1\right)\left(e^{-2\pi i f L_J(1+\hat{\Omega}\cdot\hat{p}_J)} - 1\right) F_I^A(\hat{\Omega})F_J^A(\hat{\Omega}) \tag{1.3.29}$$

is the overlap reduction function. It was shown in Anholm et al. (2009) that for PTA frequencies and typical pulsar distances, the terms involving the pulsar distances can be safely ignored and the resulting integral is

$$\Gamma_{IJ}(f) = \frac{8\pi}{3}\chi_{IJ} \tag{1.3.30}$$

where

$$\chi_{IJ} = \frac{3}{2}\left[\frac{1}{3} + \frac{1-\cos\zeta_{IJ}}{2}\left[\ln\left(\frac{1-\cos\zeta_{IJ}}{2}\right) - \frac{1}{6}\right]\right] \tag{1.3.31}$$

is the standard Hellings-Downs coefficient (Hellings & Downs 1983) with ζ_{IJ} the angular separation of the I, J-th pulsar pair. These correlation coefficients plotted in Figure 6 are the *unique* signature of the stochastic GW background that all data analysis pipelines (Jenet et al. 2005; Anholm et al. 2009; van Haasteren et al. 2009a; Yardley et al. 2011) search for. From Eqs. (1.3.25), (1.3.26), and (1.3.28) we see that the power spectrum of the GW induced residuals is

$$P_{IJ}(f) = \chi_{IJ}P_g(f), \tag{1.3.32}$$

where

$$P_g(f) = \frac{1}{24\pi^2}h_c(f)^2 = \frac{A^2}{24\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{2\alpha} f^{-3}, \tag{1.3.33}$$

where $\alpha = -2/3$ for a SMBHB stochastic background. Thus, we have shown that the characteristic signature of an isotropic, unpolarized, stationary, and Gaussian stochastic GW background is fully encoded in the pulsar timing residuals via the Hellings and Downs correlation coefficients of Eq. (1.3.31) and the GW power spectrum of Eq. (1.3.33), which in turn is a simple function of the strain spectrum of Eq. (1.2.27).

1.4 Dissertation Summary

This dissertation is composed of a selection of papers that were published during my Ph.D work as well as an extended introduction to the data analysis methods used in modern PTA data analysis. Chapter 2 begins with a brief introduction to Bayes theorem and Bayesian data analysis that is pervasive throughout this dissertation. We then discuss

the data analysis techniques used in construction of the pulsar timing residuals and corresponding pulsar timing model parameters. Next we note that if one wants model other elements in the TOAs such as gravitational waves or additional noise sources (beyond the standard template fitting uncertainty) then we must move beyond standard pulsar timing techniques. From this assertion we construct two forms of the pulsar timing likelihood function that are used in all subsequent GW and noise characterization work. The first likelihood function is derived by assuming the operation that converts *pre*-fit to *post*-fit residuals is simply a linear transformation. The second form of the likelihood comes from including all signals including the *linear* timing model, GWs, noise parameters, etc., in a joint likelihood and then marginalizing over the timing model parameters. Next, we discuss in detail the most general noise model that is used in modern PTA data analysis techniques and write the likelihood in a more computationally efficient manner. Finally we conclude this chapter by discussing Markov Chain Monte-Carlo algorithms including specifics on how to compute the Bayesian evidence and how to construct efficient jump proposals.

Chapter 3 is the first of three chapters discussing data analysis techniques and results for continuous GWs. In this chapter we introduce a frequentist detection technique known as $\mathcal{F}$-statistic, which is simply a likelihood ratio maximized over several of the parameters in the continuous GW signal. We derive a coherent and incoherent $\mathcal{F}$-statistic which we denote as the $\mathcal{F}_e$ and $\mathcal{F}_p$, respectively. The $\mathcal{F}_e$-statistic ignores the pulsar term and only models the signal with the coherent earth term and the $\mathcal{F}_p$-statistic models the total *power* in the residuals from both the earth and the pulsar term. An outline for a detection and upper limit pipeline is then discussed along with several simulations to test the efficacy of these detection statistics.

Chapter 4 details a complete Bayesian analysis pipeline for the detection and characterization of continuous GW signals in PTA data. Here we include the pulsar term completely in our analysis and map out the entire combined parameter space of GW parameters as well as pulsar distances via MCMC. We test this pipeline on semi-realistic simulated IPTA data. We show that we can efficiently map out this parameter space to perform parameter estimation and that we can also evaluate the Bayesian evidence to

perform model selection.

Chapter 5 uses the techniques developed in the previous chapters to provide a complete continuous wave analysis of the 5-year NANOGrav dataset. We begin by reviewing upper limit and detection protocols both in the Bayesian and frequentist framework and show that a truly *robust* pipeline must include the noise, GW and timing model parameters in the analysis simultaneously in order to avoid biases in parameter estimation and false detections in model selection. We then present the results of several detection pipelines in both Bayesian and frequentist frameworks. We show that there is no evidence for any continuous GWs in this dataset. Next, we compute upper limits on the strain amplitude of any continuous GW that *could* be present in the dataset. We do this using both the frequentist $\mathcal{F}_p$-statistic and the Bayesian method computing both sky-averaged and sky-dependent upper limits.

In Chapter 6 we switch from continuous GW analysis to stochastic GW background analysis. At the time of writing, full Bayesian analyses are extremely computationally intensive and completely infeasible in some cases. This is due to the fact that one must invert a very large covariance matrix (describing the correlations in the residuals for *all* pulsars) when computing the likelihood function. Since we must include the noise parameters simultaneously with the GW parameters in this search we must compute this likelihood ~ 1 million times in our MCMC in order to fully sample the posterior. In this chapter, we introduce an approximation to the inverse covariance that only require us to invert the covariance matrix for each *single* pulsar as opposed to the combined covariance matrix for all pulsars. Since matrix inversion is an $\mathcal{O}(n^3)$ process, the corresponding computational speedup is proportional to the square of the number of pulsars in the array. We carry out several tests to show that there is no significant bias in parameter estimation when compared with the full likelihood function.

Lastly, in Chapter 7 we summarize this dissertation work and discuss its implications. We also summarize current and future work in this field.

Chapter 2

Data Analysis Methods for PTAs

> "*This is one corner... of one country, in one continent, on one planet thats a corner of a galaxy thats a corner of a universe that is forever growing and shrinking and creating and destroying and never remaining the same for a single millisecond. And there is so much, so much to see.*"
>
> — The Doctor, *Doctor Who: The Power of Three*

2.1 Introduction

In general, there are two main schools of thought in the interpretation of probability: frequentist and Bayesian. In the following, we assume that we are interested in detecting and characterizing a signal in noisy data. From a frequentist viewpoint, the data are random while the signal parameters are fixed but unknown (i.e., we construct probability distributions for the data, or rather some function of the data, given a set of signal parameters), whereas in the Bayesian framework the data are fixed and the signal parameters are uncertain (i.e., we construct probability distributions of the signal parameters given a dataset). In other words, the more traditional frequentist approach is interested in the long term relative *frequency* of measuring the unknown signal parameters over many realizations of data (i.e, the experiment is repeated many times), and the Bayesian approach is interested in our *degree of belief* in the signal parameters given a single realization of data.

Furthermore, inference problems such as this can be further subdivided into two broad

categories: model selection and parameter estimation. For model selection, we are interested in determining which model of the signal is most favored by the data, and for parameter estimation we are interested in the values of the parameters that describe a given model. Both frequentist and Bayesian frameworks have tools to address each of these problems. For our work we focus primarily on Bayesian methods and will defer further discussion of frequentist methods to chapter 3.

While Bayesian parameter estimation and model selection has been commonplace in LIGO and LISA (Cornish & Crowder 2005; van der Sluys et al. 2008, 2009; Littenberg & Cornish 2009; Littenberg 2011; Veitch et al. 2012; the LIGO Scientific Collaboration et al. 2013), many PTA applications have been more frequentist in nature (Jenet et al. 2004, 2005; Anholm et al. 2009; Yardley et al. 2011, 2010; Babak & Sesana 2012; Ellis et al. 2012c; Petiteau et al. 2013) and only recently has the Bayesian framework been put to use in the PTA context (van Haasteren et al. 2009b; van Haasteren & Levin 2010; Corbin & Cornish 2010; Finn & Lommen 2010; van Haasteren & Levin 2013; Ellis et al. 2013; Lentati et al. 2013b; Taylor et al. 2012). Here we will briefly review Bayesian inference for clarity of notation. We then derive the likelihood function that is key to all of our data analysis efforts. Lastly we will review the Markov Chain Monte Carlo (MCMC) algorithm in detail as it is an essential tool for PTA data analysis.

2.2 Bayes Theorem

In the Bayesian framework, the data d are assumed to be fixed and the parameters Θ that parameterize a hypothesis (or model) $\mathcal{H}$ are assumed to follow a given prior distribution. The data are used to update our prior knowledge of the hypothesis $p(\Theta|\mathcal{H})$ via Bayes theorem,

$$p(\Theta|d,\mathcal{H}) = \frac{p(d|\Theta,\mathcal{H})p(\Theta|\mathcal{H})}{p(d|\mathcal{H})}, \tag{2.2.1}$$

where $p(\Theta|d,\mathcal{H})$ is the posterior probability distribution, that is, the probability that the set of parameters Θ for hypothesis $\mathcal{H}$ could generate the given data d. In the above expression $p(d|\Theta,\mathcal{H})$ is the likelihood function, the probability that this dataset d is drawn from a random distribution described by hypothesis $\mathcal{H}$ parameterized by Θ. Lastly, the

prior $p(\Theta|\mathcal{H})$ encompasses any prior knowledge we have about the given hypothesis and $p(d|\mathcal{H})$ is the marginalized likelihood or evidence

$$p(d|\mathcal{H}) = \int d\Theta\, p(d|\Theta, \mathcal{H}) p(\Theta|\mathcal{H}). \tag{2.2.2}$$

For the purposes of parameter estimation we can safely ignore the evidence in Bayes theorem since it is just a normalization factor that does not depend on the model parameters Θ. However, if we want to perform model selection to claim a detection or compare different signal models then computing the evidence is crucial. In this case we can make use of the Bayesian odds ratio between models "A" and "B"

$$\mathcal{O} = \frac{p(d|\mathcal{H}_A)}{p(d|\mathcal{H}_B)} \frac{p(\mathcal{H}_A)}{p(\mathcal{H}_B)}, \tag{2.2.3}$$

where the first ratio is known as the Bayes Factor, which quantifies our confidence in one model over the other based on the data (henceforth we will denote the Bayes factor as $\mathcal{B}$), and the second ratio is the prior odds ratio for models A and B, which describes our prior belief in both models. In this dissertation we consider only the Bayes factor, and assume the prior odds are even. (The choice of the prior odds will determine the false-alarm rate of a detection scheme based on the odds ratio (Vallisneri 2012)).

2.3 Pulsar Timing Data Analysis

In pulsar timing, we measure the times-of-arrival (TOAs) of radio pulses emitted from pulsars. These TOAs contain many terms of known functional form (pulse period, spin down, etc.) as well as several noise sources. Let the TOAs for a pulsar be given by

$$t^{\rm obs} = t^{\rm det}(\xi_{\rm true}) + n, \tag{2.3.1}$$

where $t^{\rm obs}$ is a vector of the observed TOAs, $t_{\rm det}(\xi_{\rm true})$ is a vector of the deterministic timing model parameterized by timing model parameters $\xi_{\rm true}$, and n is a vector of the noise in the measurements which will be taken to be Gaussian. Assuming we have an initial estimate of the true timing model parameters, $\xi_{\rm est}$ (either from information gained when discovering the pulsar or past timing observations), then we can form the pre-fit residuals

$$\delta t^{\rm pre} = t^{\rm obs} - t^{\rm det}(\xi_{\rm est}) = t^{\rm det}(\xi_{\rm true}) - t^{\rm det}(\xi_{\rm est}) + n. \tag{2.3.2}$$

If we now write the estimated parameters, $\xi_{\rm est}$, in terms of the true parameters, $\xi_{\rm true}$, as $\xi_{\rm true} = \xi_{\rm est} + \epsilon$, where ϵ is an offset parameter, then the above expression becomes

$$\begin{aligned}
\delta t^{\rm pre} &= t^{\rm det}(\xi_{\rm est} + \epsilon) - t^{\rm det}(\xi_{\rm est}) + n \\
&= t^{\rm det}(\xi_{\rm est}) + \left.\frac{\partial t^{\rm det}(\xi_{\rm est} + \epsilon)}{\partial \epsilon}\right|_{\epsilon=0} \epsilon - t^{\rm det}(\xi_{\rm est}) + n + \mathcal{O}(\epsilon^2) \\
&\approx \left.\frac{\partial t^{\rm det}(\xi_{\rm est} + \epsilon)}{\partial \epsilon}\right|_{\epsilon=0} \epsilon + n \\
&= M\epsilon + n,
\end{aligned} \tag{2.3.3}$$

where M is the design matrix and we have assumed that our initial estimate of the timing model parameters is sufficiently close to the true values so that we can approximate this as a linear system of equations in a small offset parameter vector, ϵ. In standard pulsar timing analysis, it is customary to obtain the best fit $\hat{\epsilon}$ values through a weighted least squares minimization of the pre-fit residuals. In the most general case we should be performing a *generalized* least squares fit using a general covariance matrix for the noise n; however, in most cases we have no a priori knowledge of this covariance matrix and therefore assume that it is just diagonal with elements σ_i^2, where σ_i is the uncertainty of the ith TOA. Previous work (Coles et al. 2011) has used an iterative method to estimate the covariance matrix of the residuals and apply a generalized least squares fit. Here we will perform a slightly different yet equivalent analysis. Instead of performing a least-squares fit and *minimizing* the chi-squared distribution, we will instead *maximize* the likelihood function for the residuals.

For our purposes here we will assume the the noise n follows Gaussian statistics and will defer a more detailed discussion for section 2.5. The likelihood function for Gaussian noise is

$$p(n|\vec{\phi}) = \frac{1}{\sqrt{\det(2\pi C)}} \exp\left(-\frac{1}{2} n^T C^{-1} n\right), \tag{2.3.4}$$

where $\vec{\phi}$ is a vector of parameters for a given model of n, and

$$C(\vec{\phi}) = \langle n n^T \rangle \tag{2.3.5}$$

is the covariance matrix of the noise, where $\langle \cdot \rangle$ represents the ensemble average. Using

Eq. (2.3.3) we write the likelihood function of the pre-fit residuals

$$p(\delta t^{\rm pre}|\vec{\phi},\epsilon) = \frac{1}{\sqrt{\det(2\pi C)}} \exp\left(-\frac{1}{2}(\delta t^{\rm pre} - M\epsilon)^T C^{-1}(\delta t^{\rm pre} - M\epsilon)\right). \tag{2.3.6}$$

Notice that the term in the exponent is the exact quantity that we would minimize in a generalized least squares analysis. Defining the following auxiliary variables, $g = M^T C^{-1}\delta t^{\rm pre}$ and $\Gamma = M^T C^{-1} M$, the log of the likelihood ratio becomes

$$\log\Lambda = \log p(\delta t^{\rm pre}|\vec{\phi},\epsilon) - \log p(\delta t^{\rm pre}|\vec{\phi},0) = \epsilon^T g - \epsilon^T \Gamma \epsilon. \tag{2.3.7}$$

Maximizing the log-likelihood ratio over the timing model parameter offsets, ϵ, is equivalent to minimizing the chi-squared; thus we obtain

$$0 = \frac{\partial \log\Lambda}{\partial \epsilon} = g - \Gamma\epsilon \tag{2.3.8}$$

Solving for the maximum likelihood values, $\hat{\epsilon}$, we obtain

$$\hat{\epsilon} = \Gamma^{-1} g = \left(M^T C^{-1} M\right)^{-1} M^T C^{-1} \delta t^{\rm pre} \tag{2.3.9}$$

$$\sigma_{\hat{\epsilon}} = \sqrt{\rm diag(\Gamma^{-1})} = \sqrt{\rm diag[(M^T C^{-1} M)^{-1}]}, \tag{2.3.10}$$

with $\sigma_{\hat{\epsilon}}$ are the 1-sigma uncertainties on the maximum likelihood estimates $\hat{\epsilon}$. More important for our purposes is the "transformation" from pre-fit to post-fit residuals. We can form the post-fit residuals as follows

$$\delta t^{\rm post} = \delta t^{\rm pre} - M\hat{\epsilon} = (I - M(M^T C^{-1} M)^{-1} M^T C^{-1})\delta t^{\rm pre} = R\delta t^{\rm pre}, \tag{2.3.11}$$

where I is the appropriately sized identity matrix and we have defined the oblique projection operator

$$R = I - M\left(M^T C^{-1} M\right)^{-1} M^T C^{-1}. \tag{2.3.12}$$

Thus, we see that the entire linear fitting process is completely encapsulated in the projection matrix R, with the caveat that in practice the full covariance matrix is replaced by a diagonal matrix only containing the TOA uncertainties leading to sub-optimal results if the true noise in the residuals is correlated or has additional white noise components (see e.g., Cordes & Shannon 2010). This effectively summarizes the *standard* pulsar timing procedure. However, if one wants to incorporate more complicated noise models or GWs into the mix then we must dig a bit deeper.

2.4 Beyond Standard Pulsar Timing Analysis

During the course of this dissertation we have adopted two similar but slightly different techniques for incorporating more complicated effects into pulsar timing data analysis. We now derive both methods.

2.4.1 Linear Transformation Approach

The first approach makes use of the so-called R-matrix as defined in Eq. (2.3.12) to perform a "coordinate transformation" from pre-fit to post-fit residuals in the likelihood function of Eq. (2.3.4) first explored in Demorest (2007) and later extended in Demorest et al. (2013) and Ellis et al. (2013). In essence, we wish to write the likelihood function in the *post-fit* basis. Let

$$p(\delta t^{\rm post}|\vec{\phi})d\delta t^{\rm post} = p(n|\vec{\phi})dn = p(n|\vec{\phi})\left|\frac{\partial n}{\partial \delta t^{\rm post}}\right|, \tag{2.4.1}$$

where $|\cdot|$ represents the determinant. We evaluate the Jacobian by *assuming* that R is invertible and writing $n = R^{-1}\delta t^{\rm post}$, since $\delta t^{\rm post} = Rn$. The Jacobian is then

$$\left|\frac{\partial n}{\partial \delta t^{\rm post}}\right| = |R^{-1}| = \frac{1}{|R|} = \frac{1}{\sqrt{RR^T}}. \tag{2.4.2}$$

Substituting this into Eq. (2.3.4) we obtain

$$p(\delta t^{\rm post}|\vec{\phi}) = \frac{1}{\sqrt{\det(2\pi RCR^T)}}\exp\left(-\frac{1}{2}(\delta t^{\rm post})^T(R^{-1})^T C^{-1}(R^{-1})(\delta t^{\rm post})\right). \tag{2.4.3}$$

The product RCR^T is just the covariance matrix of the residuals

$$\Sigma = \langle \delta t^{\rm post}(\delta t^{\rm post})^T\rangle = R\langle nn^T\rangle R^T = RCR^T. \tag{2.4.4}$$

Henceforth, we will drop the "pre" and "post" prefix to the residuals and simply refer to "residuals" (i.e., in practice the "residuals" are those obtained from a converged TEMPO2 fit) unless stated otherwise. We can then write the likelihood function for the residual data as

$$p(\delta t|\vec{\phi}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}}\exp\left(-\frac{1}{2}\delta t^T\Sigma^{-1}\delta t\right). \tag{2.4.5}$$

This likelihood function is then a multivariate Gaussian distribution for the residuals with covariance matrix RCR^T. The inverse of Σ does not formally exist since we have removed

degrees of freedom by fitting out the timing model. In practice, we can make use of a singular value decomposition to compute the determinant and pseudoinverse to evaluate the likelihood. Viewed in this way, the likelihood function for the residuals is simply a change of coordinates where R is a linear (but not invertible) map from $n \to \delta t = Rn$. Lastly, it is *very* important to note that R in this case only uses a noise covariance matrix composed of the error bars on the TOAs, where the covariance matrix in Σ is parameterized by $\vec{\phi}$.

2.4.2 Marginalization Approach

The second, and more self consistent, approach treats the fitting and signal characterization phase simultaneously by analytically marginalizing over the pulsar timing parameters, ϵ. In order to perform the marginalization, we first re-write the likelihood function of Eq. (2.3.6) in terms of the maximum likelihood parameter estimates, $\hat{\epsilon}$

$$\begin{aligned} p(\delta t|\vec{\phi},\epsilon) = & \frac{\exp\left(-\frac{1}{2}[\delta t^T C^{-1}\delta t - \hat{\epsilon}^T M^T C^{-1} M\hat{\epsilon}]\right)}{\sqrt{\det(2\pi C)}} \\ & \times \exp\left(-\frac{1}{2}(\epsilon-\hat{\epsilon})^T M^T C^{-1} M(\epsilon-\hat{\epsilon})\right). \end{aligned} \tag{2.4.6}$$

Since the only dependence on ϵ is in the last expression, adopting uniform priors on ϵ, we carry out the Gaussian integral analytically to obtain

$$I = \int_{-\infty}^{\infty} d\epsilon \, \exp\left(-\frac{1}{2}(\epsilon-\hat{\epsilon})^T M^T C^{-1} M(\epsilon-\hat{\epsilon})\right) = \sqrt{\frac{(2\pi)^m}{\det(M^T C^{-1} M)}}, \tag{2.4.7}$$

where m is the number of parameters in the timing model. Using this result and the definition of $\hat{\epsilon}$ in Eq. (2.3.9) the likelihood function for the residuals marginalized over the pulsar timing parameters is

$$p(\delta t|\vec{\phi}) = \frac{\exp\left(-\frac{1}{2}[\delta t^T C^{-1}\delta t - \delta t^T C^{-1} M(M^T C^{-1} M)^{-1} M^T C^{-1}\delta t]\right)}{\sqrt{(2\pi)^{N_{\rm TOA}-m}\det(C)\det(M^T C^{-1} M)}}. \tag{2.4.8}$$

In van Haasteren & Levin (2013), it was discovered that this marginalized likelihood can be written in a more visually pleasing and computationally appealing form. We begin by decomposing the design matrix via a singular value decomposition (SVD)

$$M = UDV^T, \tag{2.4.9}$$

where U and V are $N_{\rm TOA} \times N_{\rm TOA}$ and $m \times m$ orthogonal matrices, respectively and D is an $N_{\rm TOA} \times m$ diagonal matrix with $m < N_{\rm TOA}$. The first m columns of U span the column space of M and the last $N_{\rm TOA} - m$ columns span the complement of M. Let G_c be an $N_{\rm TOA} \times m$ matrix and G be an $N_{\rm TOA} \times (N_{\rm TOA} - m)$ matrix that form the total orthogonal matrix $U = [G_c\, G]$ More explicitly

$$M = \begin{bmatrix} Gc & G \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} V^T, \tag{2.4.10}$$

where S is an $m \times m$ diagonal matrix composed of the eigenvalues of M. These matrices have the following important properties

$$G_c^T G_c = I_m \tag{2.4.11}$$

$$G^T G = = I_{N_{\rm TOA}-m} \tag{2.4.12}$$

$$G_c G_c^T + G G^T = I_{N_{\rm TOA}}. \tag{2.4.13}$$

Finally, it is possible to show that the marginalized likelihood function can be written as

$$p(\delta t|\vec{\phi}) = \frac{\exp\left(-\frac{1}{2}\delta t^T G (G^T C G)^{-1} G^T \delta t\right)}{\sqrt{(2\pi)^{N_{\rm TOA}-m} \det(G^T C G)}}. \tag{2.4.14}$$

The marginalized likelihood written in this way has a similar interpretation to Eq. (2.4.5) as a multivariate Gaussian distribution for data $G^T \delta t$ and covariance matrix $G^T C G$. The matrix G^T is a projection operator that projects our data δt onto the null space of M, that is, it projects the data into a subspace orthogonal to the linearized timing model. In the timing analysis used here, dispersion measure (DM) variation and profile frequency evolution effects are part of the timing model, and these terms are included when constructing the G matrix. In this way we have fully taken into account the timing model fitting procedure.

2.4.3 Adding Other Signals to the Likelihood Function

In the above, we have assumed that the TOAs only contain a deterministic timing model and Gaussian noise, however; it is trivial to extend the likelihood functions to include any number of other deterministic (linear or non-linear) sources. For example, one may want to include continuous GW sources (Babak & Sesana 2012; Ellis et al. 2012c; Petiteau

et al. 2013; Ellis 2013) or bursts (Finn & Lommen 2010; van Haasteren & Levin 2010) or full non-linear timing model (Lentati et al. 2013a; Vigeland & Vallisneri 2013) terms (As opposed to the linear approximation $M\epsilon$). To include these extra sources we simply re-write the residuals as

$$\delta t = M\epsilon + \sum_{i=1}^{N_{\rm signals}} s_i(\vec{\lambda}_i) + n, \tag{2.4.15}$$

where $s_i(\vec{\lambda}_i)$ is the functional form of a given signal described by parameters $\vec{\lambda}_i$. One can then use this expansion of δt in Eq. (2.3.6). The resulting likelihood functions (both in the R-matrix and marginalization approaches) will be identical to Eqs. (2.4.5) and (2.4.14) with

$$\delta t \to \delta t - \sum_{i=1}^{N_{\rm signals}} s_i(\vec{\lambda}_i). \tag{2.4.16}$$

Throughout the remainder of this dissertation either the R-matrix or G-matrix likelihood functions will be used either with or without additional deterministic signals depending on the application. The reason for the use of two different likelihood functions is an artifact of timing; earlier work makes use of the R-matrix, while newer work makes use of the G-matrix approach.

2.5 Parameterized Noise Models

In section 6.2 we have derived the likelihood function used for our analysis; however, we have not specified the form of the noise covariance matrix C, or more specifically, the form of the noise in the residuals n. Here we will detail currently used noise models, explaining each component along the way. Next we will compute the covariance matrix and insert it into our likelihood function derived in Section 6.2, making use of some techniques to increase computational efficiency.

In practice, the noise in pulsar timing residuals is non-Gaussian due to interstellar medium scintillation effects which are manifest through a time varying pulse intensity, resulting in time-dependent TOA uncertainties. Nonetheless, the noise in each residual is modeled very well by a Gaussian distribution with zero mean and standard deviation equal to the uncertainty on the TOA. In other words, the noise in the *weighted* (by the individual TOA errors) residuals is very well approximated as a Gaussian distribution.

As will be detailed in the next section, we include these error bar weights in our noise covariance matrix.

2.5.1 White Noise Model

Naively one may think that modeling white noise in our pulsar timing residuals is quite simple and can be accomplished simply by assuming that the only white noise present is due to radiometer noise and is captured in the TOA uncertainty output by the template matching procedure described above. In fact, some pulsars do seem to follow this trend with no evidence of additional white noise beyond the standard template fitting errors; however, in many pulsars, especially very well timed pulsars and those with long (i.e., $T >$ 10 yr) baselines we see strong evidence for additional white noise beyond radiometer noise. Even radiometer noise may be affected by some sort of systematic uncertainty causing the error bars to either be under or over estimated, thus we include a free parameter that is a multiplier on the given TOA uncertainties. This kind of effect will likely be dependent on the given observing backend (i.e., the instrumentation used for timing) and possibly on the observing frequency, therefore, the covariance matrix of the radiometer noise, $n_{\rm rad}$, is given by

$$C_{\rm rad} = EW = \begin{bmatrix} E_1 & 0 & \cdots & 0 \\ 0 & E_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{N_{\rm back}} \end{bmatrix} \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{N_{\rm back}} \end{bmatrix}, \tag{2.5.1}$$

where $W = \mathrm{diag}\{\sigma_i^2\}$, with σ_i the uncertainty on the ith TOA, and E is also an $N_{\rm TOA} \times N_{\rm TOA}$ diagonal matrix composed of $N_{\rm back}$ sub diagonal matrices, where $N_{\rm back}$ is the number of backend/frequency combinations for a given pulsar. Each sub-matrix of E is simply the appropriately sized identity matrix multiplied by the square of the unknown error scaling factor (EFAC). For example, if we observe a pulsar with backends A and B, each at both 800 MHz and 1.4 GHz, then in general we will have four EFAC parameters, where all uncertainties for TOAs associated with backend A at 800 MHz will now become $\sigma_i \rightarrow e_1\sigma_i$, where e_1 is the EFAC parameter for that backend/frequency combination. Similarly, TOAs with the other three backend/frequency combinations will

have error bars scaled by three different EFAC parameters. Generally, measured values of EFAC are very nearly unity indicating that there is no systematic error in the TOA template matching uncertainty; however, it is desirable to use several EFAC parameters as described above in characterizing the noise just to be sure that one particular backend/frequency observing setup is not biasing the noise estimates of the others. In principle it would be possible to carry out a Bayesian model selection to determine if the data supports separate EFAC parameters of simply only 1, however; such a study is beyond the scope of this dissertation and will be left to future work.

The EFAC parameters above should capture any biases in the TOA uncertainties but it is not likely to capture an additional white Gaussian noise process that affects all TOAs in the same way. For example if the timing residuals show a large spread that is not consistent with the TOA error bars, then there is likely an additional uncertainty that we can model with an additional unknown parameter typically referred to as EQUAD because it is added in *quadrature* to the EFAC component. The covariance matrix for the EQUAD component is then

$$C_{\rm equad} = \mathcal{Q}I, \tag{2.5.2}$$

where, as above, $\mathcal{Q}$ is a block diagonal matrix composed of the the squares of the EQUAD parameters, and I is an $N_{\rm TOA} \times N_{\rm TOA}$ identity matrix. Note that again we separate out the EQUAD parameters based on their backend/frequency combination, so in the above example the *effective* uncertainty on a set of TOAs from the first backend/frequency setup becomes $\sigma_i \rightarrow \sqrt{e_1^2\sigma_i^2 + q_1^2}$, where q_1 is the EQUAD parameter.

As we noted in the Introduction when discussing averaged pulse profiles, individual pulses *jitter* within the pulse window and can lead to further uncertainty in the TOA measurement that is not fully captured by the template fitting error. For many observing scenarios, the jitter noise will simply be absorbed into the EQUAD parameter discussed above, however; in current NANOGrav datasets, the large observing bandwidth has led to the construction of several TOAs per observing epoch in order to deal with the frequency dependence of the profile shape (see e.g., Demorest et al. 2013, for more details). In this case, the jitter noise of all TOAs within a given epoch are correlated and this must be accounted for in our noise model. One may be tempted to simply perform a weighted

average of the TOAs and work with the new reduced datasets but in the Bayesian scheme we must marginalize over the timing model parameters analytically and it is unclear how to carry out this process for epoch-averaged TOAs. Because of this, we have developed a framework to essentially work backward from the marginal likelihood to derive a nearly exact averaging scheme. First we re-write our noise covariance matrix for the "jitter" parameter

$$C_J = U\tilde{J}U^T, \tag{2.5.3}$$

where $\tilde{J}$ is a $q \times q$ reduced covariance matrix with q the number of epochs[1] in our dataset and U is the "exploder" matrix that maps epochs (columns) to the full set of TOAs (rows). Again, $\tilde{J}$ has a similar form to both E and $\mathcal{Q}$ above except that each submatrix describes the variance of the *averaged* residuals as opposed to the residuals themselves and is again diagonal with constant multiplier j_i^2 where j_i is the jitter parameter for a given backend/frequency combination. In other words, this matrix is identical to the EQUAD covariance matrix for the *averaged* residuals. The U-matrix essentially turns this reduced covariance matrix into the full covariance matrix by correlating (i.e, adding non-zero off-diagonal elements to the covariance matrix) all TOAs within a given observing epoch. Finally, we note that this term will indeed capture true pulse phase jitter as described in Cordes & Shannon (2010) but it could also capture other effects such as polarization calibration, short timescale ISM effects, or red noise processes with shallow spectral indices.

2.5.2 Red Noise Model

Red noise, also referred to as *timing noise* is a time-correlated noise process with ensemble average power spectra that has greater power at lower fluctuation frequencies. In pulsar timing, the appearance of red noise could be due to a variety of effects including intrinsic pulsar instabilities which are manifest through random walks in the spin properties (Shannon & Cordes 2010), mode switching where the spin-down "switches" between two distinct values of $\dot{P}$ (Lyne et al. 2010), or time-varying dispersion measure (DM)

[1]Here we have defined an epoch to equal to the integration time as jitter will only cause correlations between TOAs created from the same folded profile.

variations[2] (Demorest et al. 2013; Keith et al. 2013). While a single realization of any of these red noise processes is likely to be non-stationary, the ensemble average behavior can be described as wide-sense stationary and obeying Gaussian statistics. If this is the case, then the ensemble average properties are completely contained in the power spectrum and in turn, the covariance matrix via the Wiener-Khinchin relation which states that the covariance matrix is the real part of the inverse Fourier transform of the power spectrum

$$[C_{\rm red}]_{ij} = \int_{-\infty}^{\infty} df \cos(2\pi f \tau_{ij}) P_{\rm red}(f), \tag{2.5.4}$$

where $\tau_{ij} = |t_i - t_j|$, where t are the TOAs, and $P_{\rm red}(f)$ is the power spectrum of the red noise process. If we assume that the power spectrum is a simple power law of the form

$$P_{\rm red}(f) = \mathcal{A} f^{-\gamma}, \tag{2.5.5}$$

with $\mathcal{A}$ the amplitude and γ the spectral index, then the covariance matrix can be found analytically

$$\begin{aligned} C_n^{\rm red} &= 2\int_{f_L}^{\infty} df\ \cos(2\pi f \tau_{ij}) \mathcal{A} f^{-\gamma} \\ &= 2\frac{\mathcal{A}}{f_L^{\gamma-1}}\Bigg[\Gamma(1-\gamma)\sin(\pi\gamma/2)(2\pi f_L \tau_{ij})^{\gamma-1} \\ &\quad - \sum_{k=0}^{\infty}(-1)^k \frac{(2\pi f_L \tau_{ij})^{2k}}{(2k)!\,(2k+1-\gamma)}\Bigg], \end{aligned} \tag{2.5.6}$$

where $\Gamma(x)$ is the gamma function and f_L is the low frequency cutoff. Physically, the low frequency cutoff is related to the Earth-pulsar light travel time but in practice, due to the timing model fit of the quadratic spin-down, we have no sensitivity to the cutoff as long as $f_L < 1/T$ where T is the length of the dataset and $\gamma \in [1, 7]$ (van Haasteren & Levin 2013). In computing the likelihood function we must invert the covariance matrix C which in general will be a dense matrix. With current datasets containing up to 20,000 TOAs, dense matrix inversions are computationally prohibitive. Furthermore, while an analytic expression can be found for a power-law spectrum, this is not the case for more complicated spectral models. For these reasons, we have chosen to adopt the formalism

[2] In general, DM variations are chromatic in that they effect different observing frequencies differently, however; if residuals are single-frequency measurements then DM variations and achromatic red noise processes become indistinguishable.

of Lentati et al. (2013b) who showed that is possible to expand the red noise realization in a Fourier series

$$n_{\rm red} = \sum_{j=1}^{N_{\rm mode}^{\rm red}} \left[a_j \sin\left(\frac{2\pi j t}{T}\right) + b_j \cos\left(\frac{2\pi j t}{T}\right) \right] = F_{\rm red}\mathbf{a}_{\rm red}, \tag{2.5.7}$$

where $\mathbf{a}_{\rm red}$ is a vector of the alternating sine and cosine amplitudes, T is the total time span of the data, and $F_{\rm red}$ is a $N_{\rm TOA} \times 2N_{\rm mode}^{\rm red}$ matrix with alternating sine and cosine terms with $N_{\rm mode}$ the number of frequencies used. Assuming that the underlying ensemble average red noise process is wide-sense stationary and can be completely described by a power-spectrum, then, by orthogonality, the Fourier coefficients $\mathbf{a}_{\rm red}$ will obey

$$\varphi_{{\rm red},ij} = \langle \mathbf{a}_{\rm red}\mathbf{a}_{\rm red}^T \rangle_{ij} = {\rm diag}(\{\varphi_{{\rm red},i}\}), \tag{2.5.8}$$

where the elements of $\varphi_{\rm red}$, denoted $\{\varphi_{{\rm red},i}\}$ are the coefficients of the theoretical power spectrum of the red noise process in the residuals. The above equation states that the Fourier modes are orthogonal, but this does not mean that they are assumed to be orthogonal in the time domain where they are sampled and this non-orthogonality is taken into account in the Fourier design matrices. In Bayesian terms, the above equation represents our prior knowledge of the power spectrum by stating that we do not know the form that the power spectrum will take but we do know that the underlying Fourier modes are orthogonal. In this framework, the red noise covariance matrix is then

$$C_{\rm red} = \langle n_{\rm red} n_{\rm red}^T \rangle = F_{\rm red} \langle \mathbf{a}_{\rm red}\mathbf{a}_{\rm red}^T \rangle F_{\rm red}^T = F_{\rm red}\varphi_{\rm red}F_{\rm red}^T. \tag{2.5.9}$$

Note that by using this formalism we can parameterize the power spectrum in any way that is desired or we can simply allow the power spectrum coefficients to be free parameters themselves, making this method extremely robust and powerful.

2.5.3 Time Varying DM Model

Time varying DM variations can be treated in a very similar manner as achromatic red noise but we can take advantage of multi-frequency observations to isolate the effects of DM. This observing frequency dependence scales as ν^{-2} where ν is the observing

frequency. We can now decompose the time varying DM signal into "Fourier" like components via

$$n_{\rm DM} = \sum_{j=1}^{N_{\rm mode}^{\rm DM}} \left[a_{DM,j} \sin\left(\frac{2\pi j t}{T}\right) D + b_{DM,j} \cos\left(\frac{2\pi j t}{T}\right) D \right] = F_{DM} \mathbf{a}_{DM}, \tag{2.5.10}$$

where D is a length $N_{\rm TOA}$ vector with components

$$D_i = \frac{1}{K \nu_i^2}, \tag{2.5.11}$$

with $K = 2.41 \times 10^{-16}$ Hz^2 cm^{-3} pc s^{-1}. Similar to the red noise case, $\mathbf{a}_{DM}$ is a vector of the alternating sine and cosine amplitudes for the DM signal and F is a $N_{\rm TOA} \times 2N_{\rm mode}^{\rm DM}$ matrix with alternating sine and cosine terms with a weighting given by D. So essentially we have decomposed the time-varying DM signal into a observing frequency dependent Fourier-like basis. As with the red noise, we can construct the covariance matrix for the DM variations as

$$C_{\rm DM} = \langle n_{\rm DM} n_{\rm DM}^T \rangle = F_{DM} \langle \mathbf{a}_{DM} \mathbf{a}_{DM}^T \rangle F_{DM}^T = F_{DM} \varphi_{DM} F_{DM}^T, \tag{2.5.12}$$

where again φ_{DM} are the coefficients of the theoretical power spectrum of the DM variations present in the residuals. Note that we also have the same freedom here to choose any mode for the DM power spectrum or allow the power spectrum coefficients to vary freely. One caveat to this way of parameterizing the DM variations is that we have no a-priori way of knowing what to choose as the lowest frequency in the Fourier-like expansion, in contrast to the red noise case where we are safe in choosing the lowest frequency to be $1/T$ since the quadratic spin-down subtraction in the timing model fit will absorb any frequencies lower than $1/T$. A natural way to ameliorate this problem is to include a quadratic fit in DM directly into the timing model. This model is then

$$Q_{\rm DM}(t_i) = \delta_0 D_i + \delta_1 t_i D_i + \delta_2 t_i^2 D_i, \tag{2.5.13}$$

where δ_0, δ_1, and δ_2 are free parameters in the model.

2.6 Modified Likelihood Function

We now will use the covariance matrices discussed above to derive a computationally efficient method of evaluating the likelihood function. Again, much of the following discussion will follow Lentati et al. (2013b). There are several ways of deriving the likelihood

function, here we will choose to explicitly write the red and DM Fourier components as a deterministic signal and choose a prior that makes use of the theoretical power spectrum. Thus, our timing residuals can be written as

$$\delta t = M\epsilon + n_{\rm white} + n_{\rm red} + n_{\rm DM}, \tag{2.6.1}$$

and the likelihood function is then

$$p(\delta t|\vec{\phi}, \mathbf{a}_{\rm red}, \mathbf{a}_{\rm DM}) = \frac{\exp\left(-\frac{1}{2}(\delta t - F_{\rm red}\mathbf{a}_{\rm red} - F_{\rm DM}\mathbf{a}_{\rm DM})^T \tilde{N}^{-1}(\delta t - F_{\rm red}\mathbf{a}_{\rm red} - F_{\rm DM}\mathbf{a}_{\rm DM})\right)}{\sqrt{(2\pi)^{N_{\rm TOA}-m}\det(G^T N G)}}, \tag{2.6.2}$$

where $N = C_{\rm rad} + C_{\rm equad}$ and $\tilde{N}^{-1} = G(G^T N G)^{-1} G^T$ and $\vec{\phi}$ is a vector of all EFAC and EQUAD parameters. Here we will not include the "jitter" term as that is a simple extension to this likelihood and will be discussed next. We can further simplify this expression by combining the red and DM terms. Defining $F = [F_{\rm red}\, F_{\rm DM}]$ and $\mathbf{a} = [\mathbf{a}_{\rm red}\, \mathbf{a}_{\rm DM}]$, where the square brackets denote concatenation. Now F is now a $N_{\rm TOA} \times (N^{\rm red}_{\rm mode} + N^{\rm DM}_{\rm mode})$ matrix and $\mathbf{a}$ is a length $N^{\rm red}_{\rm mode} + N^{\rm DM}_{\rm mode}$ vector. The parameters $\mathbf{a}$ can be though of as hyper-parameters, and assuming their ensemble average properties are described by the theoretical power spectrum coefficients

$$\varphi = \begin{bmatrix} \varphi_{\rm red} & 0 \\ 0 & \varphi_{\rm DM} \end{bmatrix}, \tag{2.6.3}$$

then the corresponding hyper-prior is

$$p(\mathbf{a}|\varphi) = \frac{\exp\left(-\frac{1}{2}\mathbf{a}^T\varphi^{-1}\mathbf{a}\right)}{\sqrt{\det\varphi}}. \tag{2.6.4}$$

The desired posterior is then

$$p(\vec{\phi}, \varphi, \mathbf{a}|\delta t) = p(\delta t|\vec{\phi}, \mathbf{a})p(\mathbf{a}|\varphi)p(\varphi)p(\vec{\phi}), \tag{2.6.5}$$

where $p(\varphi)$ and $p(\vec{\phi})$ are the prior probability distributions of the components of φ and $\vec{\phi}$, respectively. In some cases it may be interesting to fully map out this posterior; however, in this case we are more interested in the underlying power spectrum than in the Fourier coefficients themselves, thus we seek to numerically marginalize over the Fourier coefficients $\mathbf{a}$. Inspecting the likelihood in Eq. (2.6.2) we see that it is nearly

identical to the un-marginalized likelihood for the timing model parameters in Sec. 6.2 except that we have the additional Gaussian hyper-prior on $\mathbf{a}$ as opposed to the flat prior for ϵ. Nonetheless, the likelihood ratio in this case is

$$\ln \Lambda = d^T \mathbf{a} - \frac{1}{2} \mathbf{a}^T \Sigma \mathbf{a}, \tag{2.6.6}$$

where $d = F^T \tilde{N}^{-1} \delta t$ and $\Sigma = (F^T \tilde{N}^{-1} F + \varphi^{-1})$. This likelihood ratio is in exactly the same form as Eq. (2.4.14). Thus, the maximum likelihood estimator and uncertainty for the Fourier coefficients is

$$\hat{\mathbf{a}} = \Sigma^{-1} d^T \tag{2.6.7}$$

$$\sigma_{\hat{\mathbf{a}}} = \sqrt{\mathrm{diag}(\Sigma^{-1})}. \tag{2.6.8}$$

The marginalized likelihood function is then

$$\begin{aligned} p(\delta t | \vec{\phi}, \varphi) &= \int_{-\infty}^{\infty} d\mathbf{a}\, p(\delta t | \vec{\phi}, \mathbf{a}) p(\mathbf{a} | \varphi) \\ &= \frac{\exp\left[-\frac{1}{2}\left(\delta t^T \tilde{N}^{-1} \delta t - d^T \Sigma^{-1} d\right)\right]}{\sqrt{(2\pi)^{N_{\mathrm{TOA}} - m} \det(\varphi) \det(G^T N G) \det(\Sigma)}}. \end{aligned} \tag{2.6.9}$$

Thus, we have shown that by explicitly including the Fourier coefficients as part of a deterministic signal model and then analytically marginalizing over them, then we obtain this relatively simple form of the likelihood that is a function of the EFAC and EQUAD parameters, $\vec{\phi}$, and the power spectrum (both for red and DM noise) coefficients φ_i. It is possible, however to arrive at an identical expression by only including the ensemble average properties of the Fourier coefficients, that is, the total covariance matrix is then

$$C = N + F \varphi F^T. \tag{2.6.10}$$

We can then make use of the Woodbury lemma[3] to compute the inverse and determinant of C

$$G(G^T C G)^{-1} G^T = \tilde{N}^{-1} - \tilde{N}^{-1} F \left(F^T \tilde{N}^{-1} F + \varphi^{-1}\right)^{-1} F^T \tilde{N}^{-1} \tag{2.6.11}$$

$$\det(G^T C G) = \det(G^T N G) \det(\varphi) \det((F^T \tilde{N}^{-1} F + \varphi^{-1})). \tag{2.6.12}$$

Plugging this into Eq. (2.4.14) we obtain and identical expression to Eq. (2.6.9).

[3] $(A + DBE^T)^{-1} = A^{-1} - A^{-1} D (B^{-1} + E^T A^{-1} D)^{-1} E^T A^{-1}$ and $|A + DBE^T| = |A||B||B^{-1} + E^T A^{-1} D|$

In the above derivation we did not include the "jitter" term; however, we will now see that it is quite trivial to include this term. Again, we write the full covariance matrix

$$C = N + U\tilde{J}U^T + F\varphi F^T \simeq N + U(\tilde{J} + \tilde{F}\varphi\tilde{F}^T)U^T = N + U\tilde{C}U^T, \tag{2.6.13}$$

where $\tilde{F}$ is identical to the F-matrix except that the time argument is not the *averaged* TOAs. Thus the $\tilde{F}$-matrix is a $q \times N_{\rm mode}$ matrix. Essentially the above expression is assuming that the red noise process does not vary over the timescale of one epoch, which is typically a few hours. This approximation is very good since any red process that has timescales that short will be completely covariant with white noise due to our sparse sampling. The likelihood function takes the same form but with the following substitutions

$$d \rightarrow U^T\tilde{N}^{-1}\delta t \tag{2.6.14}$$

$$\Sigma \rightarrow \left(\tilde{C}^{-1} + U^T\tilde{N}^{-1}U\right) \tag{2.6.15}$$

$$\det(\varphi) \rightarrow \det(\tilde{C}), \tag{2.6.16}$$

and $\vec{\phi}$ now includes the jitter parameters as well.

2.7 Markov Chain Monte Carlo Methods

In this section we review the concept of MCMC. The appeal of MCMCs in general is that they sample directly from the posterior distribution and can efficiently explore the parameter space. The algorithm begins by specifying a point in some multidimensional parameter space $\vec{x}$. This point can be chosen at random from the prior or can be initialized in some other way if we have additional information about the posterior structure. From here, we propose a "jump" to a new point in parameter space, $\vec{y}$ via a jump proposal distribution function $q(\vec{y}|\vec{x})$. We then evaluate the posterior at this new point and accept the jump with probability $\alpha = \min(1, H)$ where H is the Hastings ratio

$$H_{\vec{x}\rightarrow\vec{y}} = \frac{p(\vec{y}|d)q(\vec{x}|\vec{y})}{p(\vec{x}|d)q(\vec{y}|\vec{x})}. \tag{2.7.1}$$

We repeat this process for many iterations until a convergence criteria is reached (i.e., autocorrelation length or Gelman Rubin R statistic (Gelman & Rubin 1992)) and the

marginalized posterior pdfs of the parameters are simply the histograms of the parameter values in the chain. The choice of proposal distribution will be very important to achieve rapid convergence and we will explore this problem in following sections.

2.7.1 Parallel Tempering and Evidence Evaluation

A major problem with generic MCMC samplers is the tendency to get trapped in a local maxima. For a standard search it is unlikely that we will know a priori where the global maxima are located in parameter space, thus we must start our chain from a random point in the prior space. We want our algorithm to then quickly locate the global maxima in the parameter space. To accomplish this in a way that satisfies detailed balance we make use of parallel tempering. This technique involves different chains exploring the parameter space simultaneously, each with a different target distribution

$$p(\Theta|d,\beta) = p(\Theta)p(d|\Theta)^{\beta}, \tag{2.7.2}$$

where $\beta \leq 1$ is the inverse "temperature" and again, Θ are our unknown model parameters. This will essentially flatten out the likelihood surface allowing the chains to more freely explore the entire prior volume. The "hot" chains will inform the "colder" chains and vice versa by proposing parameter swaps between different temperatures. A parameter swap between the ith and jth temperature is accepted with probability $\alpha = \min(1, H)$, where the multi-temperature Hastings ratio is

$$H_{i\to j} = \frac{p(d|\Theta_i,\beta_j)p(d|\Theta_j,\beta_i)}{p(d|\Theta_i,\beta_i)p(d|\Theta_j,\beta_j)}. \tag{2.7.3}$$

By swapping parameter states between different temperatures this ensures rapid location of the global maxima. While the swapping schedule varies in our analyses, depending on the number of unknown parameters, typically we perform swaps only between adjacent temperature chains every $\sim$1000 iterations. The true posterior samples will come from the $\beta = 1$ chain but the higher temperature chains can be used to evaluate the evidence via thermodynamic integration (see e.g. Littenberg & Cornish 2009, and references therein). Consider the evidence for a chain with temperature $1/\beta$ as part of a partition function

$$\begin{aligned} Z(\beta) &= \int d\Theta\, p(d|\Theta,\mathcal{H},\beta)p(\Theta|\mathcal{H}) \\ &= \int d\Theta\, p(d|\Theta,\mathcal{H})^{\beta}p(\Theta|\mathcal{H}). \end{aligned} \tag{2.7.4}$$

Now, consider the following derivative

$$\begin{aligned}\frac{\partial \log Z(\beta)}{\partial \beta} &= \frac{1}{Z(\beta)} \int d\Theta \, \frac{\partial p(d|\Theta,\mathcal{H})^\beta}{\partial \beta} p(\Theta|\mathcal{H}) \\ &= \int d\Theta \, \frac{1}{p(d|\Theta,\mathcal{H})^\beta} \frac{\partial p(d|\Theta,\mathcal{H})^\beta}{\partial \beta} \frac{p(d|\Theta,\mathcal{H})^\beta p(\Theta|\mathcal{H})}{Z(\beta)} \\ &= \int d\Theta \, \frac{\partial \log p(d|\Theta,\mathcal{H})^\beta}{\partial \beta} p(\Theta|d,\mathcal{H})^\beta \\ &= \int d\Theta \, \log p(d|\Theta,\mathcal{H}) p(\Theta|d,\mathcal{H})^\beta \\ &\equiv \langle \log p(d|\Theta,\mathcal{H}) \rangle_\beta,\end{aligned} \tag{2.7.5}$$

where we have used Bayes theorem in the third line and $\langle \log\, p(d|\Theta,\mathcal{H})\rangle_\beta$ is the expectation value of the likelihood for the chain with temperature $1/\beta$. Finally, integrating this expression we obtain the logarithm of the evidence for model $\mathcal{H}$

$$\log p(d|\mathcal{H}) = \int_0^1 d\beta \frac{\partial \log Z(\beta)}{\partial \beta} = \int_0^1 d\beta \langle \log\, p(d|\Theta,\mathcal{H})\rangle_\beta. \tag{2.7.6}$$

In practice, it is important to choose a temperature ladder such that we explore the entire likelihood surface and recover the full integrand of Eq. 2.7.6. Here we will closely follow Littenberg & Cornish (2010) in the construction of our temperature ladder and diagnostic techniques. In constructing a temperature ladder to be used with thermodynamic integration it is important to understand that there are two regimes that we are interested in (at least in the GW detection problem). The first regime is the range of temperatures in which the (tempered) likelihood is still in "contact" with the GW, that is, the data still inform on the GW parameters. Since this is where the bulk of the integrand is concentrated when a signal is present it is very important that we choose a fine temperature spacing here to resolve the point at which the likelihood loses contact with the GW. To do this we choose a geometrically spaced temperature ladder with temperature spacing

$$\Delta T = 1 + \sqrt{\frac{2}{n_{\rm dim}}}, \tag{2.7.7}$$

where $n_{\rm dim}$ is the number of dimensions in our search. Now that a temperature spacing is defined we must choose a maximum temperature $T_{\rm max}$ for this regime. This choice is based on the expected maximum SNR of a GW signal in the data. Since $\rho \propto \sqrt{\ln\, p(d|\Theta)}$, the effective SNR for a chain at temperature T is $\rho_{\rm eff} \propto \rho/\sqrt{T}$, therefore, for a chosen

maximum SNR we have

$$T_{\max} = \left(\frac{\rho_{\max}}{\rho_{\text{eff,max}}} \right)^2 , \tag{2.7.8}$$

where $\rho_{\text{eff,max}}$ is the SNR at which we lose contact with the GW signal. The values of $\rho_{\max}$ and $\rho_{\text{eff,max}}$ are very problem dependent and are usually chosen based on simulations and trial and error.

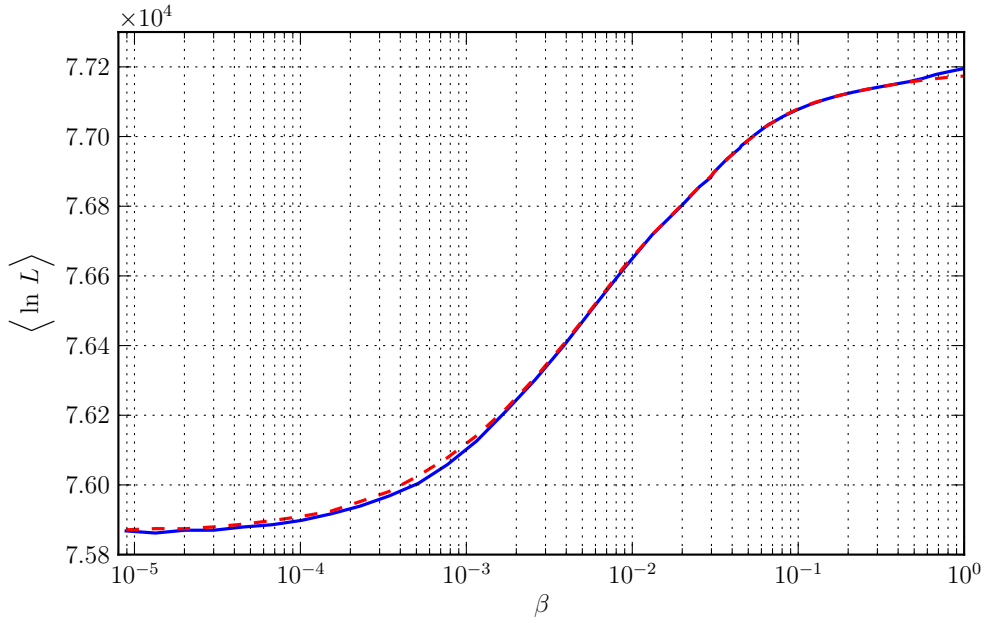


Figure 7 : Mean log-likelihood vs. β for GW plus noise (blue) and noise (green) models. Here we see that we have indeed explored a sufficient range of temperatures based on the fact that both curves become constant at small β.

If we were only interested in parameter estimation then we would cut off the temperature ladder here; however, for evidence evaluation we must explore the full parameter space. This is the second temperature regime of evidence evaluation via thermodynamic integration. Here, we must choose an overall maximum temperature such that we are effectively sampling from the prior. In other words, the temperature must be sufficiently high such that the average log-likelihood has become constant with respect to increasing temperature. In this regime we choose a more coarse temperature spacing with $\Delta T \sim 1.5$ and geometric spacing. As noted in Littenberg & Cornish (2010) a good diagnostic to ensure that we are using a high enough temperature is to plot the mean log-likelihood for each temperature chain vs. β.

For example, in Figure 7 we plot the mean log-likelihood vs. β for a continuous GW plus noise (blue) and noise (red) models. First we notice that at low temperatures (high β) the GW plus noise model fits the data better based on the higher likelihood values (the data has an SNR 10 GW injection) but that it has slightly lower values at high temperature (low β) because of the expanded prior volume due to the GW parameter space. Since the Bayesian evidence is the area under these curves the question that is being answered by computing a Bayes factor is "Does the fact that the GW plus noise model fits the data better (low temperature regime) overcome the fact that that model has a larger prior volume (high temperature regime)?". Because of this it is crucial that we include temperatures high enough so that the average log likelihood becomes constant, indicating that we are sampling the prior distribution.

2.7.2 Jump Proposals

In order to facilitate good mixing of the MCMC chains, especially in large parameter spaces, it is very important to have good jump proposals. In our implementation of the PTMCMC algorithm we use a jump proposal that is composed of a randomized cycle of sub-proposals. Here we will briefly outline the different jump proposals used in the cycle.

Correlated Jumps

An important and useful correlated jump proposal is built on the adaptive metropolis (AM) algorithm Haario et al. (2001)(hereafter HST01). This method makes use of an adaptive scheme where the *gaussian* proposal distribution is updated using the past history of the chain. By using the full past history of the chain this algorithm is indeed non-Markovian but it is shown in HST01 that it retains the correct ergodic properties and thus will give unbiased samples from the posterior probability distribution. The algorithm is actually quite simple. First we use a multidimensional proposal distribution with diagonal covariance matrix C_0, where the initial jump sizes are relatively unimportant but should be chosen small enough that we have a large initial acceptance rate and thus we will begin to build up points for later adaptation. After some number of iterations, η,

(we usually choose $\eta \sim 1000$) the covariance matrix at iteration n becomes

$$C_n = \begin{cases} C_0 & n \leq \eta \\ s_d \mathrm{Cov}(\Theta_0, \ldots, \Theta_{n-1}) & n > \eta, \end{cases} \tag{2.7.9}$$

where s_d is a parameter that depends on the dimension of the problem and $\mathrm{Cov}(\Theta_0, \ldots, \Theta_{n-1})$ is the sample covariance matrix at the nth iteration of the algorithm. HST01 suggest a value of $s_d = 2.4^2/n_{\mathrm{dim}}$, where n_{dim} is the dimension of the problem, however we have found that we need to use a smaller value to obtain optimal acceptance ratios around 25% (Gelman et al. 1996), however; we will occasionally make small jumps (scale by 0.01) or large jumps (scale by 10). As shown in the above equation, we do not perform the adaptation at every iteration of the chain but instead update the covariance matrix every η iterations, which helps shorten the runtime of the algorithm. This adaptive method will help speed convergence as the jump proposal will begin to mimic the posterior and take into account any parameter correlations. This jump is used in $\sim 20\%$ of our total jump cycle.

In large parameter spaces, as we encounter when modeling the GW and noise parameters simultaneously, the above method can result in very low acceptance and thus, slow convergence. Haario et al. (2005) introduce the Single Componentwise Adaptive Metropolis (SCAM) algorithm in which only one *uncorrelated* variable is updated in the jump proposal. If the variables are completely uncorrelated, then this method is identical to using the AM algorithm but only updating one parameter. However, if the parameters are correlated, we can define a set of uncorrelated parameters

$$y = U^T x, \tag{2.7.10}$$

where x is our original vector of parameters and U is defined by the eigenvalue decomposition $C_n = USU^T$, where U is a unitary matrix and S is diagonal. It is then easy to see that the covariance matrix of y, averaged over many steps in the chain is

$$\langle yy^T \rangle = U^T \langle xx^T \rangle U = U^T USU^T U = S. \tag{2.7.11}$$

Since $S = \mathrm{diag}\{\sigma_s^2\}$ is a diagonal matrix, each y represents an uncorrelated parameter. Therefore, we choose an uncorrelated parameter at random and propose the jump

$$y_{i+1}^j = y_i^j + 2.4\mathcal{N}(0, \sigma_s^j), \tag{2.7.12}$$

where $N(0, \sigma_s)$ is a zero mean Gaussian deviate with variance σ_s^2, i is the iteration number, and j is the parameter number. We can then relate the jump in the uncorrelated parameters back to a jump in the correlated parameters

$$x_{i+1} = U y_{i+1}. \tag{2.7.13}$$

If U is not the identity matrix (i.e., the parameters, x, are correlated) then this means that we will jump in combinations of correlated physical parameters even though we only jump in *one* uncorrelated component at a time. We have found that jumps of this kind greatly improve mixing when running with a large number of search parameters (e.g. >100). This jump is used in $\sim 40\%$ of our total jump cycle.

We also employ a third type of correlated jump proposal known as differential evolution (DE) (Braak 2006). Differential evolution is a simple genetic algorithm that also makes use of the previous history of samples in the chain. A differential evolution jump can be constructed as follows. First choose, at random, two previous iterations of the chain. Denote the parameter vector at those two new points as x_m and x_n. The DE jump is then

$$x_{i+1} = x_i + s_{DE}(x_m - x_n), \tag{2.7.14}$$

where s_{DE} is a scale factor which we choose to be $s_{DE} = 2.4^2/n_{\rm dim}$ and $s_{DE} = 1$, each with 50% probability. The first scale factor here is identical to that used in the AM jumps and the second is known as a "mode jump", that is, if x_m and x_n are located at two different modes of the posterior distribution, then the mode jump will result in a jump that stays on the same mode as x_i or jumps to the other mode. For this reason, DE jumps are usually employed if there are strong mulimodal structures in the posterior pdf. Also, since we are drawing points from the posterior, then these jumps will also "learn" about any correlations among parameters and will be taken into account in the jump proposal. This jump is used in $\sim 20\%$ of our total jump cycle.

Uncorrelated Jumps

Although we use mostly correlated jump proposals, about 15% of our jumps consist of uncorrelated jumps. In many cases, these uncorrelated jump proposals are simple draws

from the prior distribution. For prior draws, we have four different jump proposals. Since all pulsars will have a strong white noise component and likely a weak red noise component and likely no visible GW signal, we draw from the white noise, red noise and GW prior distributions separately with different weights. Red noise and GW (including the pulsar distance for continuous wave searches) prior draws account for $\sim 12\%$ of our total jump cycle. Finally, we also occasionally make white noise and full parameter space prior draws, which account for $\sim 3\%$ of our total jump cycle. Although this is quite a large percentage of jumps that draw from the prior it greatly improves mixing in our case when we have many search parameters with broad posterior distributions.

Chapter 3

Frequentist Detection Statistics for Continuous GWs

> "*Seize the time... Live now! Make now always the most precious time. Now will never come again.*"
>
> — Captain Picard, *Star Trek The Next Generation: The Inner Light*

This chapter is based on:
Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays
J. A. Ellis, X. Siemens, J. D. E. Creighton
ApJ (2012), 765, 175

3.1 Introduction

In the next few years pulsar timing arrays (PTAs) are expected to detect gravitational waves (GWs) in the frequency range 10^{-9} Hz–10^{-7} Hz. Prior to the establishment of PTAs, Jenet et al. (2004) used existing pulsar data to rule out the proposed SMBHB system 3C66B, a possible source of continuous GWs (the estimated mass of the proposed system has since been been lowered significantly (Iguchi et al. 2010) so that it is not likely to be detectable with current PTAs). In this work, the authors looked for the signature of a continuous GW in real pulsar data through the use of Lomb-Scargle periodograms and suggested a method for directed searches of known sources. Yardley et al. (2010) also relied on the Lomb-Scargle periodogram to determine the sensitivity of the PPTA to continuous GW sources as a function of GW frequency. van Haasteren & Levin

(2010) developed a bayesian framework aimed at the detection of GW memory in PTAs; however, the authors mention that the methods presented could be used for continuous GW sources as well. Sesana & Vecchio (2010) use an Earth-term only signal model to perform a study of SMBHB parameters measurable with PTAs using a Fisher matrix approach. Corbin & Cornish (2010) have developed a Bayesian Markov Chain Monte-Carlo (MCMC) data analysis algorithm for parameter estimation of a SMBHB system in which the pulsar term is taken into account in the detection scheme, thereby increasing the SNR and improving the accuracy of the GW source location on the sky. Recently, Lee et al. (2011) have developed parameter estimation techniques based on vector Ziv-Zakai bounds incorporating the pulsar term and have placed limits on the minimum detectable amplitude of a continuous GW source. In this work, the authors also propose a method of combining timing parallax measurements with single-source GW detections to improve pulsar distance measurements.

In the context of LIGO searches for continuous gravitational waves from spinning neutron stars, Jaranowski et al. (1998) developed the so-called $\mathcal{F}$-statistic, the logarithm of the likelihood ratio maximized over some of the signal parameters. Cutler & Schutz (2005) later generalized the $\mathcal{F}$-statistic to multi-detector networks. Very recently, Babak & Sesana (2012) have used the $\mathcal{F}$-statistic to show that in PTA data multiple SMBHB sources can be resolved in the sky. In this paper we build on this work, and improve on a number of aspects of prior continuous wave search methods developed for PTA data analysis.

In Section 3.2 we review the signal model. In Section 3.3 we discuss the $\mathcal{F}$-statistic in the context of PTA data. Unlike LIGO implementations of the $\mathcal{F}$-statistic, our algorithm is implemented fully in the time domain. This naturally deals with the irregular sampling of PTA data and avoids the spectral leakage problems that arise when frequency domain methods are used on such data. We also account for the timing model: fitting out pulsar parameters removes signal power at low frequencies, at frequencies near 1 yr^{-1} and 0.5 yr^{-1} due to sky location, proper motion, and parallax fitting, and for pulsars in binaries, at frequencies near the binary orbital frequency. Our approach also naturally incorporates colored noise sources, both uncorrelated and correlated (for the case when the

dominant noise source is a gravitational wave stochastic background). We also develop an incoherent detection statistic that maximizes over all pulsar dependent contributions to the likelihood. To test the effectiveness and sensitivity of our detection statistics, in Section 3.4 we perform a number of Monte-Carlo simulations. We produce sensitivity curves for PTAs of various configurations, and show that the performance of the incoherent statistic is comparable to the coherent $\mathcal{F}$-statistic. We also present an outline of the implementation of a continuous wave search pipeline. Finally, in Section 3.5 we summarize our results and conclude with a derivation of the likelihood maximized over the gravitational wave phases at the pulsar locations, which results in a vast reduction of the search parameter space. We leave the exploration of this new statistic for future work.

3.2 The Signal Model

As was discussed in Chapter 1 we write our GW induced pulsar timing residuals in the following form:

$$s(t,\hat{\Omega}) = F^{+}(\hat{\Omega})\Delta s_{+}(t) + F^{\times}(\hat{\Omega})\Delta s_{\times}(t), \tag{3.2.1}$$

where

$$\Delta s_A(t) = s_A(t_p) - s_A(t_e), \tag{3.2.2}$$

and t_e and t_p are the times at which the GW passes the Earth[1] and pulsar, respectively, and the index $A \in \{+,\times\}$ labels polarizations. The functions $F^A(\hat{\Omega})$ are the antenna pattern functions defined in Eq. (1.3.15). Also, recall that the GW contribution to the

[1]Technically, this is the time that the GW passes the SSB, however, following convention we will label this as the *Earth* time and will later refer to the *Earth*-term, keeping in mind that, in practice, all variables are referenced to the SSB.

timing residuals is

$$
\begin{aligned}
s_+(t) = \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \Big[& -\sin[2\Phi(t)](1+\cos^2\iota)\cos 2\psi \\
& - 2\cos[2\Phi(t)]\cos\iota \sin 2\psi \Big]
\end{aligned} \tag{3.2.3}
$$

$$
\begin{aligned}
s_\times(t) = \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \Big[& -\sin[2\Phi(t)](1+\cos^2\iota)\sin 2\psi \\
& + 2\cos[2\Phi(t)]\cos\iota \cos 2\psi \Big],
\end{aligned} \tag{3.2.4}
$$

where ψ is the GW polarization angle and ι is the inclination angle of the SMBHB. The orbital phase and frequency of the SMBHB are

$$
\Phi(t) = \Phi_0 + \frac{1}{32\mathcal{M}^{5/3}} \left(\omega_0^{-5/3} - \omega(t)^{-5/3} \right) \tag{3.2.5}
$$

and

$$
\omega(t) = \left(\omega_0^{-8/3} - \frac{256}{5} \mathcal{M}^{5/3} t \right)^{-3/8}. \tag{3.2.6}
$$

where Φ_0 and ω_0 are the initial values at the time of our first observation, the chirp mass is defined by $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1+m_2)^{1/5}$, where m_1 and m_2 are the masses of the two SMBHs, and d_L is the luminosity distance to the source. For reasons that will become clear later, we write the residuals for pulsar α in the following form

$$
\begin{aligned}
\delta t_\alpha(t, \hat{\Omega}) &= s_\alpha(t, \hat{\Omega}) + n_\alpha(t) = s^e_\alpha(t, \hat{\Omega}) + s^p_\alpha(t, \hat{\Omega}) + n_\alpha(t) \\
&= \sum_{i=1}^{4} \left[a_i(\zeta, \iota, \Phi_0, \psi) A^i_\alpha(t, \theta, \varphi, \omega_0) \right] \\
&+ s^p_\alpha(t, \zeta, \iota, \Phi_0, \psi, \theta, \varphi, \omega_0, L_\alpha) + n_\alpha(t),
\end{aligned} \tag{3.2.7}
$$

where $\zeta = \mathcal{M}^{5/3} d_L^{-1}$, $n_\alpha(t)$ is the noise for each pulsar and

$$
s^p_\alpha = F^+(\hat{\Omega}) s_+(t_p) + F^\times(\hat{\Omega}) s_\times(t_p). \tag{3.2.8}
$$

Hereon we will refer to the summation term as the Earth term and s^p_α as the pulsar term. We write the combination of chirp mass and luminosity distance to the binary as one parameter because the two can not be disentangled unless there is a measurement of $\dot{f}$, which we do not consider here. It is customary to label the parameters $(\zeta, \iota, \Phi_0, \psi)$ and $(\theta, \varphi, \omega_0)$ extrinsic and intrinsic parameters (Jaranowski et al. 1998), respectively. We

then define the amplitudes and time dependent basis functions

$$\begin{aligned}
a_1 &= \zeta\left[(1+\cos^2\iota)\cos 2\Phi_0\cos 2\psi + 2\cos\iota\sin 2\Phi_0\sin 2\psi\right] \\
a_2 &= -\zeta\left[(1+\cos^2\iota)\sin 2\Phi_0\cos 2\psi - 2\cos\iota\cos 2\Phi_0\sin 2\psi\right] \\
a_3 &= \zeta\left[(1+\cos^2\iota)\cos 2\Phi_0\sin 2\psi - 2\cos\iota\sin 2\Phi_0\cos 2\psi\right] \\
a_4 &= -\zeta\left[(1+\cos^2\iota)\sin 2\Phi_0\sin 2\psi + 2\cos\iota\cos 2\Phi_0\cos 2\psi\right]
\end{aligned} \tag{3.2.9}$$

and

$$\begin{aligned}
A^1_\alpha &= F^+_\alpha(\hat{\Omega})\omega(t)^{-1/3}\sin(2\Phi'(t)) \\
A^2_\alpha &= F^+_\alpha(\hat{\Omega})\omega(t)^{-1/3}\cos(2\Phi'(t)) \\
A^3_\alpha &= F^\times_\alpha(\hat{\Omega})\omega(t)^{-1/3}\sin(2\Phi'(t)) \\
A^4_\alpha &= F^\times_\alpha(\hat{\Omega})\omega(t)^{-1/3}\cos(2\Phi'(t)),
\end{aligned} \tag{3.2.10}$$

where $\Phi'(t) = \Phi(t) - \Phi_0$. Throughout this work we assume that the source is slowly evolving (i.e. the phase is independent of the chirp mass) and $\omega(t) \approx \omega_0$ and $\Phi'(t) \approx \omega_0 t$.

3.3 The Likelihood Function and the $\mathcal{F}$-statistic

Here we will introduce our formalism and derive the likelihood function and $\mathcal{F}$-statistic (the log-likelihood ratio maximized over extrinsic parameters) for PTAs. In this work, we will use the linear-transformation approach of Section 2.4.2 when writing the likelihood, extending it to multiple pulsars. We will also discuss the statistical properties of the $\mathcal{F}$-statistic in the presence and absence of a signal and show that we obtain the expected behavior for PTA data.

3.3.1 Likelihood

For a pulsar timing array with M pulsars we define the probability distribution function of the presumed Gaussian noise as multivariate Gaussian

$$p(\mathbf{n}) = \frac{1}{\sqrt{\det 2\pi\boldsymbol{\Sigma}_n}}\exp\left(-\frac{1}{2}\mathbf{n}^T\boldsymbol{\Sigma}_n^{-1}\mathbf{n}\right), \tag{3.3.1}$$

where

$$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix} \tag{3.3.2}$$

is the vector of the noise time-series, $n_\alpha(t)$, for all pulsars,

$$\mathbf{\Sigma}_n = \begin{bmatrix} C_1 & S_{12} & \dots & S_{1M} \\ S_{21} & C_2 & \dots & S_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1M} & S_{2M} & \dots & C_M \end{bmatrix} \tag{3.3.3}$$

is the multivariate covariance matrix, and

$$C_\alpha = \langle n_\alpha n_\alpha^T \rangle \tag{3.3.4}$$

$$S_{\alpha\beta} = \langle n_\alpha n_\beta^T \rangle \big|_{\alpha \neq \beta} \tag{3.3.5}$$

are the auto-covariance and cross-covariance matrices of the pulsar noise for pulsar α and pulsar pair (α,β), respectively. It is important to note that in the case of uncorrelated noise, the off-diagonal cross covariance matrices, $S_{\alpha\beta}$, vanish. In practice, we do not know the auto-covariance matrices a priori and we must estimate them from our residual data, δt_α. For this work, we note that in the small signal regime (which is almost certainly the case for our real PTA data sets) the auto-covariance matrix of the data is

$$\begin{aligned} \Sigma_{\delta t,\alpha} &= \langle r_\alpha r_\alpha^T \rangle = \langle s_\alpha s_\alpha^T \rangle + \langle n_\alpha n_\alpha^T \rangle \\ &\approx \langle n_\alpha n_\alpha^T \rangle = C_\alpha, \end{aligned} \tag{3.3.6}$$

since the amplitude of the signal is much smaller than the noise. Therefore, it is possible to make an estimate of the auto-covariance matrix of the noise, C_α from our observable data $\delta t_\alpha(t)$.

In order to time pulsars, a timing model is fit out of the pulsar TOAs via a weighted least squares fitting routine (Hobbs et al. 2006). This procedure can be expressed via a data-independent linear operator $\mathbf{R}$ (see Demorest et al. 2013 for details) so that

$$\tilde{\mathbf{r}} = \mathbf{R}\delta\mathbf{t}^{\rm pre}, \tag{3.3.7}$$

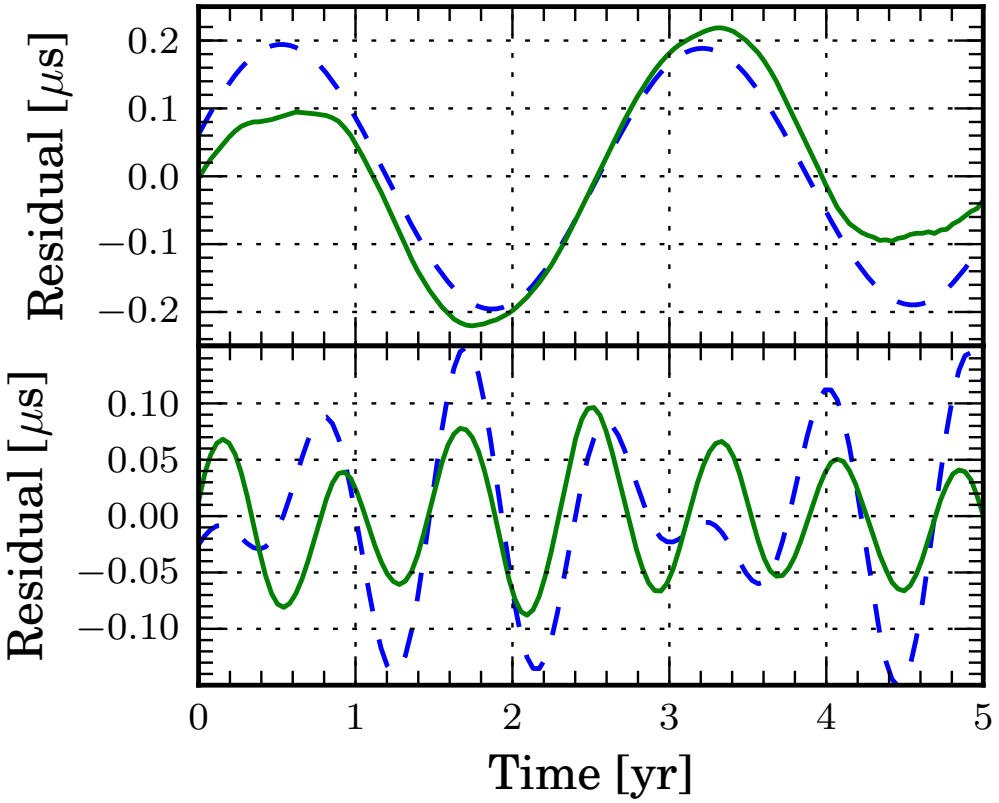


Figure 8 : SMBHB waveforms in two different regimes. Each plot shows the waveform before (dotted blue) and after fitting (solid green) for a full timing model including spin-down, astrometeric and binary parameters. Top Panel: The Earth and pulsar term modulations lie within the same frequency bin. Bottom Panel: The Earth term and pulsar term modulations are in different frequency bins.

where

$$\mathbf{R} = \begin{bmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_n \end{bmatrix} \tag{3.3.8}$$

is a matrix with sub-matrices R_α, which are $N_{\rm TOA} \times N_{\rm TOA}$ fitting operators for each pulsar and $\tilde{\mathbf{r}}$ are the post-fit residuals. From Eq. (3.3.7) we can see that $\tilde{\mathbf{n}} = \mathbf{R}\mathbf{n}$, where $\tilde{\mathbf{n}}$ is the post-fit noise. The effect of this fitting procedure on the waveforms can be seen in Figure 8 where the waveform is changed, quite significantly, from its pre-fit form. It is straightforward to show that the probability distribution function for $\tilde{\mathbf{n}}$ is

$$p(\tilde{\mathbf{n}}) = \frac{1}{\sqrt{\det 2\pi \mathbf{\Sigma}_{\tilde{n}}}} \exp\left(-\frac{1}{2}\tilde{\mathbf{n}}^T \mathbf{\Sigma}_{\tilde{n}}^{-1} \tilde{\mathbf{n}}\right), \tag{3.3.9}$$

where

$$\mathbf{\Sigma}_{\tilde{n}} = \langle \tilde{\mathbf{n}}\tilde{\mathbf{n}}^T \rangle = \mathbf{R}\langle \mathbf{n}\mathbf{n}^T \rangle \mathbf{R}^T. \tag{3.3.10}$$

The fitted residuals can therefore be written as

$$\tilde{\mathbf{r}} = \mathbf{R}\left(\mathbf{s} + \mathbf{n}\right) = \tilde{\mathbf{s}} + \tilde{\mathbf{n}}, \tag{3.3.11}$$

where $\tilde{\mathbf{s}} = \mathbf{R}\mathbf{s}$ and

$$\delta\mathbf{t}^{\rm pre} = \begin{bmatrix} \delta t_1 \\ \delta t_2 \\ \vdots \\ \delta t_M \end{bmatrix}, \qquad \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{bmatrix} \tag{3.3.12}$$

are the residual data and signal template for each pulsar, respectively. We can therefore write the likelihood of the data $\tilde{\mathbf{r}}$ given some signal template $\mathbf{s}$

$$\mathcal{L}(\mathbf{s}|\tilde{\mathbf{r}}) = p(\tilde{\mathbf{r}}|\mathbf{s}) = \frac{1}{\sqrt{\det 2\pi\mathbf{\Sigma}_{\tilde{n}}}} \exp\left(-\frac{1}{2}(\tilde{\mathbf{r}} - \tilde{\mathbf{s}})^T \mathbf{\Sigma}_{\tilde{n}}^{-1} (\tilde{\mathbf{r}} - \tilde{\mathbf{s}})\right). \tag{3.3.13}$$

We define the inner product for two time vectors $\mathbf{x}$ and $\mathbf{y}$ using the post-fit noise covariance matrix $\mathbf{\Sigma}_{\tilde{n}}$ as

$$(\mathbf{x}|\mathbf{y}) = \mathbf{x}^T \mathbf{\Sigma}_{\tilde{n}}^{-1} \mathbf{y}. \tag{3.3.14}$$

In this notation we can write the log of the likelihood ratio as

$$\ln\,\Lambda = \ln\,\frac{\mathcal{L}(\mathbf{s}|\tilde{\mathbf{r}})}{\mathcal{L}(0|\tilde{\mathbf{r}})} = (\tilde{\mathbf{r}}|\tilde{\mathbf{s}}) - \frac{1}{2}(\tilde{\mathbf{s}}|\tilde{\mathbf{s}}). \tag{3.3.15}$$

It is worth pointing out that finding the inverse of $\mathbf{\Sigma}_{\tilde{n}}$ is computationally intensive. Aside from it being a very large matrix, the fitting procedure results in loss of degrees of freedom in the data which makes $\mathbf{\Sigma}_{\tilde{n}}$ singular. Inverting this matrix therefore requires singular value decomposition. In most realistic scenarios we can assume that the off-diagonal cross-covariance matrices are small and expand the inverse of Eq. (6.2.5) in a Neumann series (see Equation 72 of Anholm et al. 2009 for details). In the simulations shown later in the paper we will assume that any correlated noise is much less than the uncorrelated part, thus we treat $\mathbf{\Sigma}_{\tilde{n}}$ as a block diagonal matrix of the auto-covariance matrices for each pulsar.

3.3.2 The Earth-term $\mathcal{F}$-statistic

We now analytically maximize over the extrinsic parameters $(\zeta, \iota, \Phi_0, \psi)$ in the signal model. A very similar calculation was first done by Jaranowski et al. (1998) in the context of LIGO, subsequently by Cornish & Porter (2007) in the context of LISA, and very recently by Babak & Sesana (2012) in the context of pulsar timing. For clarity, here we review this calculation in the notation introduced above. For this calculation we treat the pulsar term as a noise source and write our signal template in the form

$$\mathbf{s}(t, \hat{\Omega}) = \mathbf{s}^e(t, \hat{\Omega}) = \sum_{i=1}^{4} a_i(\zeta, \iota, \Phi_0, \psi)\mathbf{A}^i(t, \theta, \varphi, \omega_0), \tag{3.3.16}$$

where

$$\mathbf{A}^i = \begin{bmatrix} A_1^i \\ A_2^i \\ \vdots \\ A_M^i \end{bmatrix}. \tag{3.3.17}$$

Later we will explain the circumstances under which it is safe to drop the pulsar term. Using the Einstein summation convention we can now write the log-likelihood ratio as

$$\ln \Lambda = a_i N^i - \frac{1}{2} M^{ij} a_i a_j, \tag{3.3.18}$$

where $N^i = (\tilde{\mathbf{r}}|\mathbf{A}^i)$ and $M^{ij} = (\mathbf{A}^i|\mathbf{A}^j)$. Maximizing the log-likelihood ratio over the four amplitudes a_i gives

$$0 = \left.\frac{\partial \ln \Lambda}{\partial a_k}\right|_{\hat{a}} = N^k - M^{ik}\hat{a}_i, \tag{3.3.19}$$

yielding the maximum likelihood estimators for the four amplitudes

$$\hat{a}_i = M_{ij} N^j, \tag{3.3.20}$$

where $M_{ij} = (M^{ij})^{-1}$. Substituting these back into the log-likelihood results in the $\mathcal{F}_e$-statistic

$$2\mathcal{F}_e = N^i M_{ij} N^j. \tag{3.3.21}$$

The distribution of $2\mathcal{F}_e$ is a non-central χ^2 distribution with 4 degrees of freedom and a non-centrality parameter $\bar{\rho}^2$. It is straightforward to show that the expectation value is

$$\begin{aligned} \langle 2\mathcal{F}_e \rangle &= 4 + \bar{\rho}^2 \\ &= 4 + (\tilde{\mathbf{s}}^e|\tilde{\mathbf{s}}^e) + 2(\tilde{\mathbf{s}}^p|\tilde{\mathbf{s}}^e) + (\tilde{\mathbf{s}}^p|\mathbf{A}^i) M_{ij} (\tilde{\mathbf{s}}^p|\mathbf{A}^j), \end{aligned} \tag{3.3.22}$$

where $\mathbf{s}^p$ is the functional form of the pulsar term and the second two terms in $\bar{\rho}^2$ are due to the fact that we have only included the Earth term in our templates $\mathbf{s}$. In Figures 9(c) and 9(d) we can see that the probability distribution functions of $2\mathcal{F}_e$ follow the expected distributions in the absence and presence of a signal. While only the intrinsic parameters are formally searched over, it is also possible to get estimates of the maximized extrinsic parameters by constructing the following quantities (Cornish & Porter 2007):

$$\begin{aligned} A_+ = &\sqrt{(\hat{a}_1 + \hat{a}_4)^2 + (\hat{a}_2 - \hat{a}_3)^2} \\ &+ \sqrt{(\hat{a}_1 - \hat{a}_4)^2 + (\hat{a}_2 + \hat{a}_3)^2}, \end{aligned} \tag{3.3.23}$$

$$\begin{aligned} A_\times = &\sqrt{(\hat{a}_1 + \hat{a}_4)^2 + (\hat{a}_2 - \hat{a}_3)^2} \\ &- \sqrt{(\hat{a}_1 - \hat{a}_4)^2 + (\hat{a}_2 + \hat{a}_3)^2} \end{aligned} \tag{3.3.24}$$

and

$$A = A_+ + \sqrt{A_+^2 + A_\times^2}. \tag{3.3.25}$$

It is then possible to recover the maximized parameters

$$\iota = \cos^{-1}\left(\frac{-A_\times}{A}\right), \tag{3.3.26}$$

$$\psi = \frac{1}{2}\tan^{-1}\left(\frac{A_+\hat{a}_4 - A_\times\hat{a}_1}{A_\times\hat{a}_3 + A_+\hat{a}_2}\right), \tag{3.3.27}$$

$$\Phi_0 = -\frac{1}{2}\tan^{-1}\left(\frac{-(A_\times\hat{a}_1 - A_+\hat{a}_4)}{(A_+\hat{a}_3 + A_\times\hat{a}_2)}\right), \tag{3.3.28}$$

$$\zeta = \frac{|A|}{4}. \tag{3.3.29}$$

It is interesting to examine the case of one pulsar. In this case, Eq. (3.3.19) has no solution because the matrix M^{ij} is singular. The reason for this is that it is incorrect to write the residuals in the form of Eq. (3.2.7) with four degrees of freedom. For one pulsar, the signal has only *two* degrees of freedom: an amplitude and a phase, or equivalently, two unknown amplitudes, thereby making the maximization over *four* independent amplitudes an ill-posed problem. Thus, at least two pulsars are needed to solve Equation 3.3.19. It should be noted that it is straightforward to generalize this statistic to N GW sources, we will simply have $4N$ independent amplitudes instead of just 4 (see Babak & Sesana 2012 for more details). However, for simplicity in this work we will deal with just one GW source.

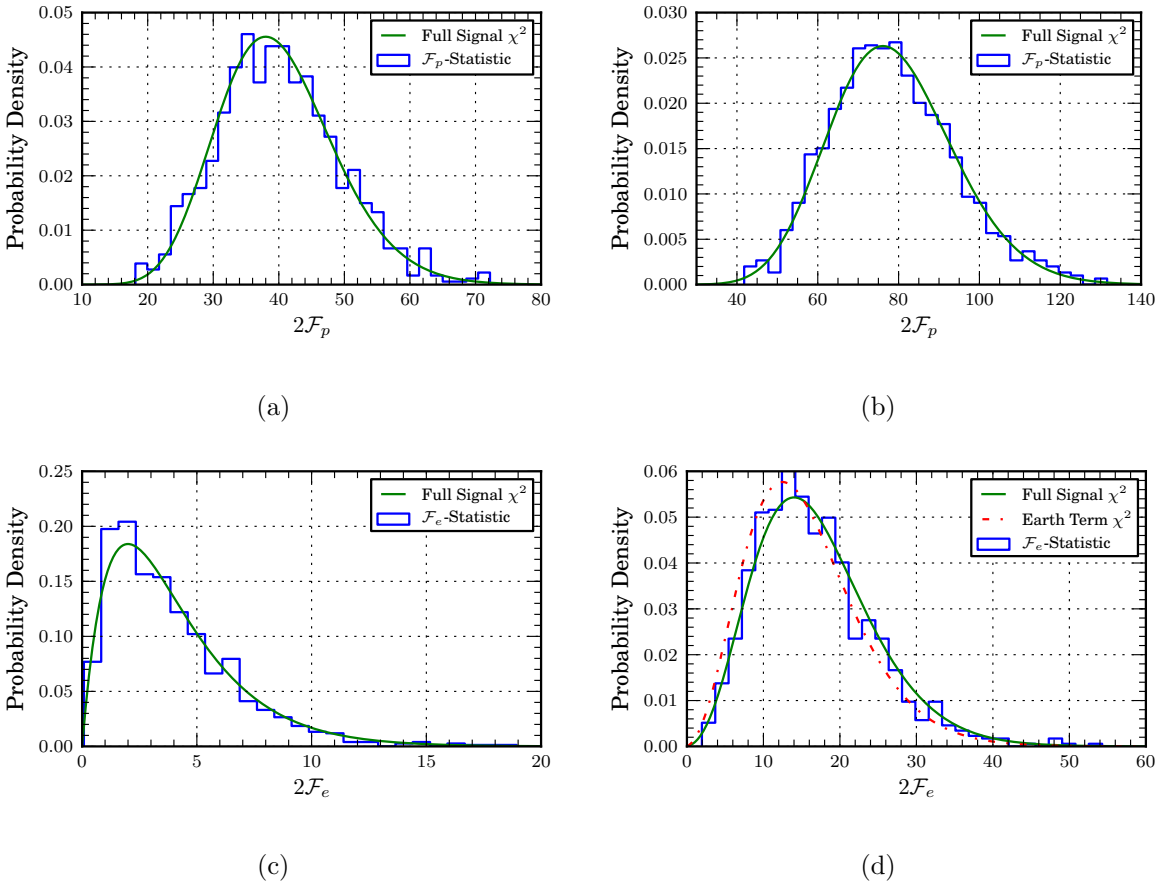


Figure 9 : Histograms and expected probability distribution functions of $2\mathcal{F}_p$ and $2\mathcal{F}_e$ in the absence and presence of a signal for 20 pulsars. Each simulation was done with the search parameters fixed and 1000 realizations of white Gaussian noise. (a): distribution of $2\mathcal{F}_p$ in the absence of a signal. (b): distribution of $2\mathcal{F}_p$ in the presence of a signal with non-centrality parameter ρ^2. (c): distribution of $2\mathcal{F}_e$ in the absence of a signal. (d): distribution of $2\mathcal{F}_e$ in the presence of a signal. The dashed (red) curve is a χ^2 distribution with a non-centrality parameter assuming that only the Earth term is present in the data. The solid (green) curve is a χ^2 distribution with non-centrality parameter $\bar{\rho}^2$ for $\mathcal{F}_e$-statistic, and ρ^2 for the $\mathcal{F}_p$-statistic.

Justification for dropping the pulsar term

There are two cases in which the pulsar term is truly a negligible contribution to the $\mathcal{F}_e$-statistic and can be dropped from the analysis with no change its statistical properties.

The first is the astrophysically likely (in terms of the resolvability of the source) scenario in which the evolution of the GW frequency is such that the Earth and pulsar terms are in different frequency bins (see e.g. Figure 2 of Sesana & Vecchio 2010). At the frequency of the Earth term the signal will build up coherently among the network of pulsars. The pulsar term signals, even if they all happen to be at the same frequency, will not because they have different phases that depend on the the pulsar distances. This

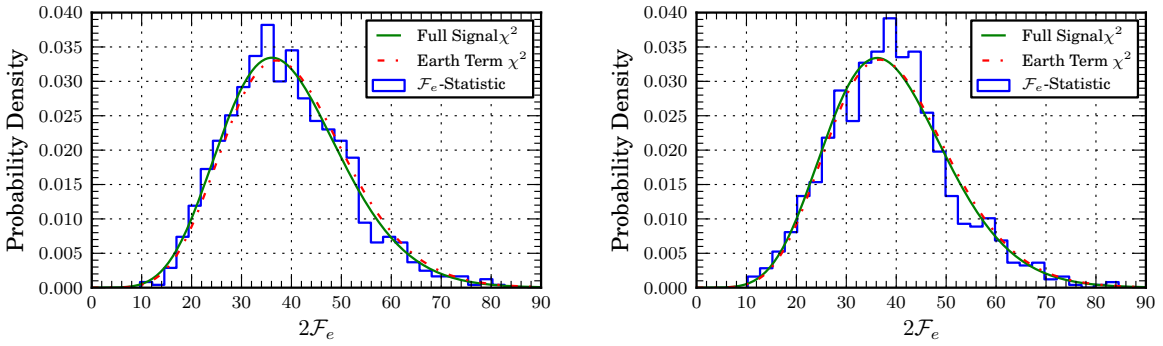


Figure 10 : Probability distribution functions for $2\mathcal{F}_e$ in the limits that the pulsar term is negligible. (a): probability distribution function in the limit that all pulsar terms lie outside the Earth term frequency bin. (b): probability distribution function in the limit of large M ($M = 50$ in this case) for overlapping Earth and pulsar term frequencies. The dashed (red) curve is a χ^2 distribution with a non-centrality parameter assuming that only the Earth term is present in the data. The solid (green) curve is a χ^2 distribution with non-centrality parameter $\bar{\rho}^2$ that takes both the Earth and pulsar term into account.

effect is illustrated in Figure 10(a) where the reference χ^2 distributions have 4 degrees of freedom and non-centrality parameters $(\tilde{\mathbf{s}}|\tilde{\mathbf{s}})$ (solid line) and $(\tilde{\mathbf{s}}^e|\tilde{\mathbf{s}}^e)$ (dashed line).

The second case is the scenario in which the Earth and pulsar term lie in the same frequency bin. Although a majority of the SMBHB sources are expected to be in this regime, they will likely have low mass and/or contribute at low frequency. Thus, these sources are less likely to be individually resolved. Nonetheless, in this case, there is still a phase difference between the Earth and pulsar terms. We expect that for a large number of pulsars the pulsar term signals will cancel because they all have different phases. We can see from Figure 9(d) that for a moderate number of pulsars ($M = 20$ in this case) the pulsar phases do not completely cancel and our measured values of the $\mathcal{F}_e$-statistic are higher than expected with just the earth term because the last two terms of Equation (3.3.22) do not sum to zero. However, in the case of large M ($M \gtrsim 50$) the pulsar term contributions sum approximately to zero, and again we have a χ^2 distribution with 4 degrees of freedom and non-centrality parameter $(\tilde{\mathbf{s}}^e|\tilde{\mathbf{s}}^e)$ (see Figure 10(b)).

If we happen to detect a signal that falls into the intermediate category mentioned above where $M < 50$ and some or all of the pulsar terms are in the same frequency bin as the Earth term, then this will create a bias in the recovered sky location but not in our ability to confidently detect the signal (see Figure 1 of Ellis et al. (2012b)).

This is because our detection criterion is based on the false alarm probability. As will be discussed in detail in Section 3.3.4, the false alarm probability only depends on the probability distribution function when the signal is *absent*, and we can see from Figure 9(a) that $2\mathcal{F}_e$ follows the expected distribution, because it is independent of the signal properties.

3.3.3 The incoherent $\mathcal{F}$-statistic

It is indeed possible to include the pulsar term in our analysis if we operate in the low frequency (or low chirp mass) regime where the frequency evolution of the source is slow enough that the frequency at the Earth and the pulsar are essentially the same so that the signal is a sum of two sinusoids of different phases: the pulsar term and the Earth term. To understand this more quantitatively, consider the Taylor series expansion of the orbital frequency of Eq. (5.A.2) evaluated at the pulsar time

$$\begin{aligned}\omega(t_p) &= \omega_0 \left(1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3} t_p\right)^{-3/8} \\ &\approx \omega_0 \left(1 + \frac{96}{5}\mathcal{M}^{5/3}\omega_0^{8/3}\left[t_e - L(1+\hat{\Omega}\cdot\hat{p})\right]\right).\end{aligned} \tag{3.3.30}$$

From this, we can see that $\omega(t_p) \approx \omega_0$ when

$$\omega_0 \ll \left(\frac{5}{96}\mathcal{M}^{-5/3}\left|T - L(1+\hat{\Omega}\cdot\hat{p})\right|^{-1}\right)^{3/8}, \tag{3.3.31}$$

where T is the total observation time. If we consider only one intrinsic parameter, ω_0, then the template for pulsar α is

$$s_\alpha(t,\hat{\Omega}) = \sum_{i=1}^{2} b_{i\alpha}(\zeta,\iota,\psi,\Phi_0,\phi_\alpha,\theta,\varphi)B_\alpha^i(t,\omega_0), \tag{3.3.32}$$

where

$$\phi_\alpha = \omega_0 L_\alpha(1+\hat{\Omega}\cdot\hat{p}_\alpha) \tag{3.3.33}$$

is the pulsar dependent phase. Making use of Eq. (3.2.9) we can now write the pulsar dependent amplitudes and basis functions as

$$\begin{aligned} b_{1\alpha} = \Bigg[& \left(F_\alpha^+ a_1 + F_\alpha^\times a_3\right)(1-\cos\phi_\alpha) \\ & - \left(F_\alpha^+ a_2 + F_\alpha^\times a_4\right)\sin\phi_\alpha \Bigg] \end{aligned} \tag{3.3.34}$$

$$\begin{aligned} b_{2\alpha} = \Bigg[& \left(F_\alpha^+ a_2 + F_\alpha^\times a_4\right)(1-\cos\phi_\alpha) \\ & + \left(F_\alpha^+ a_1 + F_\alpha^\times a_3\right)\sin\phi_\alpha \Bigg] \end{aligned} \tag{3.3.35}$$

and

$$B_\alpha^1(t) = \frac{1}{\omega_0^{1/3}}\sin(2\omega_0 t) \tag{3.3.36}$$

$$B_\alpha^2(t) = \frac{1}{\omega_0^{1/3}}\cos(2\omega_0 t), \tag{3.3.37}$$

where, again, ω_0 is the orbital angular frequency of the SMBHB. Again, using the Einstein summation convention, the log-likelihood ratio is

$$\ln\Lambda = \sum_{\alpha=1}^{M}\left[b_{i\alpha}P_\alpha{}^i - \frac{1}{2}Q_\alpha^{ij}b_{i\alpha}b_{j\alpha}\right], \tag{3.3.38}$$

where $P_\alpha = (\tilde{r}_\alpha|B_\alpha^i)$ and $Q_\alpha^{ij} = (B_\alpha^i|B_\alpha^j)$. Maximizing the log-likelihood ratio over the $2M$ amplitude parameters $b_{i\alpha}(\zeta,\iota,\psi,\phi_0,\phi_\alpha,\theta,\varphi)$ gives

$$0 = \left.\frac{\partial\ln\Lambda}{\partial b_{k\beta}}\right|_{\hat{b}} = P_\beta^k - Q_\beta^{ik}\hat{b}_{i\beta} \tag{3.3.39}$$

which yields the solution for the maximum likelihood estimators of the $2M$ amplitudes

$$\hat{b}_{i\beta} = Q_{ik}^\beta P_\beta^k. \tag{3.3.40}$$

Putting the amplitude estimators back into the likelihood ratio we obtain the $\mathcal{F}_p$-statistic

$$2\mathcal{F}_p = \sum_{\alpha=1}^{M} P_\alpha^i Q_{ij}^\alpha P_\alpha^j. \tag{3.3.41}$$

It is straightforward to then show that $2\mathcal{F}_p$ follows a χ^2 distribution with $2M$ degrees of freedom and non-centrality parameter ρ^2 and that

$$\begin{aligned} \langle 2\mathcal{F}_p\rangle &= 2M + \rho^2 \\ &= 2M + (\tilde{\mathbf{s}}|\tilde{\mathbf{s}}) \end{aligned} \tag{3.3.42}$$

where $\rho^2 = (\tilde{\mathbf{s}}|\tilde{\mathbf{s}})$ is the optimal signal-to-noise ratio (see Figure 9). Note that this is an *incoherent* detection statistic since it involves sum of the squares of the data, whereas the Earth-term $\mathcal{F}_e$-statistic is *coherent* since it involves the square of the sum of the data.

It is worth pointing out that for the case of white Gaussian noise, the $\mathcal{F}_p$-statistic is the time domain equivalent of the weighted power spectral summing technique studied in Ellis et al. (2012b). For colored Gaussian noise the statistic is the time domain equivalent to a weighted power spectral summing technique with frequency dependent weights. Another feature of this detection statistic is that it does not only apply to the low-frequency limit. If we work in the high frequency regime where the Earth and the pulsar terms are in different frequency bins, we can drop the pulsar term and arrive at the exact same maximized likelihood function. In this case the pulsar dependence of the amplitudes $b_{i\alpha}$ comes from the antenna pattern functions the not the pulsar phase. However, many of the justifications for dropping the pulsar term mentioned in the previous section do not apply in this case since the statistic is incoherent. We find that this detection statistic will often pick out the pulsar term frequency over the Earth term frequency because the residuals of Eq. (3.2.7) scale like $\omega(t)^{-1/3}$ and the pulsar term will always be at an equal or lower frequency than the Earth term frequency due to the geometry of the system. For the system of equations in Eq. (3.3.40) we have $2M$ equations and $6+M$ unknowns, so if we have 6 or more pulsars it is possible to solve for the all the parameters $(\zeta, \iota, \psi, \Phi_0, \theta, \varphi)$ along with the pulsar phases ϕ_α.

3.3.4 False alarm probability and detection statistics

Here we review the false alarm and detection probability distribution functions both when the intrinsic parameters are known and unknown. Our discussion follows closely that of Jaranowski et al. (1998) and Jaranowski & Królak (2005). In the case of known extrinsic parameters, we have shown in Sections 3.3.2 and 3.3.3 that when the signal is absent, the statistics $2\mathcal{F}_e$ and $2\mathcal{F}_p$ follow χ^2 distributions with 4 and $2M$ degrees of freedom, respectively. It was also shown that the aforementioned statistics follow a non-central χ^2 with non-centrality parameters $\bar{\rho}^2$ and ρ^2, respectively, when the signal is present.

Therefore, the probability distribution functions p_0 and p_1 when the intrinsic parameters are known and when the signal is absent and present, respectively, are

$$p_0(\mathcal{F}) = \frac{\mathcal{F}^{n/2-1}}{(n/2-1)!}\exp(-\mathcal{F}) \tag{3.3.43}$$

$$\begin{aligned} p_1(\mathcal{F},\kappa) = {} & \frac{(2\mathcal{F})^{(n/2-1)/2}}{\kappa^{n/2-1}} I_{n/2-1}\left(\kappa\sqrt{2\mathcal{F}}\right) \\ & \times \exp\left(-\mathcal{F}-\frac{1}{2}\kappa^2\right), \end{aligned} \tag{3.3.44}$$

where n is the number of degrees of freedom, $I_{n/2-1}$ is the modified Bessel function of the first kind of order $n/2-1$, and κ is ρ for $\mathcal{F}_p$ and $\bar{\rho}$ for $\mathcal{F}_e$. The false alarm probability P_F is defined as the probability that $\mathcal{F}$ exceeds a given threshold $\mathcal{F}_0$ when no signal is present. In this case, we have

$$P_F(\mathcal{F}_0) = \int_{\mathcal{F}_0}^{\infty} p_0(\mathcal{F})d\mathcal{F} = \exp(-\mathcal{F}_0)\sum_{k=0}^{n/2-1}\frac{\mathcal{F}_0^k}{k!}. \tag{3.3.45}$$

The probability of detection P_D is the probability that $\mathcal{F}$ exceeds the threshold $\mathcal{F}_0$ when the signal-to-noise ratio is κ:

$$P_D(\mathcal{F}_0,\kappa) = \int_{\mathcal{F}_0}^{\infty} p_1(\mathcal{F},\kappa)d\mathcal{F}, \tag{3.3.46}$$

however; we do not deal with the detection probability in this work. Our detection criterion is based on the false alarm probability.

We now turn to the more realistic problem of calculating the false alarm probability when the intrinsic parameters are not known. A detailed derivation and description is given in Jaranowski & Królak (2000), here we will simply review the result. The probability P_F^T that $\mathcal{F}$ exceeds $\mathcal{F}_0$ in one or more cells is given by

$$P_F^T(\mathcal{F}_0) = 1 - \left[1 - P_F(\mathcal{F}_0)\right]^{N_c}, \tag{3.3.47}$$

where N_c is the number of independent cells in parameter space. The number of independent cells can be calculated via geometrical methods described in Jaranowski & Królak (2000) and references therein.

Here we will make the following approximations. For our $\mathcal{F}_p$ statistic we will set N_c to be equal to the number of independent frequency bins defined by the Nyquist frequency. For our $\mathcal{F}_e$ statistic, we will set N_c to be equal to the number of templates

used in the search. In general the number of independent templates and the number of independent cells will be quite different. However, since we only have a three dimensional parameter space $(\omega_0, \theta, \varphi)$, and use a nested sampling algorithm to conduct the search (thereby reducing the number of templates in low likelihood regions of parameter space), setting the number of templates equal to the number of independent cells is a reasonable assumption.

3.4 Pipeline, sensitivities, and implementation

In this section we will test the $\mathcal{F}_e$ and $\mathcal{F}_p$ statistics on realistic simulated data sets. First, we will outline our detection pipeline, then we will briefly describe our simulated data sets and test the ability to confidently detect the signal and recover the injected intrinsic parameters. Finally, we perform Monte-Carlo simulations to produce sensitivity curves for PTAs of various configurations.

3.4.1 Detection Pipeline

The only inputs to our detection pipeline are the ephemeris file (typically called a "par" file) and TOA file (typically called a "tim" file) for each pulsar. The steps in the pipeline are as follows:

1. Use the standard pulsar timing package TEMPO2 (Hobbs et al. 2006) to form the residuals for each pulsar.

2. Use TEMPO2 plugin to output the design matrix for each pulsar (see Chapter 15 of Press et al. 1992 for more details). Then construct $\mathbf{R}$ from the design matrices following Demorest (2007).

3. Use a maximum likelihood method to make an estimate of $\mathbf{\Sigma}_{\tilde{n}}$. Note that the cross terms in Eq. (6.2.5) are expected to be small, so we will ignore them for this work.

4. Follow the methods described in Secs. 3.3.2 and 3.3.3 to construct the detection statistics defined in Equations (3.3.21) and (3.3.41) and search the relevant parameter space. If using the $\mathcal{F}_p$-statistic we simply grid up the frequency space for the

search. If using the $\mathcal{F}_e$-statistic we use the nested sampling package, MultiNest (Feroz & Hobson 2008; Feroz et al. 2009) to search the three dimensional intrinsic parameter space.

5. Output the maximum value of the detection statistic and number of templates used and compute the relevant false alarm probability using Eq. (3.3.47). Here we set our false alarm probability threshold to 10^{-4}. If the false alarm probability corresponding to our maximum value of $\mathcal{F}$ is less than 10^{-4} then we claim a detection.

6. Use the maximum likelihood estimators to find the extrinsic parameters (using Equations (3.3.26)–(3.3.29)), and construct the posterior probability distribution to find the intrinsic parameters by sampling the maximized likelihood (Equation (3.3.21)). As mentioned above, when using the $\mathcal{F}_p$ statistic, one could use numerical techniques to obtain estimates of the extrinsic parameters.

7. Use the maximum likelihood values of the intrinsic and extrinsic parameters to construct Gaussian prior distributions and carry out parameter estimation on the the full 7 dimensional search space, again using MultiNest, to get better estimates of SMBHB parameters.

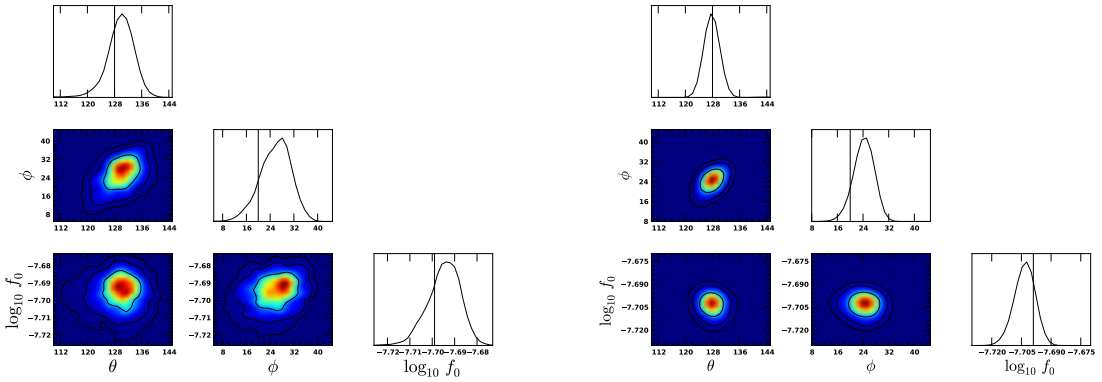
In this chapter we will only conduct steps 1–5 and leave steps 6 and 7 for future work. Although this work uses simulated datasets, nothing in this detection pipeline makes any assumptions about the spacing of the data, or the color of the noise.

In the absence of a detection we would like to set upper limits on the strain amplitude as a function of GW frequency. This can be accomplished as follows

1. Run the detection pipeline and determine the value of the $\mathcal{F}$-statistic.

2. For each frequency, choose the value of ζ corresponding to a specific strain amplitude. Then inject a SMBHB signal with randomly drawn binary orientation parameters $(\cos\iota, \psi, \Phi_0, \theta, \varphi)$.

3. Run the detection pipeline again on this injected data and measure the value of the $\mathcal{F}$-statistic.



(a) (b)

Figure 11 : Posterior probability distribution functions for sky location and orbital frequency for a network SNR=14 injection with and without red noise. Here we have used a PTA with 25 pulsars. The vertical lines indicated the injected parameters and the contours are the one, two and three sigma contours. (a): 100 ns white noise. (b): 100 ns white noise and uncorrelated red noise with amplitude $A = 4.22 \times 10^{-33}\,\mathrm{s}^{-1.1}$ and $\gamma = 4.1$. We see that the sky location and orbital frequency have all been recovered at the one-sigma level in both cases.

4. Keep the value of ζ fixed and perform a given number of injections with different binary orientation parameters (1000, for example) and determine the fraction of $\mathcal{F}$-statistic values that is larger than the value measured in the original data.

5. Repeat steps 2–4 until the strain amplitude is such that 95% of the injections give a value of the $\mathcal{F}$-statistic that is larger than the original value.

6. Record this value and repeat steps 2–5 at each frequency.

3.4.2 Simulated data sets

For this work we use a simulated pulsar timing array with sky locations drawn from uniform distributions in $\cos\theta$ and φ. All pulsars are assumed to have a distance of 1 kpc and a white noise rms of 100 ns with equal error bars. The timespan of the observations for all pulsars is 5 years with evenly spaced bi-monthly TOA measurements. Each set of residuals has been created by fitting a full timing model including spin-down, astrometric, and binary parameters (see Edwards et al. 2006 for details). As a check, in some simulations an uncorrelated red noise process with a power law spectrum $P(f) = Af^{-\gamma}$ is included in the residuals. This has no effect on our results. While these

simulated data sets do not include uneven sampling or extra fitting procedures like jumps or time varying DM variations, they do capture the essence of real timing residuals in the quadratic fitting of the spin-down parameters and the yearly and half yearly sinusoidal trends due to the sky location, proper motion and parallax fitting. Very uneven sampling is likely to reduce our sensitivity at higher frequencies and a detailed study of this problem will be presented in future work.

3.4.3 Implementation of the detection statistics

Here we will test our detection statistics on mock data sets with injected SMBHB GW signals in the presence of white and red Gaussian noise. We will focus primarily on the $\mathcal{F}_e$-statistic since, as we will show, it is a more robust detection statistic. Then, we will implement a procedure to produce a sensitivity curve for the GW strain amplitude as a function of frequency for a simulated NANOGrav (Demorest et al. 2013) array and plausible SKA arrays.

Figure 11 shows the posterior probability distributions of the intrinsic search parameters for simulated SMBHB signals in the presence of 100 ns white noise (Figure 11(a)) and uncorrelated red noise with amplitude $A = 4.22 \times 10^{-33}\,\mathrm{s}^{-1.1}$ and $\gamma = 4.1$ (Figure 11(b)). The two cases do have different realizations of the white noise, however, we can see that the $\mathcal{F}_e$-statistic does a very good job of determining the frequency and sky location of the source. In general, the $\mathcal{F}_e$-statistic is more robust than the $\mathcal{F}_p$ statistic because it produces explicit estimates of the sky location as well as the frequency, which is very important when looking for electromagnetic counterparts (Sesana et al. 2012; Tanaka et al. 2012).

It is possible to produce a sensitivity curve by a method that is similar to what we use to set upper limits. In this case we begin with simulated data with a given level of noise and no signal present. We follow the method presented in Sec. 3.4.1 except we now look for strain amplitude that gives a false alarm probability that is less than than our threshold (10^{-4} in our case) in 95% of realizations for each frequency. For clarity, we define the strain amplitude as

$$h = 2\frac{\mathcal{M}^{5/3}(\pi f_{\rm gw})^{2/3}}{d_L}, \tag{3.4.1}$$

where $f_{\rm gw} = \omega_0/\pi$. This amplitude comes from the overall scaling factor that results in differentiating Eq. (3.2.3) and (3.2.4) with respect to time. For simplicity and speed we have simplified this method for our sensitivity plots. Instead of performing a search at each frequency, we simply evaluate the $\mathcal{F}_e$ and $\mathcal{F}_p$ statistics at the values of the injected parameters. The purpose of these sensitivity plots is to illustrate the overall features of the different detection statistics and to give order of magnitude estimates of expected sensitivity for real data.

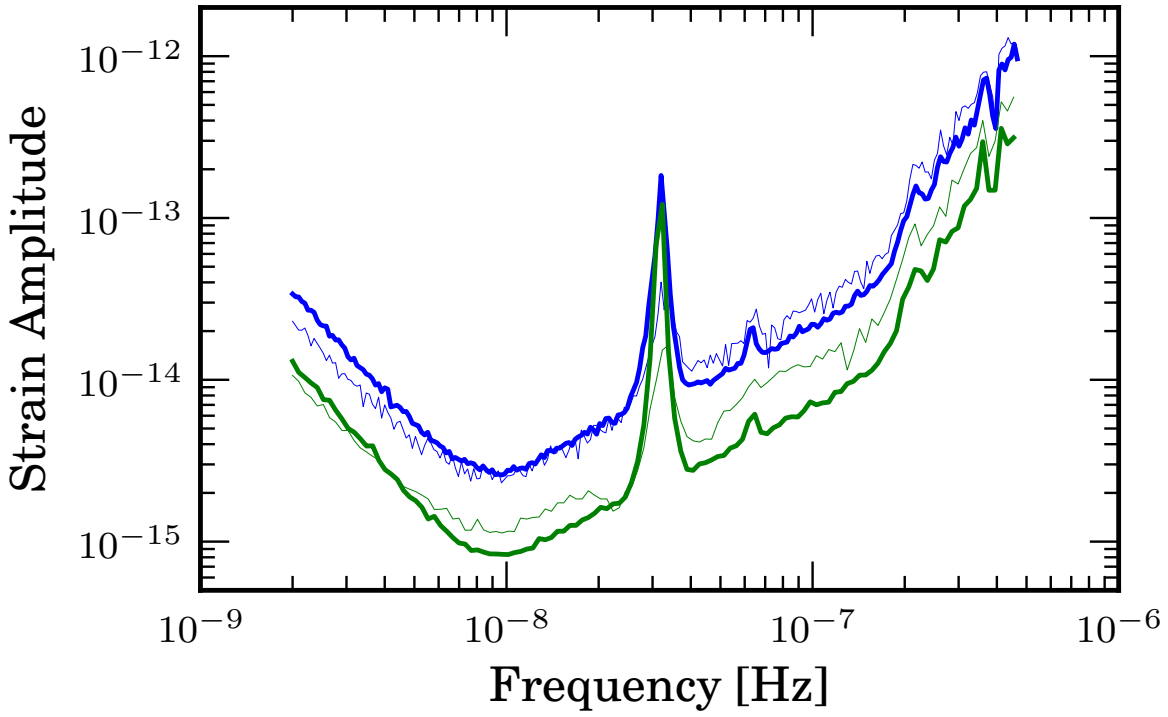


Figure 12 : Sensitivity curves for the $\mathcal{F}_e$ and $\mathcal{F}_p$ statistics for different PTA configurations (all pulsars have 100 ns residuals). The thick lines represent the $\mathcal{F}_e$-statistic and the thin lines represent the $\mathcal{F}_p$-statistic. The black(blue) curves are for a simulated PTA with 17 pulsars and the sky locations of the NANOGrav pulsars. The green(gray) curves a for a simulated PTA with 100 pulsars and random sky locations.

We have produced various sensitivity curves for both the $\mathcal{F}_e$ and $\mathcal{F}_p$ statistics in Figure 12. The two scenarios that we look at are a 17 pulsar simulated NANOGrav array in which we use the real sky location and timing models of the NANOGrav pulsars, and a simulated PTA with 100 pulsars at random sky locations. The loss in sensitivity at GW frequencies of $1\,{\rm yr}^{-1}$ and $0.5\,{\rm yr}^{-1}$ are due to the fitting of the pulsar's sky location and proper motion, and parallax, respectively. It is important to note that the sensitivity curves for the $\mathcal{F}_e$

and $\mathcal{F}_p$ statistics in the 17 pulsar case are very similar. Conversely, for the case of 100 pulsars the $\mathcal{F}_e$ statistic is more sensitive by a factor of ~ 2 for almost all frequencies. This is due to the different scaling relations of the statistics vs. the number of pulsars ($\mathcal{F}_e \propto \sqrt{M}$ while $\mathcal{F}_p \propto M^{1/4}$). However, the plot shows that the $\mathcal{F}_p$-statistic is more sensitive at lower frequencies and the $\mathcal{F}_e$-statistic is more sensitive at higher frequencies. There are two effects that contribute to this. The first is a result of our simulation and stems from the fact that we assume that for a given frequency, the maximum value of the $\mathcal{F}_e$-statistic is at the injected sky location. However, for low frequencies where the Earth and pulsar term are in the same frequency bin this assumption breaks down as the sky location will be biased (see e.g. Ellis et al. 2012b). The second effect is one inherent to our detection statistics themselves. As discussed in Sec. 3.3.3, the $\mathcal{F}_p$-statistic has different interpretations in the low and high frequency regimes. In the low frequency regime, it effectively contains the entire signal (Earth and pulsar terms), and in the high frequency regime it only contains the Earth term piece since the pulsar terms are out of that frequency bin. This distinction results in a different scaling relation for the ratio of $\mathcal{F}_e/\mathcal{F}_p$. In the low frequency case the $\mathcal{F}_e$-statistic scales coherently but it only has approximately *half* of the signal, whereas, the $\mathcal{F}_p$-statistic scales incoherently but has the *full* signal. Therefore, the ratio scales as $M^{1/4}/2$, thus the incoherent method will do better for $M \leq 16$. Conversely, in the high frequency regime, both statistics contain only *half* of the signal and the ratio scales as $M^{1/4}$. Therefore, the coherent statistic will do about a factor of 2 better than the incoherent method for $M \geq 16$.

3.5 Summary and Outlook

In this work we have adapted the standard $\mathcal{F}$-statistic (Jaranowski et al. 1998) to act as a detection statistic for continuous wave searches in realistic PTA data. We have also developed an incoherent detection statistic that maximizes over all pulsar contributions to the likelihood. Both of these detection statistics are implemented in the time domain to avoid spectral leakage problems associated with Fourier domain methods applied to irregularly sampled data. These methods take the pulsar timing model fitting into account and have been generalized to account for both correlated and uncorrelated colored

noise. Most of our analysis relies on dropping the pulsar term from our signal model as it will not add coherently. We have justified the use of this approximation in most astrophysically likely scenarios. It was shown that both detection statistics follow well known χ^2 distributions in the presence and absence of GW signals and therefore have well defined false-alarm probabilities. We have shown that the $\mathcal{F}_e$ statistic can not only confidently detect a GW signal but can also determine the sky location and frequency of the source to relatively high accuracy in the presence of white and colored Gaussian noise. A realistic implementation of a fully functional continuous GW pipeline starting from basic pulsar timing data and methods for computing upper limits on the strain amplitude were outlined in detail. Finally, we have used simulated data sets of various PTA configurations to produce sensitivity curves for our $\mathcal{F}$-statistics. From these sensitivity curves, we have shown that the sensitivity of the $\mathcal{F}_e$ and $\mathcal{F}_p$ statistics are very similar for $M \leq 25$ pulsars and that the $\mathcal{F}_e$ statistic becomes more sensitive for $M > 25$ and for higher frequencies.

As was shown in Ellis et al. (2012b), explicitly searching over the pulsar distances or somewhat equivalently, the GW phases at the pulsar locations (in the low frequency regime), is computationally prohibitive for $M \gtrsim 5$. A statistic that could maximize over these GW phases would greatly reduce the parameter space of the search, while still preserving the SNR of the full signal. The implementation of such an algorithm will be the subject of future work. However, we will give the derivation here. From Eq. (3.2.7), in the low-frequency limit we can write the signal in the following form

$$s_\alpha(t) = \sum_{i=0}^{4} \left[(\cos\Phi_\alpha - 1)\delta_i^j + \sin\Phi_\alpha \varepsilon_i{}^j\right] a_j A^i, \tag{3.5.1}$$

where $\Phi_\alpha = \omega L_\alpha(1 + \hat{\Omega}\cdot\hat{p}_\alpha)$, $a^i = a_i$ and A^i are defined in Equations (3.2.9) and (3.2.10), respectively, and the matix

$$\varepsilon_i{}^j = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3.5.2}$$

After some algebra, the log-likelihood ratio of Eq. (3.3.15) can be written as

$$\begin{aligned}\ln \Lambda = \sum_{\alpha=1}^{M} \big[& b_\alpha(\cos^2 \Phi_\alpha - \sin^2 \Phi_\alpha) + c_\alpha \cos \Phi_\alpha \\ & + d_\alpha \sin \Phi_\alpha + f_\alpha \sin \Phi_\alpha \cos \Phi_\alpha \big],\end{aligned} \tag{3.5.3}$$

with

$$b_\alpha = -\frac{1}{2} M_\alpha^{ij} a_i a_j \tag{3.5.4}$$

$$c_\alpha = N_\alpha^i a_i + M_\alpha^{ij} a_i a_j \tag{3.5.5}$$

$$d_\alpha = N_\alpha^i \varepsilon_{ij} a^j \tag{3.5.6}$$

$$f_\alpha = -M_\alpha^{ij} \varepsilon_{\ell j} a_i a^\ell \tag{3.5.7}$$

where M_α and N_α are defined by the following relations

$$M_\alpha^{ij} = (A_\alpha^i | A_\alpha^j) \tag{3.5.8}$$

$$N_\alpha^i = (r_\alpha | A_\alpha^i). \tag{3.5.9}$$

Maximizing the log-likelihood with respect to the pulsar phases Φ_β, we obtain

$$\begin{aligned}0 = \left. \frac{\partial \ln \Lambda}{\partial \Phi_\beta} \right|_{\hat{\Phi}} = & f_\beta(\cos^2 \hat{\Phi}_\beta - \sin^2 \hat{\Phi}_\beta) + 2 b_\beta \cos \hat{\Phi}_\beta \sin \hat{\Phi}_\beta \\ & - c_\beta \sin \hat{\Phi}_\beta + d_\beta \cos \hat{\Phi}_\beta = 0.\end{aligned} \tag{3.5.10}$$

Setting $x = \cos \hat{\Phi}_\alpha$, this expression reduces to a quartic equation of the form

$$\begin{aligned}0 = & (4 f_\alpha^2 + 16 b_\alpha^2) x^4 + (4 f_\alpha d_\alpha + 8 c_\alpha b_\alpha) x^3 \\ & + (c_\alpha^2 - 4 f_\alpha^2 - 16 b_\alpha^2) x^2 + (-2 f_\alpha d_\alpha - 8 c_\alpha b_\alpha) x \\ & + f_\alpha^2 - c_\alpha^2\end{aligned} \tag{3.5.11}$$

which is guaranteed to have at least one unique solution although it is unknown whether we are guaranteed to have a solution in the range $-1 \leq x \leq 1$ which corresponds to a physical value of $\hat{\Phi}_\beta$. This maximization results in a monumental reduction in the parameter space that needs to be searched. It takes on the order of $\sim 10^{2M}$ templates just to cover the pulsar phases (Ellis et al. 2012b). In practice, we could construct the various quantities M_α, N_α, $\mathbf{a}$, b_α, c_α, d_α, and f_α, solve Eq. (3.5.11) numerically to find the maximum likelihood estimators for all the pulsar phases. Substituting these solutions

back into our likelihood Eq. (3.5.3) still leaves us with the problem of searching over a 7 dimensional parameter space (since the amplitudes $\mathbf{a}$ depend on 4 parameters $(\zeta, \iota, \Phi_0, \psi)$ and the basis functions $\mathbf{A}$ depend on 3 parameters $(\theta, \varphi, \omega_0)$). We note, however, that this can be easily handled with a Markov chain Monte-Carlo (MCMC) or nested sampling algorithm.

The above derivation is valid but is quite challenging to implement practically. A very similar type of analysis is proposed in Taylor et al. (2013) in which the pulsar phase is maximized numerically as opposed to analytically as above.

Chapter 4

A Bayesian Analysis Pipeline for Continuous GWs

> "*Now you can luxuriate in a nice jail cell, but if your hand touches metal, I swear by my pretty floral bonnet I will end you.*"
>
> — Malcom Reynolds, *Firefly: Our Mrs. Reynolds*

This chapter is based on:
A Bayesian analysis pipeline for continuous GW sources in the PTA band
J. A. Ellis
CQG, (2013) 30, 224004

4.1 Introduction

A significant amount of work has gone into the detection problem for continuous GWs from SMBHBs. Both Jenet et al. (2004) and Yardley et al. (2010) use a Lomb-Scargle periodogram based approach to essentially measure the excess power that a continuous GW would induce compared to a noise only model. van Haasteren & Levin (2010) developed a Bayesian framework aimed at the detection of GW memory in PTAs; however, the authors mention that the methods presented could be used for continuous GW sources as well. Most recently, a maximized likelihood based approach has been developed by Babak & Sesana (2012); Ellis et al. (2012c) and was later extended to include multiple resolvable sources in Petiteau et al. (2013).

Many authors have focused on determining the parameter accuracy that we may hope to extract from a future detection of a continuous GW from a SMBMB. Sesana & Vecchio

(2010) use an Earth-term (to be defined in next section) only signal model to perform a study of SMBHB parameters that are measurable with PTAs using a Fisher matrix approach. Corbin & Cornish (2010) have developed a Bayesian Markov Chain Monte-Carlo (MCMC) data analysis algorithm for parameter estimation of a SMBHB system in which the pulsar term is taken into account in the detection scheme, thereby increasing the signal-to-noise-ratio (SNR) and improving the accuracy of the GW source location on the sky. Recently, Lee et al. (2011) have developed parameter estimation techniques based on vector Ziv-Zakai bounds incorporating the pulsar term and have placed limits on the minimum detectable amplitude of a continuous GW source. In the aforementioned work, the authors also propose a method of combining timing parallax measurements with single-source GW detections to improve pulsar distance measurements.

In this work we introduce a fully functional Bayesian pipeline aimed at both detection and parameter estimation of single continuous GWs. To this end, we make use of MCMC augmented with Parallel Tempering, an adaptive jump proposal scheme and thermodynamic integration for evidence evaluation. Previous work has made use of the Fisher matrix or similar techniques to either estimate parameter uncertainties or propose jumps in an MCMC algorithm. Since it is known that the Fisher matrix is limited in use and only applies to large SNR (Vallisneri 2008), we choose to use an Adaptive Metropolis (AM) approach first developed in Haario et al. (2001); Andrieu & Thoms (2008) and later applied to cosmology and GW parameter estimation in van der Sluys et al. (2008); Taylor & Gair (2012); Taylor et al. (2012).

The layout of the chapter is as follows. In section 4.2 we introduce the signal model and notation used in this work. In section 4.3 we briefly review MCMC techniques, adaptive metropolis, parallel tempering, thermodynamic integration and introduce our likelihood function and priors. In section 4.4 we introduce the semi-realistic simulated datasets that we use to test our algorithm. In section 4.5 we test our algorithm on simulated data and make a few statements about the measurability of SMBMB parameters in realistic PTA data sets. Finally, we briefly mention future work and conclude in section 4.6.

4.2 The signal model

Here we use the signal model of Eqs. 3.2.3 and 3.2.4. From this signal model and the following frequency evolution equations

$$\Phi(t) = \Phi_0 + \frac{1}{32\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega(t)^{-5/3}\right) \tag{4.2.1}$$

and

$$\omega(t) = \left(\omega_0^{-8/3} - \frac{256}{5}\mathcal{M}^{5/3}t\right)^{-3/8}, \tag{4.2.2}$$

we see that our parameter space is $(8 + N_{\rm psr})$ dimensional where, $N_{\rm psr}$ is the number of pulsars in the array and the continuous wave parameter space vector is

$$\vec{\lambda} = \{\theta, \varphi, \Phi_0, \psi, \iota, \mathcal{M}, d_L, \omega_0\}, \tag{4.2.3}$$

where Φ_0 and ω_0 are the initial values at the time of our first observation, the chirp mass is defined by $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}(1 + z)$, where m_1 and m_2 are the masses of the two SMBHs, z is the redshift, d_L is the luminosity distance to the source, ψ is the polarization angle and (θ, φ) are the polar and azimuthal sky locations of the GW source. However, since typical pulsar distance uncertainties are on the order of tens of percent (Verbiest et al. 2012), in order to attain phase coherence in our search algorithm, we must allow the pulsar distance to vary as a search parameter as well. Henceforth, we will adopt the notation that $\vec{\lambda}_\alpha = \{\vec{\lambda}, L_\alpha\}$ in order to denote the fact that the pulsar distance is a search parameter.

Eqs. 4.2.1 and 4.2.2 are true in general and can be applied when the frequency evolves appreciably over the total observing time. However, it is very useful to work under the assumption of slowly evolving binaries where $T_{\rm chirp} \gg T$, with T the observing time and

$$T_{\rm chirp} = \frac{\omega_0}{\dot{\omega}} = 3.2 \times 10^5\,{\rm yr}\left(\frac{\mathcal{M}}{10^8\,{\rm M}_\odot}\right)^{-5/3}\left(\frac{f_0}{1 \times 10^{-8}\,{\rm Hz}}\right)^{-8/3}, \tag{4.2.4}$$

where

$$\dot{\omega} = \frac{96}{5}\mathcal{M}^{5/3}\omega_0^{11/3}. \tag{4.2.5}$$

Since typical PTA observations are on the order of 10–20 years and $T/T_{\rm chirp} \sim 10^{-4}$, this is a safe assumption for a broad range of masses and initial orbital frequencies of interest.

With this approximation we can write the orbital frequency and phase for the earth term simply as

$$\Phi_e(t) = \Phi_0 + \omega_0 t \tag{4.2.6}$$

$$\omega_e(t) = \omega_0. \tag{4.2.7}$$

However, for the pulsar term we are dealing with the retarded time and must include the first order corrections to the orbital frequency and phase

$$\Phi_p(t) = \Phi_0 + \omega_0 t - \omega_0 L(1 + \hat{\Omega} \cdot \hat{p}) - \dot{\omega} L(1 + \hat{\Omega} \cdot \hat{p})t \tag{4.2.8}$$

$$\omega_p(t) = \omega_0 - \dot{\omega} L(1 + \hat{\Omega} \cdot \hat{p}), \tag{4.2.9}$$

where L is on the order of a kpc and the last term in the pulsar phase containing $\dot{\omega}$ terms is responsible for any frequency evolution over the earth-pulsar light crossing time. As we will see later, writing the pulsar phase in this way will become very useful.

4.3 Implementation

Our use of MCMC in this work is motivated by the fact that our parameter space will be at least 9-dimensional for a PTA comprised of one pulsar and we will gain another parameter for every pulsar that is used in the search. For typical PTAs (20 pulsars), this means that the parameter space will be $\sim$ 28 dimensional. While a rudimentary template based search technique was explored in Ellis et al. (2012b), no real effort has gone into an efficient template based search technique. Although this parameter space is high dimensional, it may indeed be possible to efficiently cover the parameter space using methods such as lattice covering (Prix 2007), stochastic template banks (Harry et al. 2009), or random template banks (Messenger et al. 2009). Such investigations of these methods applied to the PTA problem, however; are beyond the scope of this work, and as we will show, using MCMC as a search technique is still quite efficient in this case.

4.3.1 Jump Proposals

As mentioned in Chapter 2 we use an AM scheme to update the covariance matrix for multidimensional gaussian jumps. However since our parameter space is quite large

$(8 + N_{\rm psr})$ we do not always update all parameters simultaneously. In ∼70% of jumps we will jump in subsets of correlated parameters such as the sky location parameters and pulsar distance as well as the chirp mass and distance. In ∼20% of jumps we update all parameter simultaneously and in the remaining ∼10% of jumps we choose one parameter at random and propose large jumps in parameter space.

In order to ensure proper mixing and exploration of our chains we have chosen to expand the parameter space in the following way. If we introduce the initial pulsar phase

$$\phi_p = \omega_0 L(1 + \hat{\Omega} \cdot \hat{p}) \tag{4.3.1}$$

and then solve for the pulsar distance

$$L = \frac{\phi_p}{\omega_0(1 + \hat{\Omega} \cdot \hat{p})} + \frac{2\pi n}{\omega_0(1 + \hat{\Omega} \cdot \hat{p})} = L^{\rm small} + L^{\rm big}, \tag{4.3.2}$$

where n is the number of times the phase has wrapped around 2π (typically 1000s). By writing the distance to the pulsar in this fashion we can separate out the very small scale fluctuations ($L^{\rm small}$) that are important for coherence and are typically less than a pc, and the large scale fluctuations ($L^{\rm big}$) that are on the order of a kpc are important for determining the frequency evolution of the binary. These two components are essentially independent and explain physics on vastly different scales. So now re-writing Eq. 4.2.8 we have

$$\Phi_p(t) = \Phi_0 + \omega_0 t - \phi_p - \dot{\omega} L(1 + \hat{\Omega} \cdot \hat{p})t, \tag{4.3.3}$$

where we jump in *both* ϕ_p and $L \approx L^{\rm big}$. It should be noted that for certain parameter combinations (particularly large chirp mass and high frequency) the approximation to the pulsar phase in Eq. 4.3.3 may differ from the true phase by a *constant* that depends on the pulsar distance. However, since this is only a constant offset, it can be absorbed into ϕ_p and this approximation is still completely valid.

4.3.2 Likelihood and Priors

Following section 2.4.3 we write the pulsar timing residuals as

$$\delta t = M\delta\boldsymbol{\xi} + n + s, \tag{4.3.4}$$

where δt are the timing residuals, M is the design matrix, $\delta\boldsymbol{\xi}$ is the parameter offset between the true pulsar timing parameters and our best fit parameters, n is the noise present in the TOAs (radiometer noise, red noise, etc.), and s is our continuous GW signal. The likelihood function can be marginalized over the timing model parameters $\delta\boldsymbol{\xi}$ to obtain

$$p(\delta t|\vec{\theta},\vec{\lambda}) = \frac{\exp\left[-\frac{1}{2}\left(\delta t - s\right)^T G(G^T C G)^{-1} G^T \left(\delta t - s\right)\right]}{\sqrt{(2\pi)^{n-m}\det(G^T C G)}}, \tag{4.3.5}$$

where G is an $n \times (n-m)$ matrix with n the number of TOAs and m the number of fitted parameters in the timing model.

For this work we will assume that the noise parameters $\vec{\theta}$ are know from some noise estimation done beforehand (see e.g. van Haasteren & Levin (2013); Ellis et al. (2014)) and will only focus on characterizing the continuous GW parameters $\vec{\lambda}$. We will also assume that the residuals between pulsars are uncorrelated. In other words, we are assuming that the stochastic GW background will be negligible compared to the intrinsic noise in each pulsar. In general this is not likely to be a good assumption when we would expect a detection of a single GW source. The effects of omitting the correlations in the likelihood function are unknown and will be the subject of future work. Under these assumptions, the likelihood function for the full PTA can be written as

$$p(\delta t|\vec{\lambda}) = \prod_{\alpha=1}^{N_{\rm psr}} p(\delta t_\alpha|\vec{\lambda}_\alpha), \tag{4.3.6}$$

where δt_α and $\vec{\lambda}_\alpha$ and the residuals and model parameters for the αth pulsar, respectively. Since we are assuming the noise is fixed (and known) then we can write the log-likelihood ratio of a model with a single continuous GW to a model with just noise as

$$\ln\,\Lambda = \sum_{\alpha}^{N_{\rm psr}} \left[\left(\delta t_\alpha|s(\vec{\lambda}_\alpha)\right) - \frac{1}{2}\left(s(\vec{\lambda}_\alpha)|s(\vec{\lambda}_\alpha)\right)\right], \tag{4.3.7}$$

where the inner product between two time-series x and y is

$$(x|y) = x^T G(G^T C G)^{-1} G^T y. \tag{4.3.8}$$

We choose isotropic priors on the angular parameters and flat priors in the log of the chirp mass, luminosity distance, and frequency of the GW. For the pulsar distance prior

we use the current electromagnetic (EM) measurements either from timing parallax or Very Long Baseline Interferometry (VLBI) to contain the prior space as follows

$$p(\vec{L}) = \prod_{\alpha=1}^{N_{\rm psr}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left(-\frac{(L_\alpha - L_\alpha^{\rm EM})^2}{2\sigma_\alpha^2}\right), \tag{4.3.9}$$

where $L_\alpha^{\rm EM}$ is the best measured distance for the αth pulsar and σ_α is the 1-sigma uncertainty on that distance measurement.

4.4 Simulated data sets

In this work we will simulate "toy model" datasets that represent realistic yet optimistic present day residuals. We have chosen an array of 10 pulsars that are meant to represent the best 10 IPTA pulsars in terms of timing precision. The datasets have uneven sampling, varying error bars, and time spans corresponding to the real pulsar observing span. The data is summarized in Table 1. To create this data we use the mean RMS from the IPTA

Table 1 : Simulated IPTA pulsar datasets. The RMS values are measured from the data with no injected signal. The pulsar distances are taken from Verbiest et al. (2012) if available. Otherwise the pulsar distances were taken from the ATNF catalog.

Pulsar Name	RMS [ns]	Time Span [yr]	Pulsar Distance [kpc]
J0437−4715	69	14.8	0.156 ± 0.001
J1909−3744	100	9.0	1.26 ± 0.03
J1713+0747	136	18.3	1.05 ± 0.06
J1939+2134	141	16.3	5.0 ± 2.0
J1744−1134	366	16.9	0.42 ± 0.02
J1857+0943	402	14.9	0.9 ± 0.2
J1640+2224	410	14.9	1.19 ± 0.24
J2317+1439	412	14.9	1.89 ± 0.38
J1824−2452	602	5.7	3.6 ± 0.72
J0030+0451	792	12.7	0.28 ± 0.1

pulsars and draw each residual from a gaussian distribution centered on the RMS with

a standard deviation of 50% of the RMS. This way we are taking into account varying error bars and assuring that we only have gaussian white noise. We then simulate a continuous GW signal as in section 4.2 and add it to our simulated noise. Finally, in an attempt to take into account the most important part of the timing model, we fit out a 2nd order polynomial from the data. The pulsar distances and uncertainties used in this analysis are the best measured values taken from Verbiest et al. (2012) if available, otherwise, we use the values from the Australia National Telescope Facility (ATNF) pulsar catalog (Manchester et al. 2005)[1] and assume a 20% uncertainty. The rough cadence is chosen to simulate bi-monthly sampling. In order to present an idealistic yet plausible representation of current IPTA data sets, we have chosen to not include any intrinsic red noise which would only act to decrease sensitivity at low frequencies, therefore; the results presented here are likely to be optimistic.

4.5 MCMC simulations

In this section we wish to test the efficacy of our algorithm by injecting continuous GW signals into our simulated datasets described above. Although our main goal is to test our algorithm, we also wish to add a certain level of realism to these simulations. For this reason we have used mock IPTA datasets and will focus any astrophysical statements mostly to low SNR sources (SNR ~ 7) as this represents a realistic possibility in the next decade. We also include injections at higher SNR and mimic these injections in ideal datasets (10 pulsars timed for 10 years all with 100 ns RMS drawn from an isotropic distribution on the sky) which have been used in previous parameter estimation work for PTAs (Corbin & Cornish 2010; Sesana & Vecchio 2010; Lee et al. 2011).

Recent work has shown that there may be potential single GW source "hot spots" in the Virgo, Fornax and Coma clusters (Simon et al. 2013). Since our purpose here is only to illustrate the efficacy of our algorithm, we have randomly chosen to inject GW sources at the sky location corresponding to the Fornax cluster with a chirp mass of $\mathcal{M} = 7 \times 10^8 M_\odot$ and initial orbital period of 3.16 yr. The distance to the GW source is

[1] `http://www.atnf.csiro.au/people/pulsar/psrcat/`

then scaled such that we achieve the desired SNR defined by

$$\mathrm{SNR}^2 = \sum_{\alpha} \Big(s(\vec{\lambda}_{\mathrm{inj}}) | s(\vec{\lambda}_{\mathrm{inj}}) \Big)_{\alpha} , \tag{4.5.1}$$

where the sum is over the number of pulsars and $\vec{\lambda}_{\mathrm{inj}}$ are the injected source parameters. This choice of injected parameters is justified since the amplitude of our GW induced residuals scales as $\mathcal{M}^{5/3}\omega^{-1/3}$ and the stochastic GWB and other potential red noise sources will lower our sensitivity at lower frequencies. Therefore, we are likely to detect a source with high chirp mass and high frequency. See table 2 for a list of the different GW sources and parameters used in this work. For each source, the same noise realization was

Table 2 : Simulated GW source parameters. These sources are injected at the sky location of the Fornax cluster and the distance is scaled such that we achieve the desired SNR.

SNR	θ [rad]	φ [rad]	ψ [rad]	ι [rad]	Φ_0 [rad]	$\mathcal{M}$ [$M_\odot$]	D_L [Mpc]	f_{gw} [Hz]
7	2.17	0.95	1.26	1.57	0.99	7.0×10^8	223.4	2×10^{-8}
14	2.17	0.95	1.26	1.57	0.99	7.0×10^8	111.7	2×10^{-8}
20	2.17	0.95	1.26	1.57	0.99	7.0×10^8	78.2	2×10^{-8}

used so that relative parameter accuracies do not depend on this specific noise realization. In general we would like to do a much more detailed analysis with many different noise realizations and many different injected sources. Indeed, this will be the subject of future work, however; here we simply want to test the various steps of our algorithm, that is, the search phase where we find the global maxima in the multi-dimensional parameter space, the sampling phase where we obtain samples from the underlying posterior distribution, and finally the evaluation phase where we compute the evidence and Bayes factors to make choices about detection.

4.5.1 Searching for global maxima

Since we have little information about the SMBMB population, we want to carry out a blind search of the parameter space making no assumptions about the underlying SMBMB source parameters. Therefore, it is very important that our algorithm be able to quickly

find the global maxima of the log-likelihood function and the true parameters so that the sampling process can begin. The trace plots of one SNR = 20 injection is shown in Figure 13 where we have plotted the measurable parameters (excluding the pulsar distance) as

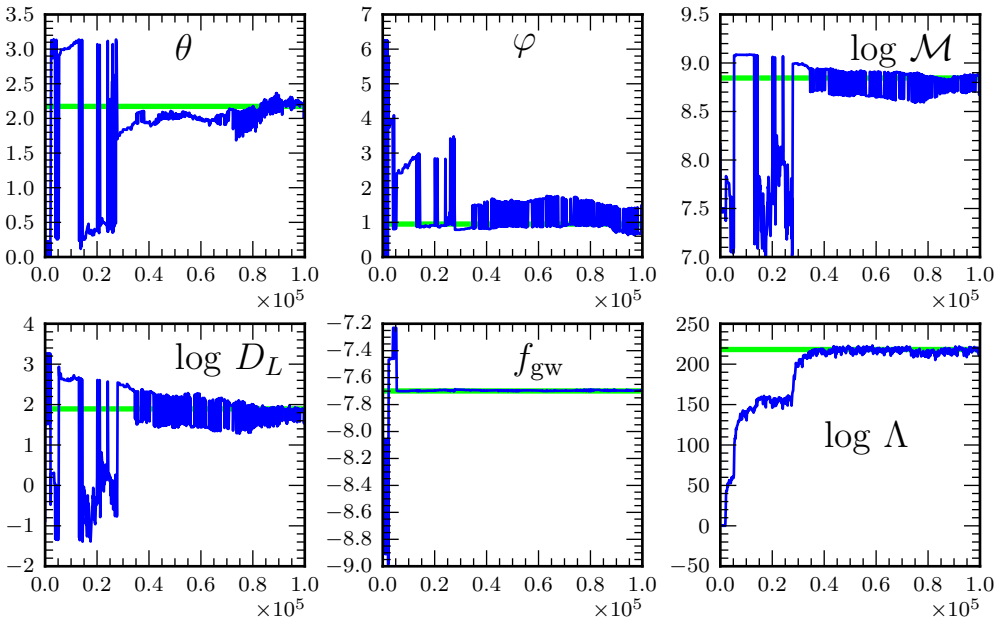


Figure 13 : Trace plots for the measurable parameters (the inclination angle, initial phase and polarization angle are not well constrained for this realization) for an SNR=20 injection for the first 10^5 steps. In all cases the black(green) line represents the injected parameters and the gray(blue) is the chain trace. We can see that the parallel tempering scheme has allowed us to locate the global maxima of the log-likelihood and all parameters within the first $\sim 6 \times 10^4$ steps.

well as the log-likelihood as a function of chain iteration for the $T = 1$ chain. Here we do not plot the polarization angle, initial phase, or inclination angle as they are not well constrained by the data and contribute little to the overall log-likelihood for this case. We can see from the figure that the algorithm has correctly found the true source parameters within the first $\sim 6 \times 10^4$ MCMC iterations. We note that the true value of the frequency is found quickly (within the first 10^4 steps of the algorithm) and we reach the true value of the log-likelihood within the first 4×10^4 steps. There are several ways that we could improve this step such as choosing a more suitable starting jump proposal distribution before starting adaptation or even starting adaptation sooner, however for the purpose of this work we believe that this is sufficient as the algorithm can still collect $\sim 2 \times 10^6$ samples with 8 chains in about 4 hours running on a 2.7 GHz quad core MacBook Pro.

It is also important to note that in practice we will have carried out a simpler search algorithm such as an $\mathcal{F}$-statistic Babak & Sesana (2012); Ellis et al. (2012c); Petiteau et al. (2013) search prior to this Bayesian analysis. If any signal is detected, then we will have a very good idea of the frequency of the GW source and can therefore seed our MCMC algorithm much closer to the true value. Since the frequency contributes heavily to the log-likelihood, it is likely that this could reduce the number of samples required for this search phase by at least an order of magnitude.

4.5.2 Sampling and parameter estimation

For each injected source we run 4 serial chains all with 8 temperatures and starting positions chosen at random from the prior, thereby assuring that our algorithm can indeed locate the global maxima. Each serial chain was run for $\sim 1.5 \times 10^6$ iterations and 25% of each chain was discarded as burn in. The resulting post burn-in chains were then concatenated to form a single chain with $\sim 4.5 \times 10^6$ posterior samples.

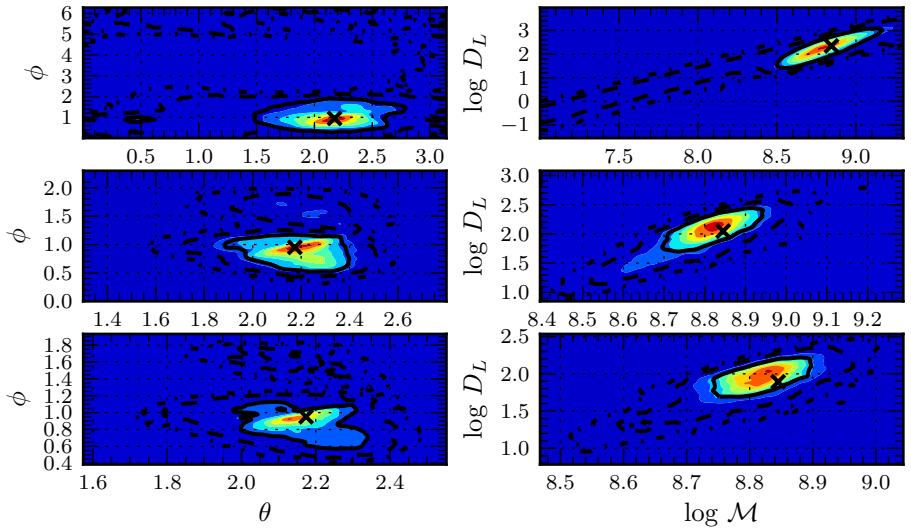


Figure 14 : Marginalized 2-D posterior pdfs in the sky coordinates (θ, ϕ) and the log of the chirp mass and distance $(\log \mathcal{M}, \log D_L)$ for injected SNRs of 7, 14, and 20 shown from top to bottom. Here the injected GW source is in the direction of the Fornax cluster with chirp mass $\mathcal{M} = 7 \times 10^8 M_\odot$. The distance to the source is varied to achieve the desired SNR. Here the "×" marker indicates the injected parameters and the solid, dashed and dot-dashed lines represent the 1, 2, and 3 sigma credible regions, respectively.

Figure 14 shows marginalized 2-D posterior pdfs of the sky coordinates (θ, φ) and the

log of the chirp mass and distance ($\log \mathcal{M}, \log D_L$) for injected SNRs of 7, 14, and 20 (shown from top to bottom) for a source injected at the sky position of the Fornax cluster. The "×" marker indicates the injected parameters and the solid, dashed and dot-dashed lines represent the 1, 2, and 3 sigma credible regions, respectively. The first thing to note from this figure is that the injected value lies well within the 1-sigma credible regions in all three cases. We also note that since we have injected a relatively high mass and high frequency GW, we can measure $\dot{f}$ and therefore; we can can break the degeneracy between chirp mass and distance as is seen in the plots on the right in the above figure. Since we know the true injected values, it is possible to determine how much each pulsar contributes to the log-likelihood function. For the aforementioned injection, four pulsars contribute more than 1% to the likelihood function for the SNR 7 injection and only three pulsars contribute more than 1% to the likelihood function for the SNR 14 and 20 injections. While this number does depend on the relative sky locations of the pulsars and the GW source as well as the specific noise realization, it is also a very strong function of the RMS of the noise in each pulsar ($\langle \ln \Lambda \rangle \propto \sigma_{\rm RMS}^{-2}$). In fact, we can see the results of this in Figure 14 where there is a bit of multi modality in the posterior for sky position because we essentially only have three and four baselines (detectors) for the SNR 7 and 14 and 20 cases, respectively.

This type of parameter degeneracy due to the small number of baselines differs from previous parameter estimation studies Corbin & Cornish (2010); Sesana & Vecchio (2010); Lee et al. (2011) where the simulated PTA consisted of a large number (20 or more) of pulsars all timed to the same accuracy. For this reason, quoted SMBHB parameter accuracies that can be obtained from PTAs should be interpreted cautiously as it is extremely unlikely that future era PTAs will even approach this ideal situation. To illustrate this point we have also simulated an ideal data set of 10 pulsars drawn uniformly on the sky with 100 ns RMS in each with baselines of 10 years. We also chose distances drawn uniformly from the range $L \in [0.5, 1, 5]$ kpc with 10% uncertainties. We have then used the same injection as in the simulated IPTA data at SNRs of 7, 14 and 20. The sky resolution Cutler (1998) and fractional uncertainties on the chirp mass and distance for the simulated IPTA dataset are $\Delta\Omega = (2357.9, 122.2, 67.2)$ deg^2, $\Delta\mathcal{M}/\mathcal{M} = (48.8\%, 9.5\%, 6.3\%)$ and $\Delta D_L/D_L =$

$(81.2\%, 28.2\%, 19.9\%)$, respectively. Whereas, for our ideal simulated datasets the corresponding values are $\Delta\Omega = (1085.9, 23.7, 12.8)$ deg^2, $\Delta\mathcal{M}/\mathcal{M} = (47.9\%, 4.4\%, 3.0\%)$ and $\Delta D_L/D_L = (79.7\%, 15.9\%, 13.2\%)$, respectively. Again, these results are not robust, in that we have only done one injection (with varying SNR) into one noise realization. Nonetheless, it should be clear that our simulated IPTA data do not yield nearly as precise sky resolution or chirp mass and distance fractional uncertainties as an ideal data set.

4.5.3 Evaluating the evidence

After we have carried out our parallel tempering MCMC search we can make use of the different temperature chains to calculate the evidence integral via Eq. (2.7.6). Since we have measured the noise parameters before conducting our search, we use the log-likelihood ratio defined in Eq. (5.4.2) as our log-likelihood. By doing this we can compute the Bayes factor comparing our GW and noise models simply by calculating the evidence using the log-likelihood ratio. Figure 15 shows the log of the Bayes factor computed from

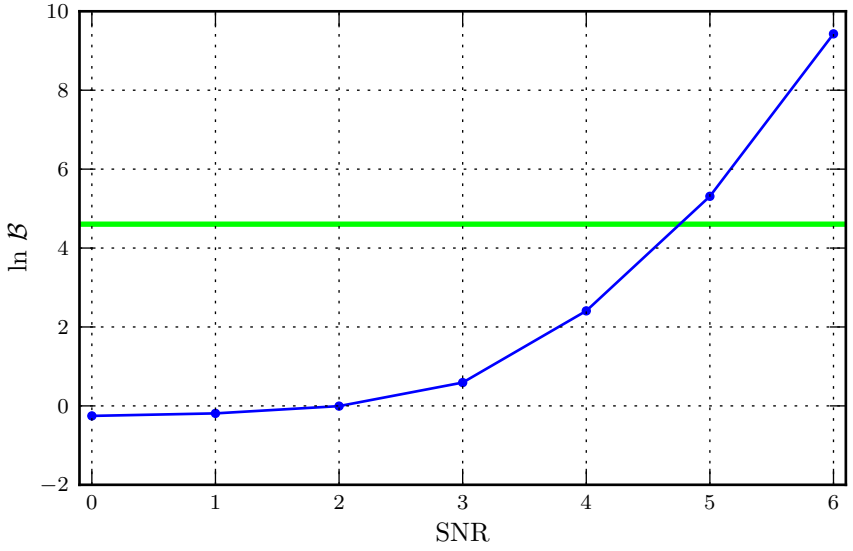


Figure 15 : Log of the Bayes factor plotted against injected SNR for the same signal and noise realization. The gray(green) horizontal line is the threshold in the log of the Bayes factor in which we can claim a detection and the black(blue) points are the log Bayes factor calculated from thermodynamic integration.

thermodynamic integration for injections at different SNRs. Here we have done injections into the same noise realization of out simulated IPTA data using the same GW source

(again with the distance scaled to give the desired SNR) as above. The MCMC sampler was run with 10 temperature chains for $\sim 2 \times 10^6$ iterations. In the figure, the green horizontal line represents our threshold in the log of the Bayes factor of $\ln 100$, above which there is decisive evidence for a GW source (Jeffreys 1961) and the black(blue) points are the computed log Bayes factor for each injection. There are two important things to note. First, notice that the log of the Bayes factor is above the threshold for injected sources with SNR ≥ 5 which agrees well with a frequentist interpretation of the SNR as a detection statistic in gaussian noise, where 5-sigma is usually required for a definitive detection. Secondly, as was discussed in Littenberg & Cornish (2009), the Bayes factor is about unity for the zero to low SNR injections. This is because of the nature of the question that we are asking. In this case we are asking "Is there evidence for *any* continuous GW source in the data?". Framed in this way, the result makes perfect sense because a low SNR signal is nearly indistinguishable from pure noise, therefore the odds of a low SNR GW are about 50/50 indicated by a Bayes factor of 1. If we were to ask the question "Is there a continuous GW source with SNR ≥ 5 in the data?", then we would expect the Bayes factor to become much less than unity at low SNR.

4.6 Conclusions and future work

We have developed a robust MCMC algorithm that makes use of an Adaptive Metropolis scheme and parallel tempering for use in PTA detection and parameter estimation of single sources of GWs from SMBHBs. We have tested the algorithm on a fairly realistic simulated IPTA dataset that has many of the features of real data including uneven sampling, varying error bars and overall noise levels, poor pulsar distance measurement uncertainty and varying data span. For comparison we have also run the algorithm on ideal datasets, similar to those that have been considered in the literature. The algorithm has shown to perform well in the three stages of our Bayesian analysis pipeline, namely the search, sampling and evaluation phase. When seeded from a random point in parameter space, the algorithm can quickly locate the global maxima through the use of parallel tempering. Posterior samples are then collected efficiently through the use of Adaptive Metropolis and special jump proposals in an extended parameter space. Finally, we have

shown that this algorithm can also be used for detection through the use of parallel tempering and thermodynamic integration to calculate the Bayesian evidence.

From the few simulations and comparisons of realistic vs. ideal data done in this work we can say that parameter estimation from current generation PTAs, counter to previous work on the subject, is likely to suffer due to the fact that few pulsars contribute to the total network SNR, resulting in a lower number of effective "detectors" than the number of pulsars in the array. A much more detailed study of the parameter estimation problem in current generation PTAs with more realistic noise models (including effects such as time varying Dispersion Measure) is underway and will be the subject of a future paper.

Chapter 5

NANOGrav Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits

> "*Which side are we on? We're on the side of the demons, chief. We're evil men in the gardens of paradise, sent by the forces of death to spread devastation and destruction wherever we go.*"
>
> — Colonel Saul Tigh, *Battlestar Galactica: Precipice*

This chapter is based on:
NANOGrav Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits
NANOGrav Collaboration
arXiv 1404:1267 (2014), Submitted to ApJ

5.1 Introduction

The direct detection of Gravitational Waves (GWs) is a major goal of experimental physics and astrophysics. One of the most promising means of detecting GWs is through the precise timing of an array of millisecond pulsars (MSPs). PTAs are most sensitive to GWs with frequencies in the nanohertz regime (i.e., 10^{-9} Hz – 10^{-7} Hz). The community has thus far focused mostly on stochastic backgrounds produced by a variety of sources; however, sufficiently nearby individual SMBHBs may produce detectable continuous waves

with periods on the order of years for masses in the range $10^8 M_\odot$–$10^{10} M_\odot$ (Wyithe & Loeb 2003; Sesana et al. 2009; Sesana & Vecchio 2010). Several upper limits have been placed on the strength of the stochastic background (Kaspi et al. 1994; Jenet et al. 2006; van Haasteren et al. 2011; Demorest et al. 2013; Shannon et al. 2013) and continuous waves (Jenet et al. 2004; Yardley et al. 2010) but no successful detection has yet been made.

In this chapter we will use current-generation frequentist (Ellis et al. 2012c) and Bayesian (Ellis 2013) data analysis pipelines to compute upper limits on the strain amplitude of continuous GWs from SMBHBs in circular orbits. We make use of the 5-year, 17 pulsar data set obtained as part of the NANOGrav project (Demorest et al. 2013). In Section 5.2 we briefly review the radio observations and timing analysis. In Section 5.3 we describe the signal model used to describe the continuous GWs in the PTA band. In Section 5.4 we describe, in detail, the time domain likelihood function, the noise model, and the frequentist and Bayesian search pipelines. In Section 5.5 we apply our search and upper limit pipelines to the NANOGrav dataset and report our findings. In section 5.6 we summarize our results. In the Appendices we derive the form of the frequency evolution of SMBHBs, and give full details on the computational implementation of our Bayesian code.

5.2 Observations and Timing Analysis

The observational data used for this analysis are the same as those presented by Demorest et al. (2013); the reader is referred to that paper for a detailed description of the observations and timing analysis. Here we present a brief review of the relevant features of the data set. The timing data used here were acquired during 2005–2010 using two radio telescopes, the 305 m Arecibo telescope, and the 100 m Robert C. Byrd Green Bank Telescope (GBT). A total of 17 pulsars (8 at Arecibo, 10 at the GBT, with J1713+0747 observed by both telescopes) were monitored using a typical observational cadence of 4–6 weeks between sessions. At each observing epoch, every pulsar was observed using two separate receiver systems operating at widely separated radio frequencies ranging from 327 MHz to 2.3 GHz. The typical observation length was 30 minutes per pulsar per

receiver. All data were recorded using the identical ASP (at Arecibo) and GASP (at the GBT) pulsar backend systems (Demorest 2007). These systems processed a typical radio bandwidth of 64 MHz using real-time coherent dedispersion and pulse period folding, resulting in 2048-bin full-Stokes pulse profiles averaged over 1–3 minutes in 4 MHz channels.

Pulse profile calibration, integration, and time of arrival (TOA) determination was done using standard techniques via the PSRCHIVE[1] software package (Hotan et al. 2004). For each pulsar all profiles in a given epoch were integrated in time to form a single set of profiles across radio frequency. From these, TOAs were measured *separately* in each 4 MHz radio frequency channel. This resulted in a set of ∼20–30 multi-frequency TOAs at each epoch, or ∼500–2000 TOAs total for each pulsar over the full data set. Before searching for the presence of GW in these data, the rotational, orbital, astrometric and interstellar medium properties specific to each pulsar–effects collectively known as the timing model, must first be determined from the TOA data. For this we analyzed the TOAs using both the TEMPO[2] and TEMPO2[3] (Hobbs et al. 2006) timing software packages and obtained identical results with both. Notable features of the timing models used here include: Spin frequency and spin-down rate, but no higher frequency derivatives, were fit for all pulsars; all five astrometric parameters (sky position, proper motion and parallax) were fit for all pulsars[4]; time-variable dispersion measure (DM); was included by fitting foran *independent* DM value at each epoch, using the multi-frequency TOAs;[5] intrinsic profile shape evolution with frequency was included as a constant-in-time offset for each frequency channel, and Keplarian and relativistic orbital elements, as appropriate for pulsars in binary systems. All TOA data and final timing solutions for this data set are publicly available online.[6]

[1] `http://psrchive.sourceforge.net`

[2] `http://tempo.sourceforge.net`

[3] `http://tempo2.sourceforge.net`

[4] Parallax was not fit for in PSR J1640+2224.

[5] Models for pulsars J1853+1308, J1910+1256 and B1953+29 did not include DM variation measurement as only single-frequency data were available for these.

[6] `http://data.nanograv.org`

5.3 GWs From Supermassive Black Hole Binaries

Here, again, we will use the signal model presented in chapter 1 including both the earth and pulsar terms as well as the full frequency evolution. From the signal model presented above, we see that our parameter space is 8 dimensional and the continuous wave parameter space vector is

$$\vec{\lambda}_0 = \{\theta, \varphi, \Phi_0, \psi, \iota, \mathcal{M}, d_L, \omega_0\}. \tag{5.3.1}$$

However, because typical pulsar distance uncertainties are on the order of tens of percent (Verbiest et al. 2012), in order to attain phase coherence in our search algorithm, we must allow the pulsar distance to vary as a search parameter as well. Henceforth, we will adopt the notation that $\vec{\lambda}_\alpha = \{\vec{\lambda}_0, L_\alpha\}$, where L_α is the distance to the αth pulsar, in order to denote the fact that the pulsar distance is a search parameter. The above parameter set represents the default parameters used in our search; however, when setting upper limits we wish to parameterize the upper limit in terms of the inclination averaged strain amplitude

$$h_0 = 4\sqrt{\frac{2}{5}}\frac{\mathcal{M}^{5/3}(\pi f_{\rm gw})^{2/3}}{d_L}. \tag{5.3.2}$$

Since the luminosity distance, d_L is only a scale parameter we use h_0 as a free parameter in the waveform instead of luminosity distance when computing upper limits.

5.4 Search Techniques

5.4.1 Likelihood Function for PTAs

For this work we will use the marginalized likelihood of Eq. (2.4.14) and assume that the residuals between pulsars are uncorrelated. In other words, we are assuming that the stochastic GW background will be negligible compared to the intrinsic noise in each pulsar. In general this is not likely to be a good assumption when we would expect a detection of a single GW source. Furthermore, terrestrial clock errors (Hobbs et al. 2012)

and errors in solar system ephemerides (Champion et al. 2010)[7] can also cause correlations between residuals from different pulsars, with however different angular correlation properties on the sky than are expected from GWs. The effects of omitting the correlations in the likelihood function are unknown and will be the subject of future work. Under these assumptions, the likelihood function for the full PTA can be written as

$$p(\delta t|\vec{\lambda}) = \prod_{\alpha=1}^{N_{\rm psr}} p(\delta t_\alpha|\vec{\lambda}_\alpha), \tag{5.4.1}$$

where δt_α and $\vec{\lambda}_\alpha$ and the residuals and model parameters for the αth pulsar, respectively and $\vec{\lambda}$ is the full CW parameter vector including pulsar distances for all pulsars. In cases where we fix the noise values, we can write the log-likelihood ratio of a model with a single continuous GW to a model with just noise as

$$\ln \Lambda = \sum_{\alpha}^{N_{\rm psr}} \left[\left(\delta t_\alpha | s(\vec{\lambda}_\alpha) \right) - \frac{1}{2} \left(s(\vec{\lambda}_\alpha) | s(\vec{\lambda}_\alpha) \right) \right], \tag{5.4.2}$$

where the inner product between two time-series x and y is

$$(x|y) = x^T G (G^T C G)^{-1} G^T y. \tag{5.4.3}$$

In the remainder of the chapter we will refer to the signal-to-noise ratio in the following form

$$\rho = \sqrt{2\langle \ln \Lambda \rangle} = \left(\sum_{\alpha}^{N_{\rm psr}} \left(s(\vec{\lambda}_\alpha) | s(\vec{\lambda}_\alpha) \right) \right)^{1/2}, \tag{5.4.4}$$

where the angle brackets denote the expectation value over many noise realizations.

5.4.2 Noise Model

In section 2.5 we have derived a general parameterized noise model for PTA data analysis. Here we usea slightly simplified variant of that noise model. Previous Bayesian analysis schemes (van Haasteren et al. 2009a; van Haasteren & Levin 2010; van Haasteren et al. 2011; Ellis et al. 2012c; van Haasteren & Levin 2013; van Haasteren 2013; Ellis et al. 2013; Ellis 2013) have used a power-law red noise model and an EFAC (constant multiplier on the TOA uncertainties) and EQUAD (additional Gaussian white noise added

[7] Note that current uncertainties in the ephemerides are small enough that they will likely not pose any problems for GW analyses.

in quadrature to EFAC noise) parameters to describe the white noise, with a covariance matrix of the form

$$C = E^2 W + \mathcal{Q}^2 \mathbb{I} + C^{\rm red}(A_{\rm red}, \gamma_{\rm red}), \tag{5.4.5}$$

where E is the EFAC parameter, $W = \mathrm{diag}\{\sigma_i^2\}$, with σ_i the errorbar on the ith TOA, $\mathcal{Q}$ is the EQUAD parameter and $C_{\rm red}$ is an analytic expression of the red noise amplitude $A_{\rm red}$ and spectral index $\gamma_{\rm red}$. It is worth noting that we use no EFAC or EQUAD parameters in our pulsar timing model fit but instead include them directly in our noise model. The EFAC is simply a parameter that quantifies any additional uncertainty in the TOA uncertainties and the EQUAD parameter quantifies any additional white noise that is not related to the formal TOA uncertainties. In principle, a different EFAC value should be used for each pulsar timing backend as this parameter is related to intrinsic receiver noise; however, in this 5-year NANOGrav dataset, only one backend per telescope was used[8]. Therefore, we are justified in only using one EFAC parameter per pulsar. This noise model is quite general and works well for many pulsars; however, the size of the matrices is quite large (on the order of $10^3 \times 10^3$) and inversion is a large bottleneck in the analysis pipelines. Furthermore, current NANOGrav observing schemes produce large sets of multifrequency observations that are essentially simultaneous. One may be tempted to simply perform a weighted average of the TOAs and work with the new reduced datasets but in the Bayesian scheme we must marginalize over the timing model parameters analytically and it is unclear how to carry out this process for epoch-averaged TOAs. Because of this, we have developed a framework to essentially work backward from the marginal likelihood to derive a nearly exact averaging scheme. First we re-write our noise covariance matrix

$$C = N + U\tilde{\mathcal{C}}U^T, \tag{5.4.6}$$

where $\tilde{\mathcal{C}}$ is a $q \times q$ reduced covariance matrix with q the number of epochs[9] in our dataset, N is a white noise covariance matrix of the EFAC and EQUAD terms, and U is the "exploder" matrix that maps epochs (columns) to the full set of TOAs (rows). If we now

[8]PSR J1713+0747 is observed at both Arecibo and GBT; however, we find that there is very little difference in the measured EFAC parameters for the two telescopes.

[9]Here we have defined an epoch to be one day.

make use of this new formalism, the likelihood function is then

$$p(\delta t|\vec{\phi},\vec{\lambda}) = \frac{\exp\left[-\frac{1}{2}\left((\delta t-s)^T\tilde{N}^{-1}(\delta t-s) - d^T\Sigma^{-1}d\right)\right]}{\sqrt{(2\pi)^{n-m}\det(\tilde{\mathcal{C}})\det(G^TNG)\det(\Sigma)}}, \tag{5.4.7}$$

where we have used the Woodbury Lemma[10] to compute the inverse and determinant of C, $\tilde{N}^{-1} = G\left(G^TNG\right)^{-1}G^T$, $d = U^T\tilde{N}^{-1}(\delta t - s)$, and $\Sigma = \left(\tilde{\mathcal{C}}^{-1} + U^T\tilde{N}^{-1}U\right)$. Note that d here are essentially daily averaged residuals. For NANOGrav datasets the number of epochs per pulsar is on the order of 30–100, while the total number of TOAs per pulsar is on the order of 10^3, thus the inversions (here N is diagonal and $\tilde{N}^{-1}$ can be pre-computed, thus the only dense matrix inversion is Σ^{-1}) required in this likelihood function scale as q^3 as opposed to n^3, resulting in computational speedups of several orders of magnitude. Furthermore, the epoch-averaged covariance matrix $\tilde{\mathcal{C}}$ can take on several forms depending on the red noise model used; however, as long as it is a slowly varying function of the TOAs (i.e., a truly red process) then this formalism is completely valid.

In order to attain further computational speedups and to gain more control over the low frequency component of our noise model we make use of the methods described in Lentati et al. (2013b), but now applied to a single pulsar instead of the full PTA. This method relies on explicitly splitting up the red and white components of the residuals, so that the residuals are now written as

$$\delta t = M\delta\boldsymbol{\xi} + n_{\rm white} + n_{\rm red} + s, \tag{5.4.8}$$

where $n_{\rm white}$ and $n_{\rm red}$ are the white and red components of the residuals, respectively. It is possible to expand the red noise piece in a Fourier series

$$n_{\rm red} = \sum_{j=1}^{N_{\rm mode}}\left[a_j\sin\left(\frac{2\pi jt}{T}\right) + b_j\cos\left(\frac{2\pi jt}{T}\right)\right] = F\mathbf{a}, \tag{5.4.9}$$

where $\mathbf{a}$ is a vector of the alternating sine and cosine amplitudes, T is the total time span of the data, and F is a $N_{\rm TOA}\times 2N_{\rm mode}$ matrix with alternating sine and cosine terms with $N_{\rm mode}$ the number of frequencies used. Now, we assume that the underlying ensemble average red noise process is wide-sense stationary and can be completely described by a

[10] $(A+DBE^T)^{-1} = A^{-1} - A^{-1}D(B^{-1}+E^TA^{-1}D)^{-1}E^TA^{-1}$ and $|A+DBE^T| = |A||B||B^{-1}+E^TA^{-1}D|$

power-spectrum. Then, by orthogonality, the Fourier coefficients $\mathbf{a}$ will be diagonal with components

$$\varphi_{ij} = \langle \mathbf{a}\mathbf{a}^T \rangle_{ij} = \mathrm{diag}(\{\varphi_i\}), \tag{5.4.10}$$

where the elements of φ, denoted $\{\varphi_i\}$ are the coefficients of the theoretical power spectrum of the red noise process in the residuals. If the red noise process is wide-sense stationary, then this relation is always true irrespective of the sampling as all information about the uneven sampling here comes from the Fourier design matrix F. Thus, we can write the covariance and epoch-averaged covariance matrices, respectively, as

$$C = N + F\varphi F^T \tag{5.4.11}$$

$$\tilde{\mathcal{C}} = \tilde{F}\varphi\tilde{F}^T, \tag{5.4.12}$$

where $\tilde{F}$ is a $q \times N_{\mathrm{mode}}$ matrix and is constructed in the same manner as F but the epoch-averaged TOAs are used as opposed to the full set of TOAs. As is done in Lentati et al. (2013b), it is possible to treat each diagonal element of φ as a free parameter; however, for this work we choose to parameterize it by a power-law

$$\varphi_i = \frac{1}{T}\frac{A_{\mathrm{red}}^2}{12\pi^2}\left(\frac{f_i}{f_{\mathrm{yr}}}\right)^{3-\gamma_{\mathrm{red}}} f_i^{-3}, \tag{5.4.13}$$

where f_i is the ith Fourier frequency assuming Nyquist sampling. In general, any Fourier based method with finite length datasets and especially with irregular sampling will suffer from spectral leakage whereby power from the lowest frequencies will leak into the higher frequencies. In effect, this makes Fourier based methods sensitive to the low-frequency cutoff. However, it was shown in van Haasteren & Levin (2013) that by including the effects of the timing model (specifically the quadratic spin-down in this case) in our analysis acts as a window function that fully removes any sensitivity to the low-frequency cutoff, thereby also removing any spectral leakage. We have done extensive simulations to test this notion and have found no evidence for spectral leakage and no bias in red noise parameter estimation and waveform reconstruction.

In the course of our single pulsar noise analysis (Ellis et al. 2014) we found that the addition of an extra white noise parameter was needed to accurately describe the data. This new white noise term incorporates a correlation among frequency channels

(within a given epoch) while still remaining independent of other epochs. In other words, this white noise term accounts for epoch-to-epoch fluctuations as opposed to fluctuations within an epoch. We defer to another paper the inclusion of pulse-jitter noise from pulsar magnetospheric activity but point out that our inferred extra term may be the same as jitter noise known to be present in all well-studied pulsars (Cordes & Shannon 2010). This parameter is quite easy to incorporate as it is simply an EQUAD like parameter in the epoch-averaged sense, that is

$$J = U\tilde{J}U^T = \mathscr{J}^2 U\mathbb{I}_q U^T, \tag{5.4.14}$$

where $\mathscr{J}$ is our frequency correlated EQUAD parameter and $\mathbb{I}_q$ is the identity matrix in the epoch-averaged space. With this, we have our final noise model with a total covariance matrix of

$$C = N + U\left(\tilde{F}\varphi\tilde{F}^T + \mathscr{J}^2\mathbb{I}_q\right)U^T, \tag{5.4.15}$$

and noise parameter vector

$$\vec{\phi} = \{E, \mathcal{Q}, \mathscr{J}, A_{\rm red}, \gamma_{\rm red}\}. \tag{5.4.16}$$

Throughout the remainder of the chapter, this noise model is always used for all pulsars.

5.4.3 $\mathcal{F}_p$-Statistic

The $\mathcal{F}_p$-statistic was first derived in Ellis et al. (2012c) (hereafter ESC12) as a "total-power" frequentist detection statistic. First we define the following harmonic basis functions:

$$B^1_\alpha(t) = \frac{1}{\omega_0^{1/3}}\sin(2\omega_0 t) \tag{5.4.17}$$

$$B^2_\alpha(t) = \frac{1}{\omega_0^{1/3}}\cos(2\omega_0 t), \tag{5.4.18}$$

where, again, ω_0 is the orbital angular frequency of the SMBHB. Following ESC12, the $\mathcal{F}_p$-statistic is written as

$$2\mathcal{F}_p = \sum_{\alpha=1}^{M} P^i_\alpha Q^\alpha_{ij} P^j_\alpha, \tag{5.4.19}$$

where we have assumed Einstein Summation notation over latin indices, $P^i_\alpha = (\delta t|B^i_\alpha(t))$, $Q^\alpha_{ij} = (B^\alpha_i|B^\alpha_j)$ and the formal sum is over all pulsars in the array. An intuitive way to

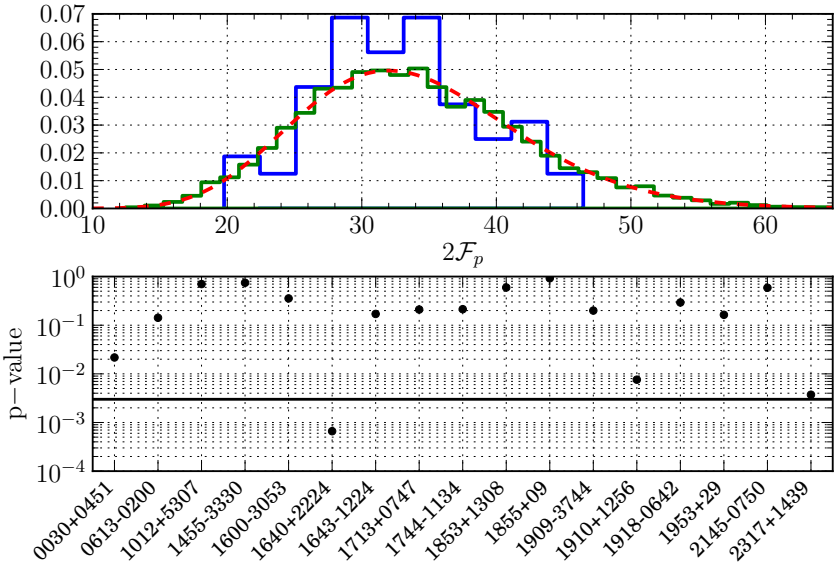


Figure 16 : Histogram of $\mathcal{F}_p$-statistic values (top panel) across all independent frequencies (black(blue) histogram) and for 100,000 realizations of simulated data with noise parameters measured from the real data (gray(green) histogram). The red dashed curve is the probability distribution function for a chi-squared distribution with 34 (i.e., $2N_{\rm psr}$) degrees of freedom. The lower panel shows the p-value from a KS test comparing the $\mathcal{F}_p$-statistic for each pulsar to a chi-squared distribution with 2 degrees of freedom. The solid line represents the 3-sigma threshold for the p-value.

think of this statistic is a weighted (by the noise power spectral density) sum of the power spectrum of the residual data done in the time domain by making use of a harmonic time domain basis. It was shown in ESC12 that $2\mathcal{F}_p$ follows a chi-squared distribution with $2N_{\rm psr}$ degrees of freedom and non-centrality parameter ρ^2 such that

$$\langle 2\mathcal{F}_p \rangle = 2N_{\rm psr} + \rho^2. \tag{5.4.20}$$

Figure 16 shows the distribution of the $\mathcal{F}_p$-statistic (top panel) for both real and simulated data as well as as a p-value test (bottom panel) for each pulsar, where we compare the single-pulsar $\mathcal{F}_p$ distribution to the expected chi-squared distribution. To compute the $\mathcal{F}_p$-statistic, we have used the maximum a-posteriori noise values obtained in a previous single-pulsar noise analysis to construct the noise covariance matrix. Since we do not have independent realizations of our data, we compute the $\mathcal{F}_p$-statistic for each independent[11] frequency bin and then construct a histogram of the results. If our noise model is a good description of the true noise in our data and there is no GW present in the data

[11]Note that the frequencies are not completely independent since our data are irregularly sampled. The frequency bins were chosen here assuming a cadence of two observing sessions per month.

then this distribution should follow the correct chi-squared distribution. We see from Figure 16 that the $\mathcal{F}_p$-statistic values do indeed follow a chi-squared distribution with $2N_{\rm psr}$ degrees of freedom. The black(blue) curve in the top panel of Figure 16 shows the aforementioned histogram along with the chi-squared distribution in the dashed gray(red) line. The p-value that results from a Kolmogorov-Smirnov (KS) test comparing the $2\mathcal{F}_p$ and chi-squared (with 34 degrees of freedom) distributions is 0.33 showing good agreement between our data and the expected chi-squared distribution. As a cross-check, we have also simulated 100,000 datasets with the measured noise parameters and have evaluated the $\mathcal{F}_p$-statistic for each. This distribution is plotted as a gray(green) histogram in the figure and it is obvious that this distribution follows a chi-squared distribution with 34 degrees of freedom nearly perfectly. We have also performed a similar test but for each pulsar separately. In the lower panel of Figure 16 we carry out the same KS-test mentioned above but now compute the $\mathcal{F}_p$-statistic values for each pulsar individually and then compare to a chi-squared with 2 degrees of freedom. The solid line corresponds to the p-value at which we should reject the null hypothesis that the two distributions are the same with 99.7% confidence. We see that with the exception of one pulsar, J1640+2224, all others lie above this threshold value. This indicates that our noise model for all pulsars except J1640+2224 provide a good description of the true noise in the dataset. Better noise models for this pulsar are currently being explored (Ellis et al. 2014) but since our full 17-pulsar $\mathcal{F}$-statistic distribution is totally consistent with the expected chi-squared distribution we just use the standard noise model described in Section 5.4.2.

For the detection problem, we are interested in the false-alarm-probability (FAP), that is, the probability that a measured value $\mathcal{F}_p$ exceeds a given threshold $\mathcal{F}_{p,0}$ when no signal is present. From ESC12, the probability distribution of $\mathcal{F}_p$ when the signal is absent is

$$p_0(\mathcal{F}_p) = \frac{\mathcal{F}_p^{n/2-1}}{(n/2-1)!}\exp(-\mathcal{F}_p), \tag{5.4.21}$$

where n is the number of degrees of freedom ($2N_{\rm psr}$ in this case). The corresponding FAP is then written as

$$P_F(\mathcal{F}_{p,0}) = \int_{\mathcal{F}_{p,0}}^{\infty} p_0(\mathcal{F}_p)d\mathcal{F}_p = \exp(-\mathcal{F}_{p,0})\sum_{k=0}^{n/2-1}\frac{\mathcal{F}_{p,0}^k}{k!}. \tag{5.4.22}$$

In a search over GW frequencies (the only free parameter in the $\mathcal{F}_p$-statistic) we will incur a trials factor such that the resulting FAP for the search is

$$P_F^T(\mathcal{F}_{p,0}) = 1 - [1 - P_F(\mathcal{F}_{p,0})]^{N_f}\,, \tag{5.4.23}$$

where N_f is the number of independent frequencies. For this work we place our detection threshold on $\mathcal{F}_p$ such that the corresponding FAP is less than 10^{-4}. The results of performing this search on the 5-year NANOGrav dataset will be presented in the next section.

5.4.4 Bayesian Method

The Bayesian search pipeline in this work is very similar to that of Ellis (2013) (hereafter E13). Here we use an MPI enabled Parallel-Tempered Markov Chain Monte-Carlo (PTMCMC) sampler[12]. In this work we use two "modes" of operation for the Bayesian search. The first is the most general in which we evaluate the full likelihood function of Eq. (5.4.7) and allow both the GW parameters, $\vec{\lambda}$, and the noise model parameters, $\vec{\phi}$ to vary simultaneously. In principle, this is the more desirable setup as it allows the uncertainty in our noise model to propagate into the measured GW parameters and also accounts for any correlations between the noise and GW parameters. This mode does require significantly more computational power as the number of search parameters in the MCMC is quite large. The total parameter space consists of 8 GW parameters, $N_{\rm psr}$ pulsar distances, and $5 \times N_{\rm psr}$ noise parameters; this comes to 110 parameters for the full 17-pulsar array.

The second mode is when we fix the noise parameters to their maximum a-posteriori values obtained from a previous single pulsar analysis. All previous GW searches for single sources have been performed in this manner (Jenet et al. 2004; Yardley et al. 2010; Babak & Sesana 2012; Ellis et al. 2012c; Petiteau et al. 2013; Ellis 2013) which is justified *if* the noise model only contains white noise and the GW signal present in any single dataset is weak. If the noise model contains only white noise, there will be little to no correlation between the GW parameters and the noise parameters, and if the

[12] `https://github.com/jellis18/PAL`

signal is weak then it will not affect the single pulsar noise analysis. However, there is some evidence of red noise in our pulsars and because of the highly varying noise levels among pulsars, it is likely that a *detectable* source would be seen in the best timed pulsars individually. Therefore, this type of Bayesian analysis is not robust and could possibly lead to biased results; nonetheless, we will carry out this mode for comparison purposes in this study. Note that we will have the same problem with the $\mathcal{F}_p$-statistic. Possible methods to ameliorate this problem in fixed-noise searches are being explored and will be the subject of a future paper.

Priors

In a Bayesian analysis, especially when using parallel tempering and thermodynamic integration, it is very important to choose reasonable priors so that we are not exploring areas of parameter space that have been ruled out by previous experiments. We choose isotropic priors on all angular parameters and uniform priors in the log of the chirp mass with $\mathcal{M} \in [10^8, 10^{10}]\,\mathrm{M}_\odot$, luminosity distance with $d_L \in [1, 10^4]$ Mpc, and frequency of the GW with $f_{\rm gw} \in [6\times10^{-9}, 4\times10^{-7}]$ Hz. We impose an additional condition on the average strain amplitude such that $h_0(\mathcal{M}, d_L, f_{\rm gw}) \leq h_{\rm ref}(f_{\rm gw}/f_0)^{2/3}$, where $h_{\rm ref} = 1\times10^{-13}$ and $f_0 = 10^{-8}$ Hz. This value is chosen so that the maximum strain is well above the level of detection. Essentially this is a cheap way to impose a correlated prior on chirp mass, luminosity distance, and GW frequency. The normalization is computed through Monte Carlo integration. For the pulsar distance prior we use the current electromagnetic (EM) measurements either from timing parallax or Very Long Baseline Interferometry (VLBI) corresponding to the best measured values taken from Verbiest et al. (2012) (10 pulsars) if available, otherwise, we use the values from the Australia National Telescope Facility (ATNF) pulsar catalog (Manchester et al. 2005)[1] which have distances based on dispersion measure and the NE2001 Galactic electron-density model (Cordes & Lazio 2002, 2003). For pulsars without parallax distances we assume a 20% uncertainty on the distance.

[1] `http://www.atnf.csiro.au/people/pulsar/psrcat/`

Using this information, we write the distance prior as follows

$$p(\vec{L}) = \prod_{\alpha=1}^{N_{\rm psr}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left(-\frac{(L_\alpha - L_\alpha^{\rm EM})^2}{2\sigma_\alpha^2}\right), \tag{5.4.24}$$

where $L_\alpha^{\rm EM}$ is the best measured distance for the αth pulsar and σ_α is the 1-sigma uncertainty on that distance measurement. In principle it would be more correct to use a Gaussian prior for the parallax, which is proportional to L^{-1}. If the variance on the parallax is quite large then the corresponding prior on distance will differ significantly, namely it will have a long tail towards higher distances. However, for the pulsars used in this analysis, the distance uncertainty is small enough that the two prior distributions are effectively the same and we are safe in using a Gaussian prior on the pulsar distance itself; however, for future analyses we will move to Gaussian priors in L^{-1}. As was noted in E13, constructing an efficient jump proposal for the pulsar distance is quite difficult. See Appendix 5.B for the implementation used in this work.

For our noise parameters, we use priors that are uniform in the EFAC in the range $[0.5, 5]$, uniform in the log of the EQUAD with EQUAD $\in [10^{-9}, 10^{-5}]$ s, uniform in the log of the jitter value with the same range as the EQUAD, uniform in the log of the red noise amplitude with $A_{\rm red} \in [10^{-18}, 10^{-11}]$, where the amplitude is in GW units, and uniform in the red noise spectral index with $\gamma_{\rm red} \in [1, 7]$. We impose a further prior on the red noise such that the variance $\sigma_{\rm red}^2$ is less than the *unweighted* standard deviation of the pulsar timing residuals, where

$$\sigma_{\rm red} = \int_{1/T}^{\infty} df P(f) = 2.05 \frac{1}{\sqrt{\gamma_{\rm red} - 1}} \left(\frac{A_{\rm red}}{10^{-15}}\right) \left(\frac{T}{1\,{\rm yr}}\right)^{\frac{\gamma_{\rm red}-1}{2}} \text{ ns}, \tag{5.4.25}$$

with T the total observation time and $P(f)$ the power spectrum of the red noise. This prior essentially restricts the model from considering red noise dominated residuals, which is a very good approximation (Perrodin et al. 2013a; Ellis et al. 2014). This prior is chosen because it leads to much more computationally efficient runs by allowing us to run fewer high temperature chains in the Thermodynamic Integration scheme (See section 2.7.1 for more details). In principle this red noise prior is illegal in the sense that it uses the data (i.e., the variance of the residuals); this prior restricts access to an area of parameter space that is not supported by the likelihood function. That is, by omitting this area

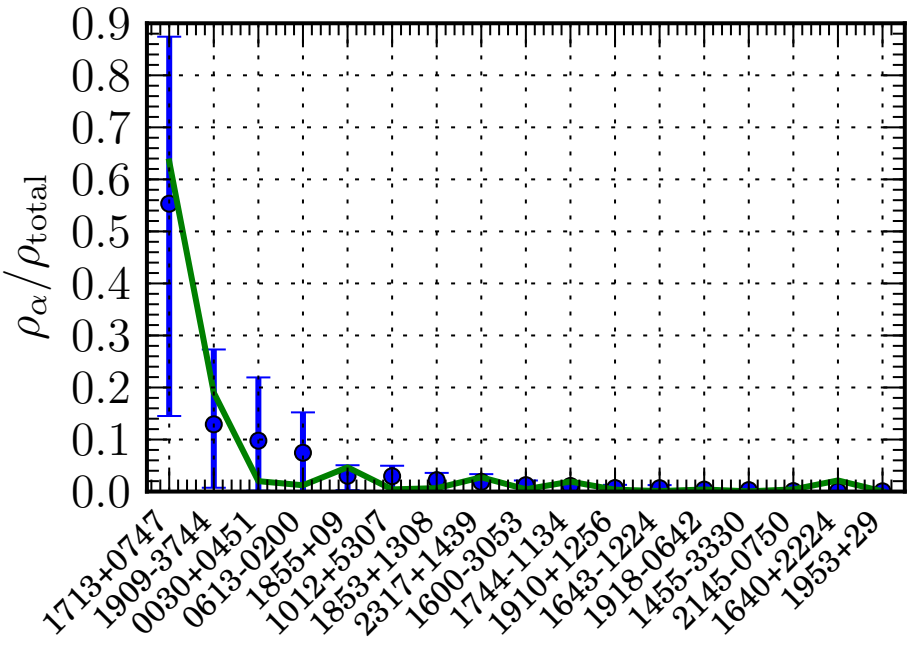


Figure 17 : Fraction of SNR that each pulsar contributes (black(blue) points). We see that PSR J1713+0747 dominates the total SNR. The gray(green) curve is a simple $1/\sigma_\alpha$ scaling which matches the measured SNR values quite well showing that the overall variance of the noise for each pulsar is the dominating factor in determining the overall SNR.

of parameter space the evidence calculation for each model, $\mathcal{H}_1$ and $\mathcal{H}_0$, will be biased slightly low but, the likelihood function evaluated at this area of parameter space is essentially zero, and this slight bias will be negligible.

5.5 Results

In this section we report the results of our frequentist and Bayesian searches, provide verification of the pipeline on injected signals and report on several upper limits.

5.5.1 Verification

First, it is interesting to determine how much each pulsar in the 17-pulsar array will contribute to the overall SNR (signal-to-noise ratio) when a GW is present. In Figure 17 we plot the fraction $\rho_\alpha/\rho_{\rm total}$, where ρ_α is the single pulsar SNR, for each pulsar in the array. To compute this fraction we simulate 5000 SNR = 10 GW realizations (with parameters drawn from isotropic distributions in all angles and distributions uniform in

the log of chirp mass and frequency) and calculate the single pulsar and total PTA SNR from Eq. (5.4.4). The black(blue) points in the plot show the mean and standard deviation of the aforementioned ratio for each pulsar and the gray(green) curve is a simple naive scaling of $1/\sigma_\alpha^2$, where σ_α is the weighted RMS of the αth pulsar's TOA uncertainties. It is obvious that J1713+0747 contributes more than 55% of the SNR on average, and PSRs J1909−3744, J0030+0451, and J0613−0200 contribute $\sim 10\%$ on average. As we see from the gray(green) curve, this is very consistent with the overall scaling with the inverse of the variance of the noise; however, PSRs J0030+0451 and J0613−0200 carry a higher percentage because they are located opposite to the bulk of other pulsars on the sky, and therefore will contribute more to the SNR for GWs coming from that side of the sky due to the antenna pattern response. This calculation does not mean that we advocate only timing the pulsars with the highest timing precision. Although many of the lower timing precision pulsars do not help with continuous GW detection or parameter estimation, they are *essential* for detection and parameter estimation of a stochastic GW background (see e.g., Siemens et al. 2013).

The fact that one pulsar dominates the total SNR means that it will be harder to make a confident GW detection as we require the same GW signal (with quadrupolar correlations) to be present in all pulsars. In other words, if the GW is only "seen" in one or two pulsars then it is hard to distinguish it from some other effect due to the pulsar timing model, ISM effects or some other systematic effect. This also implies the need to run a Bayesian analysis where both the noise and GW parameters are allowed to vary simultaneously. This does not mean that a continuous GW would not be a valid interpretation of a loud sinusoidal signal in one pulsar, only that statistically we do not have enough information to confidently claim a detection. Furthermore, if we did have a loud *detectable* signal, parameter estimation would be quite poor with the current NANOGrav PTA as there would be large degeneracies in the sky location (due to the small effective number of detector baselines), making sky localization and binary orientation estimates very poor. However, NANOGrav is currently timing 43 pulsars with microsecond or better precision. Also, new ultra-wideband receivers (DuPlain et al. 2008) have increased timing precision by a factor of $\sim$1.7 for many of the pulsars in this

5-year data set. More pulsars and better timing precision could help ameliorate some of the limitations we have with the 5-year data set.

Despite the potential limitations discussed above, we verify the efficacy of our pipeline by running both the frequentist and Bayesian pipeline on a synthetic dataset with an injected GW source. To create the synthetic dataset we first compute the residuals of our 17 NANOGrav pulsars using the TEMPO2 (Hobbs et al. 2006) package. Next we subtract these residuals from the site arrival times, thereby producing a new set of arrival times that match our timing model perfectly. To each set of idealized TOAs we then add a Gaussian noise process with the same characteristics as those measured in the real data, and a GW signal using the fully evolving signal model. We then use these new TOAs to produce a set of synthetic residuals. For this simulation we have chosen to inject a signal with SNR 10 and parameters $\vec{\lambda} = \{\theta = 2.07, \varphi = 5.4, f_{\rm gw} = 4 \times 10^{-8}\,{\rm Hz}, \mathcal{M} = 5 \times 10^{8}{\rm M}_{\odot}, d_L = 1.0\,{\rm Mpc}, \psi = 0.78, \iota = 0.26, \Phi_0 = 0.53\}$.

The $\mathcal{F}_p$-statistic pipeline was run on this synthetic dataset. Since we are treating this injection as if it were a true blind search, we must first run a single-pulsar noise analysis to determine the maximum a posteriori noise parameters; however, since a strong continuous GW and red noise will be covariant we have included a single frequency sinusoid as part of our noise analysis for each pulsar. This is implemented by simply adding a free amplitude and frequency parameter to the noise model discussed above. While this may appear to be special treatment for the injected signal, we have run the same noise model on the real data and find no evidence for any sinusoidal features. After we have obtained the maximum a-posteriori noise parameters (not including the sinusoid parameters), we use these values to construct the noise covariance matrix for use in the $\mathcal{F}_p$-statistic as well as the fixed-noise Bayesian search. By performing the noise search with an included sinusoid but not including it in our noise covariance matrix in the subsequent GW analysis we are sidestepping the problem of the GW being absorbed into red noise parameters.

We have carried out this analysis and the results are shown in Figure 18 where we plot $\mathcal{F}_p$ vs. GW frequency when using the measured noise values (black) and the true injected noise values (grey). The vertical dashed line indicates the injected frequency and the horizontal dashed line represents our detection threshold corresponding to a FAP of 10^{-4}.

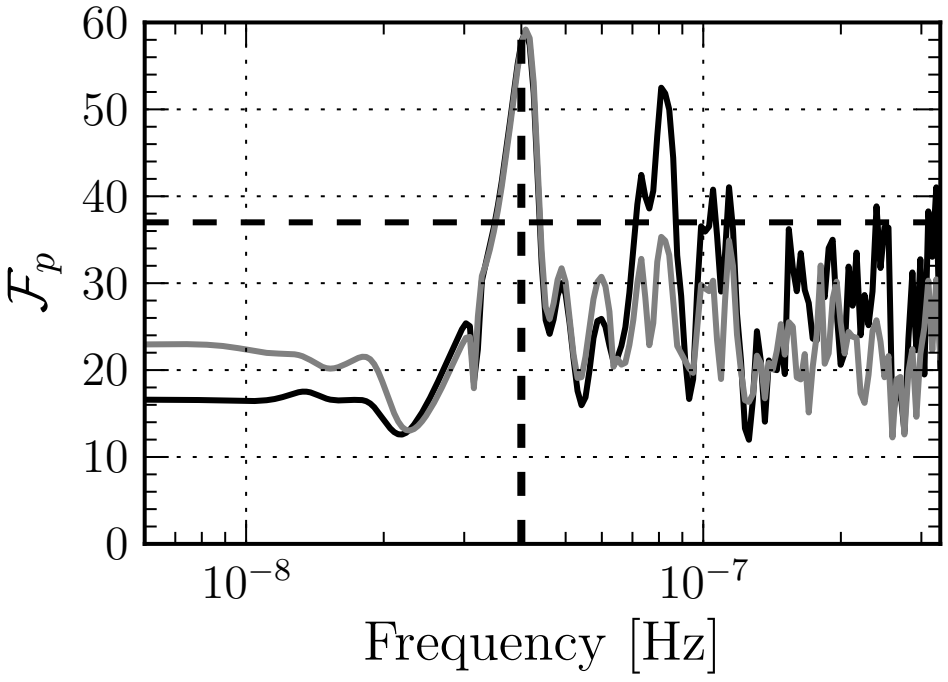


Figure 18 : $\mathcal{F}_p$-statistic evaluated over the frequency range $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ Hz. The horizontal dashed line corresponds to our detection threshold of $FAP = 10^{-4}$ and the vertical dashed line denotes the injected frequency. The black and gray curves are the $\mathcal{F}_p$-statistic values when using the measured and true noise parameters, respectively. See text for more details.

To compute the total FAP over all frequencies we make use of Eq. (5.4.23) and choose $N_f = 324$, resulting in a total FAP of 1.6×10^{-8} which is a decisive detection. The number of independent frequencies is difficult to calculate when we are using many datasets with very irregular sampling. In this work we have chosen $N_f = 324$ as this corresponds to $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ and $\Delta f_{\rm gw} = 10^{-9}$ Hz. The upper limit on frequency was chosen because our approximate observing cadence is (2.5 weeks)$^{-1}$ and the frequency spacing was chosen by imposing the condition that the autocorrelation function of $\mathcal{F}_p$ when no signal is present drops to half of its maximum value at that frequency lag. This analysis shows us that we can indeed detect a continuous GW if it is present in our data by conducting a fully blind search; however, we also see that our results will not be conclusive as there are several frequencies at which the FAP is above our threshold value. From comparison with true-noise case, we see that the uncertainty (and residual correlations between GW and noise parameters) in the noise parameters can lead to confusing results. This again, is mostly due to the fact that our sensitivity is dominated by a small number of pulsars. Because of this, we caution against using a fixed-noise method to make final detection statements but instead advocate these methods as a first round in a suite of analyses.

Both Bayesian pipelines (with and without varying noise parameters) were run on this synthetic dataset. For both runs we have used PTMCMC and thermodynamic integration as discussed in section 2.7.1. Due to the large parameter spaces when using the full GW and noise model, we have chosen to use only the pulsars that contribute more than 1% to the injected SNR, resulting in 6 pulsars, J1713+0747, J1909−3744, B1855+09, J0030+0451, J0613−0200, and J1012+5307. Here we use the same noise parameters as mentioned above for our fixed-noise search. Even though these estimates are different from the true noise parameters, we nonetheless achieve a log-Bayes factor of 27.4 for the fixed-noise search (a log-Bayes factor greater than 5 is considered decisive evidence). However, as we mentioned earlier, we should not totally trust this level of evidence as it does not fully incorporate our uncertainty in the noise model. When we run an analysis where we allow the noise and GW parameters to vary simultaneously we only achieve a log-Bayes factor of 5.35. While still decisive, this search is much less sensitive to the

GW; nonetheless, this search is the most robust and will be the real test as to whether or not one can trust a real GW detection candidate. Of course these results could change depending on the noise realization or GW parameter combinations. A more detailed study of this is warranted but beyond the scope of this chapter. Nonetheless, the large spread of overall noise levels in modern PTAs will most likely make the confident detection of a continuous wave GW very difficult.

5.5.2 Search Results

First we will discuss the results of the $\mathcal{F}_p$-statistic search on the real 17-pulsar NANOGrav data. To carry out the analysis we have computed $\mathcal{F}_p$ for many frequencies with $f_{\rm gw} \in [1/T, 3.3\times10^{-7}]$ Hz. These frequencies were chosen based on the fact that the approximate cadence is 2.5 weeks^{-1}. The results of this search are shown in Figure 19 where the solid black line is the value $\mathcal{F}_p$ at each frequency, the dotted, dash-dotted, and dashed lines are the value of $\mathcal{F}_p$ corresponding to a 1.0%, 0.5%, and 0.1% FAP, respectively, where these values are calculated from Eq. (5.4.22). Furthermore if we maximize $\mathcal{F}_p$ over frequencies then the total FAP, accounting for the trials factor N_f is very nearly 1, indicating that we should fail to reject the null hypothesis (no visible GW signal) with very high confidence.

We will now briefly discuss the results of both Bayesian searches. To carry out this analysis we have run our PTMCMC and computed the Bayes factors for each case. In the first case we allow the noise parameters and GW parameters to vary and explicitly compute the Bayesian evidence via thermodynamic integration for a model with a GW and noise and a model with just noise. In the second case, we fix the noise parameters to the maximum a-posteriori obtained from single pulsar analyses and only allow the GW parameters to vary. As mentioned above, the second case is not reliable since there is likely to be correlations between the GW and noise parameters; however, we give the results of both searches for completeness. As above, in the case of a true continuous GW signal we can get very different results from a fixed-noise versus a varying noise search. However, in our case the log-Bayes factor for searches with and without varying noise parameters is -0.55 and -0.1, respectively, both indicating that there is no evidence for a continuous GW and a model consisting of noise is preferred. We further note that this

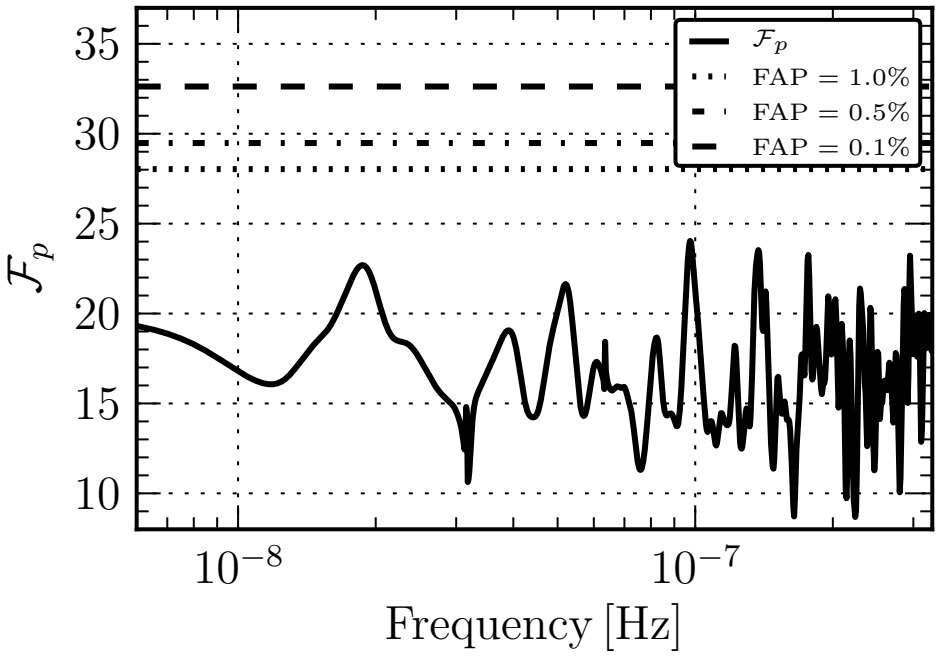


Figure 19 : $\mathcal{F}_p$-statistic evaluated over the frequency range $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ Hz. These frequencies were chosen based on the fact that the approximate cadence is 2.5 weeks^{-1}. The dashed, dash-dotted, and dotted lines represent the value of $\mathcal{F}_p$ that gives a FAP of 0.1%, 0.5%, and 1%, respectively. Here we note that there is no evidence for a detection and the data are consistent with the null hypothesis.

is completely consistent with our frequentist analysis.

5.5.3 Upper Limits

In this section we will outline the procedures used to compute both the frequentist and Bayesian upper limits on the strain amplitude, h_0. First we wish to state that an x% upper limit on the strain amplitude does not mean that we would have detected a signal with that amplitude with x% confidence, it simply means that the *true* value of the amplitude is less than the upper limit with x% probability. In the following sections we will discuss the mathematics of upper limit computation in the frequentist and Bayesian frameworks, and then we will lay out our computational procedure.

Frequentist Approach

From a frequentist viewpoint, the data are random while the signal parameters are fixed but unknown (i.e., we construct probability distributions for the data, or rather some

function of the data, given a set of signal parameters), whereas in the Bayesian framework the data are fixed and the signal parameters are uncertain (i.e., we construct probability distributions of the signal parameters given a dataset). From the above statement it then follows that frequentist upper limits are derived from integrating the probability distribution of some statistic of the data (the $\mathcal{F}_p$-statistic in this case) at a fixed value of the parameter of interest, and Bayesian upper limits are derived from integrating the probability distribution of the parameter of interest for the given data set.

More formally, the probability distribution of the $\mathcal{F}_p$-statistic given a value of the strain amplitude h_0 is

$$p(\mathcal{F}_p|h_0) = \int p(\mathcal{F}_p|h_0, \tilde{\lambda}, \mathbf{n}) p(\tilde{\lambda}) p(\mathbf{n})\, d\tilde{\lambda}\, d\mathbf{n}, \tag{5.5.1}$$

where $\tilde{\lambda} = \{\theta, \varphi, f_{\rm gw}, \mathcal{M}, \iota, \psi, \Phi_0\}$ is a reduced parameter space vector, $p(\tilde{\lambda})$ is the sampling distribution of $\tilde{\lambda}$ (these sampling distributions are identical to the prior probability distributions in the bayesian case), $\mathbf{n}$ is a noise timeseries drawn from the distribution

$$p(\mathbf{n}) = \frac{1}{\sqrt{\det 2\pi C}} \exp\left(-\frac{1}{2}\mathbf{n}^T C^{-1} \mathbf{n}\right), \tag{5.5.2}$$

with C the covariance matrix of the noise in the pulsar timing residual timeseries, and $p(\mathcal{F}_p|h_0, \tilde{\lambda}, \mathbf{n})$ is the probability distribution function for the $\mathcal{F}_p$ statistic for given values of h_0 and $\tilde{\lambda}$ and a given noise realization $\mathbf{n}$. An upper limit on h_0 at confidence level α is then computed by solving the equation

$$\begin{aligned} \alpha &= \int_{\mathcal{F}_{p,0}}^{\infty} p(\mathcal{F}_p|h_0)\, d\mathcal{F}_p \\ &= \left\langle \frac{1}{N} \sum_{i=1}^{N} \begin{cases} 1 & \text{if } \mathcal{F}_{p,i} \geq \mathcal{F}_{p,0} \\ 0 & \text{otherwise} \end{cases} \right\rangle, \end{aligned} \tag{5.5.3}$$

for h_0, where the N observables $\mathcal{F}_{p,i}$ are drawn from the "signal distribution", $p(\mathcal{F}_p|h_0)$, and the average, $\langle\cdot\rangle$, is over that distribution. In other words, we integrate the probability distribution of the $\mathcal{F}_p$-statistic over the so called "signal space" (i.e., from the measured value $\mathcal{F}_{p,0}$ to infinity), that is, we count the number of signal realizations that gives an $\mathcal{F}_p$-statistic value larger than the one measured in the actual dataset. This integral can take on any value $\alpha \in [0, 1]$ for a given h_0; therefore, the integral must be repeated with different values of h_0 until $\alpha = 0.95$ for a 95% upper limit.

In practice, we carry out the following computational procedure:

1. Measure the value $\mathcal{F}_{p,0}$ from the real 17-pulsar NANOGrav dataset as described in Section 5.4.3.

2. Simulate a synthetic noise vector $\mathbf{n} = L\mathbf{w}$ for each pulsar, where L is the Cholesky decomposition of the noise covariance matrix C, and $\mathbf{w}$ is a unit variance, zero-mean, vector.

3. Choose strain amplitude, h_0 and construct a GW waveform $\mathbf{s}(t, h_0, \tilde{\lambda})$ for each pulsar where the parameters, $\tilde{\lambda}$ are drawn from the distribution $p(\tilde{\lambda})$.

4. Construct a new set of residuals for each pulsar $\delta t_{\rm sim} = R\left(\mathbf{n} + \mathbf{s}(t, h_0, \tilde{\lambda})\right)$, where R is the so called fitting projection matrix introduced in Demorest et al. (2013) and Ellis et al. (2013).[13]

5. Now measure the value $\mathcal{F}_{p,i}$ for the simulated dataset.

6. Repeat steps 2–5 10,000 times and measure the number of realizations that result in $\mathcal{F}_{p,i} > \mathcal{F}_{p,0}$.

7. Repeat steps 2–6 with different values of h_0 until 95% of simulations result in $\mathcal{F}_{p,i} > \mathcal{F}_{p,0}$.

In the remainder of the chapter we will choose to compute upper limits on the strain amplitude as a function of GW frequency or GW sky location at a fixed GW frequency. To facilitate such upper limits we simply fix the parameters (either GW frequency or sky location) when simulating waveforms in step 3.

Bayesian Approach

As mentioned above, in the Bayesian framework we do not rely on simulations as we treat the data as fixed and integrate the posterior pdf of the parameter of interest to

[13]We choose to create residuals with the R matrix rather than re-fitting the timing model with TEMPO2 in order to simulate many datasets quickly. We have done many tests to make sure that we get the same results using both the R matrix and using a full TEMPO2 run.

compute the upper limit. In principle, a Bayesian upper limit is much more simple and intuitive than a frequentist upper limit. To compute a Bayesian upper limit we compute an integral that is analogous to Eq. (5.5.3)

$$\begin{aligned}\alpha &= \int_0^{h_{\rm up}} dh_0 \int d\tilde{\lambda}\, d\vec{\phi}\, p(\delta t|h_0,\tilde{\lambda},\vec{\phi})p(h_0)p(\tilde{\lambda})p(\vec{\phi}) \\ &= \int_0^{h_{\rm up}} dh_0\, p(\delta t|h_0)p(h_0),\end{aligned} \tag{5.5.4}$$

where $p(\delta t|h_0,\tilde{\lambda},\vec{\phi})$ is the likelihood function, $p(h_0)$, $p(\tilde{\lambda})$, $p(\vec{\phi})$ are the prior probability distributions on h_0, $\tilde{\lambda}$, and $\vec{\phi}$, respectively, where $\vec{\phi}$ denotes the noise model parameters. In words, we simply integrate the marginalized posterior distribution of h_0 until the desired credible region corresponding to a probability of α is reached at $h = h_{\rm up}$. As in the frequentist case, we want upper limits on the strain amplitude as a function of GW frequency or sky location. In this case we simply fix the parameters and then marginalize over the others. In practice, to compute the Bayesian upper limits we carry out a separate MCMC run for fixed values of frequency and/or fixed sky locations and then compute the 95% upper limit for each. The choice of prior on h_0 is very important and can lead to very different upper limits. Such a detailed analysis of priors is beyond the scope of this work but will be addressed in a future paper. In principle, our prior distribution should come from population synthesis models (Sesana 2013b); however, since we wish our upper limits to be informed by our data and not dominated by our prior distribution we use a very conservative[14] prior that is uniform in h_0 with $h_0 \in [0, 10^{-11}]$.

Sky Averaged Strain Upper Limits

In Figure 20 we report the 95% upper limits on the strain amplitude, h_0, as a function of GW frequency computed using the methods described above for the frequentist and Bayesian pipelines. The gray(red), thick black(blue) and thin black(purple) curves are the 95% upper limits on strain amplitude computed using the $\mathcal{F}_p$-statistic, Bayesian method with fixed-noise values, and Bayesian method with varying noise values, respectively.

[14]On a logarithmic scale this prior prefers higher strain values a priori; however, it is conservative in the sense that the corresponding upper limit will not overestimate our sensitivity and the limit will not depend on the lower bound of the prior as is the case for logarithmic priors.

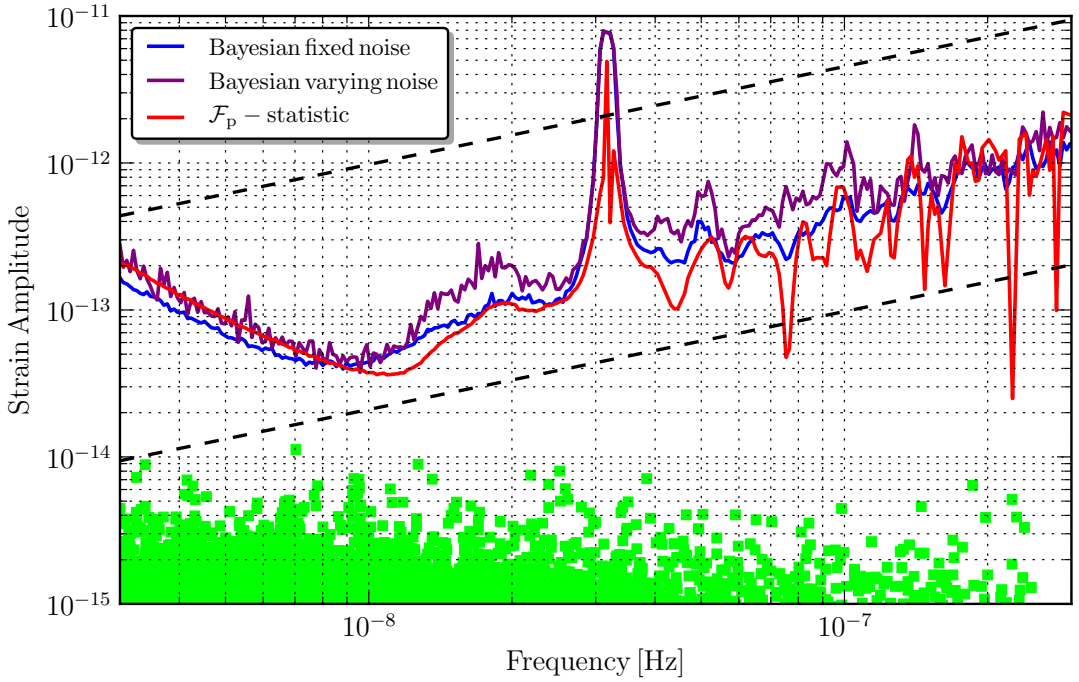


Figure 20 : Sky-averaged upper limit on the strain amplitude, h_0 as a function of GW frequency. The Bayesian upper limits are computed using a fixed-noise model (thick black(blue)) and a varying noise model (thin black(purple)) and the frequentist upper limit (gray(red)) is computed using the $\mathcal{F}_p$-statistic. The dashed curves indicate lines of constant chirp mass for a source with a distance to the Virgo cluster (16.5 Mpc) and chirp mass of $10^9 M_\odot$ (lower) and $10^{10} M_\odot$ (upper). The gray(green) squares show the strain amplitude of the loudest GW sources in 1000 monte-carlo realizations using an optimistic phenomenological model of Sesana (2013b). See text for more details.

There are several features in the plot that require explanation. First, the decrease in sensitivity at $f_{\rm gw} = 1{\rm yr}^{-1}$ and $f_{\rm gw} = 2{\rm yr}^{-1}$ is due to the sky position and parallax fitting in the timing model, respectively. The upward trend at lower frequencies is due to the quadratic spin-down model fit. The noisiness of the frequentist upper limit is due to the fact that $\mathcal{F}_p$-statistic distribution at higher frequencies is indeed quite noisy when computed using the real data, and since our upper limits compare the value measured in real data to values measured in simulated data, this noisiness is to be expected.

If we compare our results to those of Yardley et al. (2010), we see that the upper limits using the 5-year NANOGrav datasets are a factor of 2 to 3 times more constraining. The main reason for this improvement is the higher timing precision of the NANOGrav dataset as compared to the older PPTA data sets (Verbiest et al. 2009). Although the procedures for setting frequentist and Bayesian upper limits is quite different, our results are very similar. In part, this is due to the fact that we have used a uniform prior on the strain amplitude, h_0, making our Bayesian analysis very similar to a pure likelihood analysis. Since the $\mathcal{F}_p$-statistic is just the likelihood (ratio) maximized over amplitudes, we would expect a likelihood analysis to give similar results. Note that the Bayesian upper limits when varying the noise parameters are somewhat less constraining than the fixed-noise case. This is to be expected since at lower frequencies the GW amplitude and red noise amplitude are somewhat correlated and at higher frequencies the GW amplitude and jitter parameter are somewhat correlated. Both correlations will result in slightly worse upper limits on the GW amplitude when allowing the noise parameters to vary.

In Figure 20, the dashed curves indicate lines of constant chirp mass for a source with a distance to the Virgo cluster (16.5 Mpc) and chirp mass of 10^9 and 10^{10}, respectively and the gray(green) squares are the strain amplitude of the loudest GW events in 1000 Monte Carlo realizations using an optimistic phenomenological model of Sesana (2013b). The model used here produces a stochastic GW background with dimensionless strain amplitude of $\sim 2 \times 10^{-15}$, just below the current upper limits presented in Shannon et al. (2013). Astrophysically, these upper limits tell us that we can essentially rule out any source with $\mathcal{M} \geq 10^{10}{\rm M}_\odot$ at the distance to the Virgo cluster (16.5 Mpc); however, our horizon distance falls just short of the Virgo cluster for sources with $\mathcal{M} \leq 10^9{\rm M}_\odot$.

Furthermore, we see that all sources from monte-carlo realizations (gray(green) squares) have strain amplitudes below our upper limits indicating that it is very unlikely that we will see a resolvable source at the current sensitivity (consistent with our search results). It is important to note; however, that these strain amplitude upper limits are averaged over sky location and inclination angle (either through marginalization in the Bayesian case, or from Monte Carlo sampling in the frequentist case), both of which play a large part in the *overall* amplitude of the signal. Therefore, these results have the caveat that they make statements about the *average* sensitivity to such GW sources; however, it is still unlikely (i.e., probability of detection $\lesssim 50\%$) that we could detect even the loudest optimally oriented source shown in Figure 20. For face on systems (i.e., $\iota = \pi/2$) and sky location near the best timed pulsars, the overall amplitude of the GW can be ~ 5 times larger than the averaged strain amplitudes reported here.

Angular Upper Limits

In Figures 21 and 22 we report the 95% *lower* limit on the luminosity distance as a function of sky location computed using the frequentist and Bayesian techniques, respectively. We have chosen to present our results in terms of the luminosity distance instead of the strain amplitude as it is a true *physical* parameter and it gives a more intuitive feel as to what the data can constrain. To compute this lower limit we carry out the same procedure as above but we fix the frequency to $f_{\rm gw} = 10^{-8}$ Hz and compute an upper limit on the strain amplitude as a function of sky location; we can then use Eq. (5.3.2) to convert an upper limit on strain amplitude into a lower limit on luminosity distance. The values in the color bar are calculated assuming a chirp mass of $\mathcal{M} = 10^9 \mathrm{M}_\odot$ and a frequency of $f_{\rm gw} = 10^{-8}$ Hz but this can be scaled to determine the minimum luminosity distance for any chirp mass value and GW frequency. In Figures 21 and 22 the white diamonds represent the locations of the 17 NANOGrav pulsars used in the analysis and the black(white) stars are the sky locations of potential GW hotspots (Simon et al. 2013) and possible GW source candidates (Valtonen et al. 2008; Iguchi et al. 2010; Ju et al. 2013).

We will now discuss the features of this sky-dependent upper limit computed using

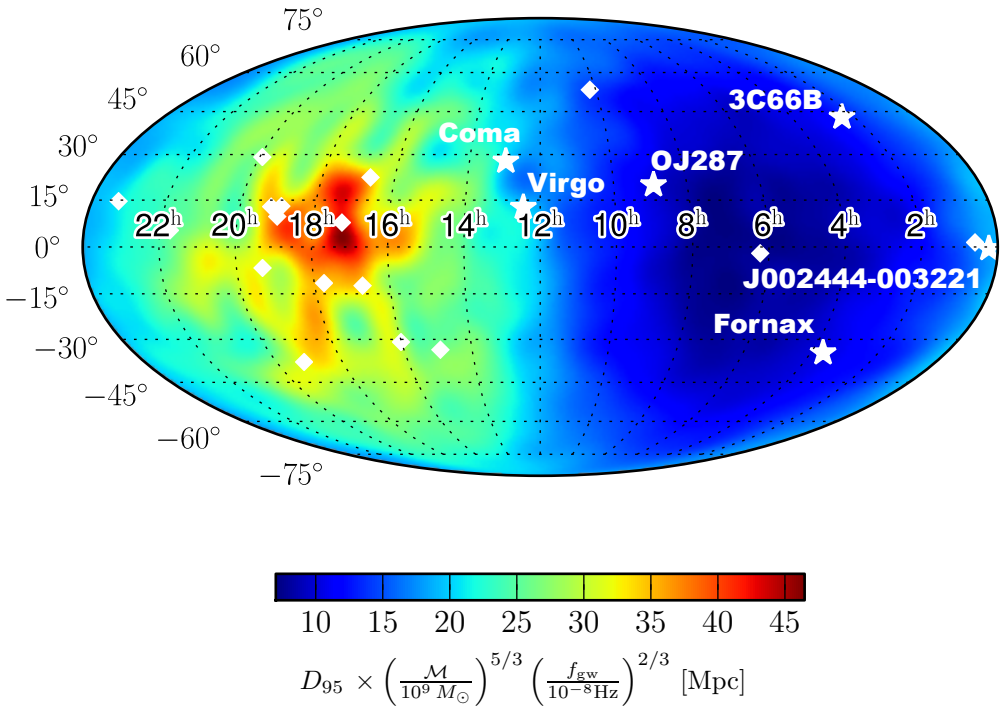


Figure 21 : 95% lower limit on the luminosity distance as a function of sky location computed using the $\mathcal{F}_p$-statistic plotted in equatorial coordinates. The values in the colorbar are calculated assuming a chirp mass of $\mathcal{M} = 10^9 \mathrm{M}_\odot$ and a GW frequency $f_{\rm gw} = 1 \times 10^{-8}$ Hz. The white diamonds denote the locations of the pulsars in the sky and the black(white) stars denote possible SMBHBs or clusters possibly containing SMBHBs. As expected from the antenna pattern functions of the pulsars, we are most sensitive to GWs from sky locations near the pulsars. The luminosity distances to the potential sources are 92.3, 1575.5, 2161.7, 16.5, 104.5, and 19 Mpc for 3C66B, OJ287, J002444−003221, Virgo Cluster, Coma Cluster, and Fornax Cluster, respectively.

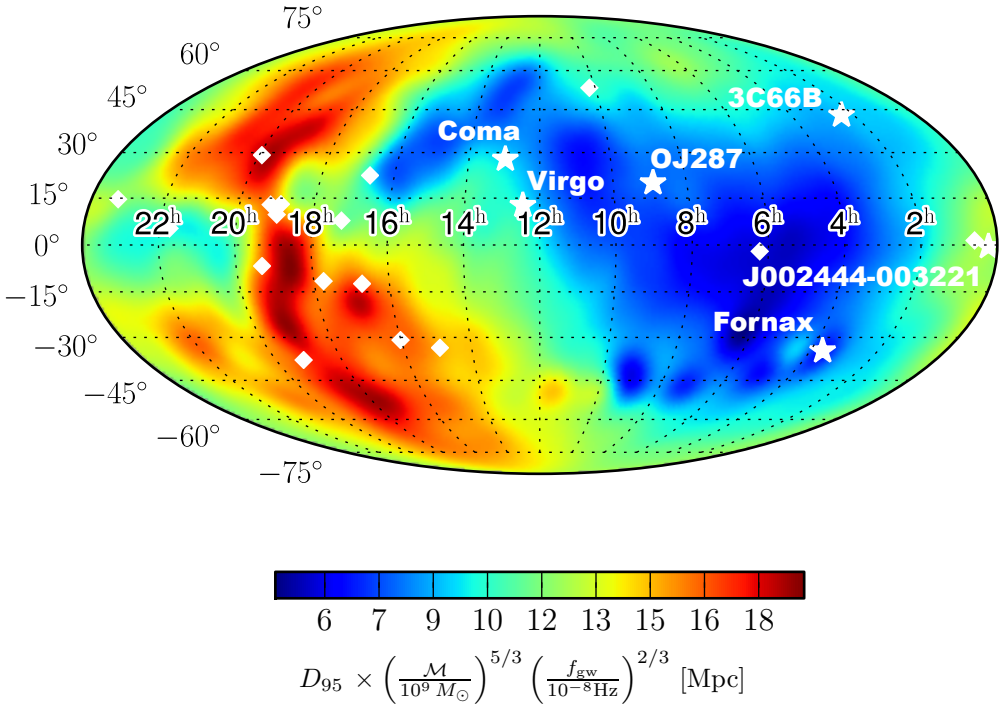


Figure 22 : 95% lower limit on the luminosity distance as a function of sky location computed using the Bayesian method including the noise model. The values in the colorbar are calculated assuming a chirp mass of $\mathcal{M} = 10^9 M_\odot$ and a GW frequency $f_{\rm gw} = 1 \times 10^{-8}$ Hz. The white diamonds denote the locations of the pulsars in the sky and the black(white) stars denote possible SMBHBs or clusters possibly containing SMBHBs.

the frequentist $\mathcal{F}_p$-statistic. Firstly, we notice that the overall distribution is quite similar to the antenna pattern response (i.e., $1 + \cos\mu$) as is to be expected in the case of no detection. Due to this, we are most sensitive (larger lower limit on luminosity distance) at sky locations near the best timed pulsars (i.e., J1713+0747, B1855+09, J1909-3744) and least sensitive in the opposite direction. More quantitatively, we note that in the most sensitive areas of the sky we can constrain the luminosity distance $d_L \gtrsim 47$ Mpc for $\mathcal{M} = 10^9 \mathrm{M}_\odot$. Furthermore, it is possible to constrain the luminosity distance $d_L \gtrsim 2$ Gpc in the most sensitive sky locations if we consider $10^{10}\mathrm{M}_\odot$ chirp mass sources. It should be noted that the Bayesian fixed-noise search gives nearly identical results to the fixed-noise frequentist search.

We now move to the sky-dependent upper limit computed using the full Bayesian technique where the GW and noise parameters are varied simultaneously. The first observation that we make is that the overall scale is about a factor of 2 lower than the fixed-noise frequentist or Bayesian upper limit. At first this may be surprising given the general agreement of the sky-averaged upper limits of Figure 20; however, full Bayesian sky-dependent upper limits exacerbate the problem of relatively few pulsars contributing to the overall PTA sensitivity as shown in Figure 17. Another difference in this upper limit, as opposed to the frequentist upper limit, is that it does not quite match the expected antenna pattern response function. These differences are due to the fact that we are simultaneously varying the GW and noise parameters, and when only one or a few pulsars contribute to the PTA sensitivity, there is a degeneracy between intrinsic red-noise processes in the pulsar and a common GW among all pulsars. In other words, it is very difficult to distinguish between a low-frequency continuous GW and a red noise process if only a small number of pulsars have sufficiently low noise levels to resolve the GW.

Because Bayesian upper limits marginalize or integrate over all parameters except the amplitude, the correlations between the GW and the red noise amplitude will broaden the 1-d pdf of the amplitude and thus will result in larger upper limits compared to the fixed-noise case. As is clear from Figure 22, the aforementioned effect is very strong for GW sky locations near our best timed pulsars. For example, we are not most sensitive

to GWs around the sky location of PSR J1713+0747 because this pulsar contributes a very large percentage of the overall SNR of the GW in this case and thus results in a very large correlation between the GW and red noise amplitudes.

Since, at the moment, we have no way of measuring the noise properties of the pulsars independently of any GWs that may be present in the data, to perform a completely robust upper limit or search we must allow both to vary simultaneously. Given this reality, we must view any fixed-noise results with the caveat that they assume that the noise parameters are measured perfectly and are independent of any GWs in the data.

Unfortunately, many of the GW hotspots and potential SMBHB sources are located at insensitive sky locations, for both frequentist and Bayesian analyses, where our lower limit on distance only allows us to constrain $10^{10}\mathrm{M}_{\odot}$ sources. This fact is a great argument for aggressive pulsar search campaigns and the addition of new pulsars to the PTA at sky locations that are currently insensitive (Burt et al. 2011).

Constraints on the SMBHB Coalescence Rate

A non-detection of continuous GWs, as we have presented here, allows us to compute an upper limit on the rate of SMBH coalescences using methods presented in Wen et al. (2011). Since we have made no detections, we assume Poisson statistics for the probability of an event (i.e., a detectable signal) occurring, that is, the probability of no events is $e^{-\langle N \rangle}$, where $\langle N \rangle$ is the expected number of events. We use this probability distribution function to place a 95% upper limit on the expected number of events such that $\exp(-N_{95}) = 0.05$, telling us that $\langle N \rangle \leq \langle N_{95} \rangle = 3$. Therefore, if the expected number of events were greater than 3, at least one source would have been detected with 95% probability. Now, following Wen et al. (2011), the expected number of events is

$$\begin{aligned} \langle N \rangle = \int \frac{d^2 R}{d\log_{10}(1+z) d\log_{10}(\mathcal{M}_r)} \left(\frac{df}{dt}\right)^{-1} \\ \times P_d(\mathcal{M}_r, z, f)\, d\log_{10}(1+z)\, d\log_{10}(\mathcal{M}_r)\, df, \end{aligned} \tag{5.5.5}$$

where R is the coalescence rate, the $P_d(\mathcal{M}_r, z, f)$ is the probability of detecting an SMBHB with chirp mass $\mathcal{M} = \mathcal{M}_r(1+z)$, redshift z, and observed GW frequency f. Following the derivation in Wen et al. (2011) and making the assumption that the differential

coalescence rate does not vary significantly over the range $\Delta \log_{10} \mathcal{M}_r = 1$ and $\Delta \log_{10}(1+z) = 0.2$, it is possible to show that

$$\frac{d^2R}{d\log_{10}(1+z)\log_{10}(\mathcal{M}_r)} < \frac{15}{\int \left(\frac{df}{dt}\right)^{-1} P_d(\mathcal{M}_r, z, f)\, df}. \tag{5.5.6}$$

In order to compute the detection probability P_d, we make use of the $\mathcal{F}_p$ statistic. We use the same method that we have used for the upper limits, except now we compare the the value of $\mathcal{F}_p$ computed using simulated data with injections to a specified threshold based on a FAP of 10^{-4}. We use 10,000 realizations at each value of z and f. After the probability of detection is computed, we numerically integrate the above expression to obtain a limit on the differential coalescence rate. It should be noted that we will be able to place more meaningful constraints on the coalescence rate using upper limits on the amplitude of a stochastic background of SMBHBs; however, this is beyond the scope of this work and will be addressed in a future paper. In Figure 23 we plot our constraints on the differential coalescence rate as a function of redshift. Since we have made the assumption that this differential coalescence rate does not vary significantly over an order of magnitude in chirp mass, the results presented here are for the $\mathcal{M}_r = 10^{10}\mathrm{M}_\odot$ case. We are unable to place meaningful constraints on less massive systems. The light gray(red) shaded area is constructed using the model presented in Jaffe & Backer (2003) along with measurements from the Sloan Digital Sky Survey (Wen et al. 2009). The medium gray(blue) shaded area is constructed by considering the different galaxy merger rates based on observations (Sesana 2013b) along with the most recent MBH-sigma relation from McConnell & Ma (2013). The dashed line comes from an a posteriori implementation of the McConnell & Ma (2013) MBH-sigma relation into the semi-analytic model of Guo et al. (2011) assuming accretion onto both SMBHs before merger. The black(green) shaded region is constructed by using the observed evolution of the galaxy mass function combined with the MBH-M-stars relation from McConnell & Ma (2013) to calibrate an analytical model for evolving the mass function via mergers (McWilliams et al. 2012). The figure shows that the coalescence rate for MBHs of $\sim 10^{10}\mathrm{M}_\odot$ is poorly constrained. This is mostly because of the steepness of the galaxy mass function at such high masses: a small change in the slope results in a large variation in the sparse population of $10^{10}\mathrm{M}_\odot$ black holes. The intrinsic scattering (e.g. Gültekin et al. 2009b) and poor knowledge of

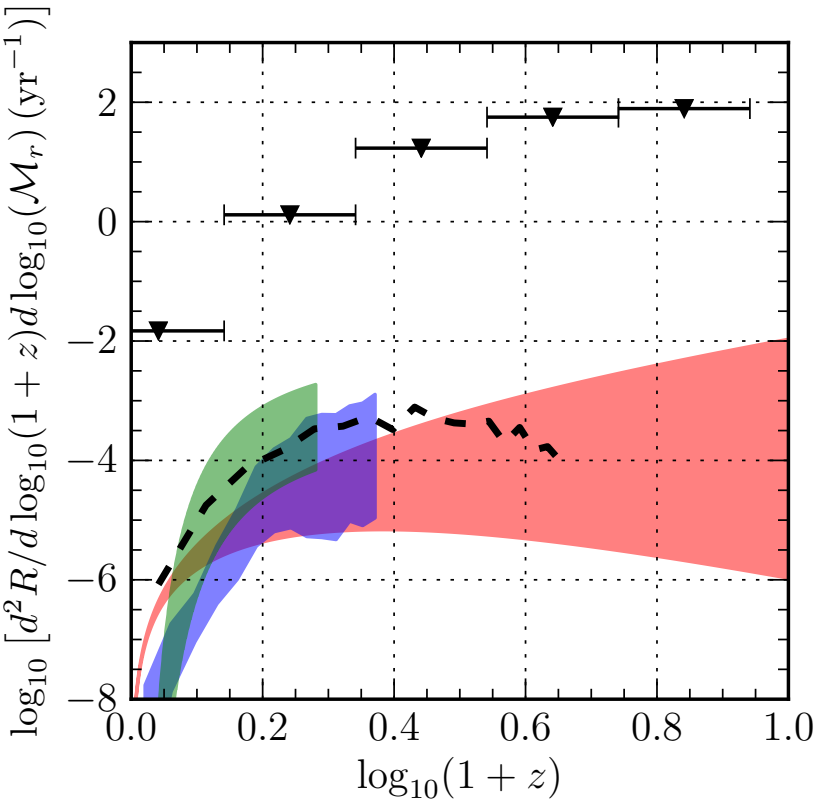


Figure 23 : Differential coalescence rate of SMBHBs per redshift per chirp mass with mass bin centered on $10^{10}\mathrm{M}_{\odot}$ and width 1 dex. We have chosen to explore only the highest masses since these high mass sources are the ones likely to be detected by GW searches in the future. The black triangles represent our upper 95% upper limits, the light gray(red) shaded area show expected coalescence rate estimates obtained from Jaffe & Backer (2003) as well as data from the Sloan Digital Sky Survey (Wen et al. 2009). The medium gray(blue) shaded region comes from the phenomenological models of Sesana (2013b) and the black dashed line comes from an a posteriori implementation of the McConnell & Ma (2013) MBH-sigma relation into the semi-analytic model of Guo et al. (2011). The black(green) shaded region is constructed by using the observed evolution of the galaxy mass function combined with the MBH-M-stars relation from McConnell & Ma (2013) to calibrate an analytical model for evolving the mass function via mergers (McWilliams et al. 2012).

the behavior of the MBH-galaxy relations at the high mass end (e.g. Lauer et al. 2007) add further uncertainties, making the coalescence rate estimate problematic. As is clear from the figure, we are unable to place any constraints on the physical models mentioned above; however, as our GW sensitivity improves with time, we will begin to place meaningful constraints on physical models pertaining to the coalescence rate of SMBHBs.

5.6 Discussion

5.6.1 Future Improvements

Predicting the future sensitivity of PTAs to continuous GWs is quite difficult and depends on a number of poorly constrained factors that make up the entire noise budget for each pulsar (Cordes & Shannon 2010). A detailed study of these effects on the future performance of a PTAs sensitivity to continuous GWs is beyond the scope of this work. Here we simply derive a rough scaling law for the SNR of a continuous GW measured by a PTA and make a few statements about expected future performance. The square of the SNR is defined to be[15]

$$\rho^2 = \sum_\alpha (s_\alpha | s_\alpha) = \sum_\alpha 2\mathrm{Re} \int_{-\infty}^{\infty} df \frac{|\tilde{s}_\alpha(f)|^2}{S_\alpha^n(f)}, \tag{5.6.1}$$

where $\tilde{s}_\alpha(f)$ is the Fourier transform of the GW induced timing residuals and $S_\alpha^n(f)$ is the power spectral density of the noise for the αth pulsar. As we mentioned above, the frequency of the GW will not vary over the observation time and the waveform is approximated by a sine wave at a single frequency. We will further assume that the pulsar term is at the same frequency for this scaling law computation. The SNR then becomes

$$\begin{aligned} \rho^2 &\approx \sum_\alpha 2\mathrm{Re} \int_{-\infty}^{\infty} df \frac{A(f)^2 a_\alpha^2 \delta(f-f')^2}{S_\alpha^{\mathrm{red}}(f) + \sigma_\alpha^2/c_\alpha} \\ &\approx A(f')^2 \sum_\alpha \frac{T_\alpha}{2} \frac{a_\alpha^2}{S_\alpha^{\mathrm{red}}(f') + \sigma_\alpha^2/c_\alpha}, \end{aligned} \tag{5.6.2}$$

where $\delta(f-f')$ is the Dirac delta function, $A(f')$ is the pulsar independent amplitude of the GW, a_α is a geometric factor that depends on the antenna pattern functions for each pulsar, c_α is the observing cadence for each pulsar, and $S_\alpha^{\mathrm{red}}(f')$ and σ_α are the red noise

[15] We ignore timing model fitting here and make use of the Fourier domain for ease of computation.

power spectral density at $f = f'$ and white noise RMS, respectively. Our sensitivity to continuous GWs is proportional to the SNR, thus this expression for the SNR can serve as a proxy for how our upper limits and sensitivity will improve with various quantities. It is interesting to examine this scaling law in the white noise and red noise dominated regime

$$\rho_{\rm white} \propto \left(\sum_{\alpha} \frac{T_{\alpha} a_{\alpha}^2 c_{\alpha}}{\sigma_{\alpha}^2} \right)^{1/2} \tag{5.6.3}$$

$$\rho_{\rm red} \propto \left(\sum_{\alpha} \frac{T_{\alpha} a_{\alpha}^2}{S_{\alpha}^{\rm red}(f')} \right)^{1/2} . \tag{5.6.4}$$

The above scaling laws tell us that many pulsars distributed across the sky with high timing precision, high observing cadences, and low red noise levels observed over a long baseline will result in the best possible sensitivity to continuous GWs. New pulsar timing backends at Arecibo and the GBT (DuPlain et al. 2008) give roughly a factor of two higher timing precision for many pulsars which will translate into an expected upper limit on the amplitude of continuous GWs that is a factor of two more constraining. As we acquire more data on our currently timed and newly discovered pulsars we will gain more sensitivity and will be able to probe to lower frequencies. Access to IPTA data would essentially serve to increase the observing cadence, and thus our sensitivity, since we would have complementary data from many different observatories measured at different times. Furthermore, current pulsar search campaigns (Lynch & Green Bank North Celestial Cap Survey Collaborations 2013) are discovering new MSPs in our least sensitive sky locations (see Figure 21 and Figure 22) which will dramatically increase our overall sky coverage and will allow for better distinction between GW and noise models. Finally, advanced detectors, such as the Square Kilometer Array (SKA; Lazio 2013) are expected to time tens of pulsars at or below the 100 nanosecond level, which will likely solve many of the current problems that we face with poor angular sensitivity and inability to distinguish between single GW source and intrinsic pulsar noise.

In the red noise dominated regime, the cadence of observations and the overall white noise level is negligible and we essentially only gain sensitivity through the addition of new pulsars and continued timing. Here we note again that there is very little evidence for red noise in the 5-year NANOGrav data set but this may change in the future as we become

more sensitive to the stochastic GW background which would induce a common red noise signal in all pulsars. Of course, red noise from a stochastic background of GWs or from intrinsic pulsar spin-noise (Shannon & Cordes 2010) is likely to have a steep spectrum and this red-noise dominated regime would only apply at the lowest frequencies.

For many pulsars in the 5-year dataset we are in the white noise dominated regime and since the SNR consists of the sum of the inverses of the white noise RMS values, we see that only the best timed pulsars will contribute (as we have seen throughout this paper) and one may argue that we should focus all observing time on the best pulsars. However, as was shown in Siemens et al. (2013), our sensitivity to the stochastic background has a significantly weaker dependence on observing cadence and white noise RMS but has a *linear* dependence on the number of pulsars in the array. Thus, it is difficult to realistically optimize a PTA for both continuous and stochastic GW sources.

5.6.2 Conclusions

In this chapter we have performed various searches for continuous GWs from non-spinning SMBHBs in circular orbits using both frequentist and Bayesian techniques. Specifically, we have run a fixed-noise frequentist and Bayesian pipeline, as well as a varying noise Bayesian pipeline. In the absence of any detections we have placed upper limits on the strain amplitude of continuous GWs as a function of GW frequency. We have also computed a lower limit on the distance to such SMBHBs as a function of sky location, as well as placing constraints on the differential coalescence rate of such SMBHBs. Our sky-averaged upper limits on strain amplitude as a function of frequency are a factor of ~ 3 times more constraining than the previously published upper limits (Yardley et al. 2010) and we see good agreement between all three data analysis methods. Although improving, our limits still lie well above the amplitudes of individual sources produced from several realizations of an optimistic SMBHB population. We have shown that with good estimates of the intrinsic noise we can rule out any sources with luminosity distance < 2 Gpc and a chirp mass of $\sim 10^{10}\mathrm{M}_{\odot}$. Unfortunately we are not yet able to place any constraints on predictions for the coalescence rate of SMBHBs obtained from both theory and observations.

Throughout the chapter we have made several statements about what is needed for completely *robust* data analysis techniques and what will be required from future PTAs in order to secure a *confident* detection of a continuous GW. These statements can be summarized as follows:

1. Currently we have no way to confidently separate intrinsic noise in the residuals from any GW that may be present. Therefore, it is necessary to include both noise and GW parameters in any data analysis pipeline that aims to be truly robust. This is not to say that fixed-noise methods should not be used; instead we advocate a hierarchical approach where the faster fixed-noise methods are used as a first-pass and then followed up with a full GW plus noise search. Lastly, a signal with more information, such as that from an eccentric system, could help break this degeneracy between signal and noise models and will be the subject of a future paper.

2. Even with simultaneous noise and GW characterization, unless we have several well timed pulsars (with very similar timing precision on all) with decent sky coverage, a *confident* detection of a continuous GW is unlikely even if the signal is loud.

While not as likely as a detection of a stochastic GW background, with continually improving timing precision, the addition of new pulsars to PTAs and improved data analysis techniques, prospects are good for obtaining astrophysically constraining GW limits, or possibly even a detection of a continuous GW, over the next decade.

Appendix 5.A Pulsar Term Frequency Evolution

In chapter 1 we derived the general form of frequency evolution due to emission of GWs leading to the following expression for the angular phase and orbital frequency evolution

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

and

$$\omega(t) = \left(\omega_0^{-8/3} - \frac{256}{5}\mathcal{M}^{5/3}t\right)^{-3/8}. \tag{5.A.2}$$

Eqs. 5.A.2 and 5.A.1 are true in general and can be applied when the frequency evolves appreciably over the total observing time. However, it is very useful to work under the

assumption of slowly evolving binaries where $T_{\rm chirp} \gg T$, with T the observing time and

$$T_{\rm chirp} = \frac{\omega_0}{\dot{\omega}} = 3.2 \times 10^5\,{\rm yr} \left(\frac{\mathcal{M}}{10^8\,{\rm M}_\odot}\right)^{-5/3} \left(\frac{f_0}{1 \times 10^{-8}\,{\rm Hz}}\right)^{-8/3}. \tag{5.A.3}$$

Since typical PTA observations are on the order of 10 – 20 years and $T/T_{\rm chirp} \sim 10^{-4}$, this is a safe assumption for a broad range of masses and initial orbital frequencies of interest. With this approximation we can write the orbital frequency and phase for the earth term simply as

$$\Phi_e(t) = \Phi_0 + \omega_0(t - t_0) \tag{5.A.4}$$

$$\omega_e(t) = \omega_0. \tag{5.A.5}$$

However, for the pulsar term we are "seeing" the phase and frequency at a retarded time $t_p = t - L(1 - \cos\mu)$, where L is the pulsar distance and μ is the angle between the GW and the pulsar on the sky. Because pulsar distances are on the order of a few kpc, this means that the total time baseline is on the order of thousands of years and we would expect frequency evolution over those timescales. However, just because the pulsar "sees" a different frequency than the earth, this does not mean that the frequency at the pulsar changes over the observation time. For this reason we can write the phase and frequency at the pulsar in a similar manner

$$\Phi_p(t) = \Phi_{p,0} + \omega_p t \tag{5.A.6}$$

$$\omega_p(t) = \omega_p. \tag{5.A.7}$$

We can determine the "pulsar frequency" by evaluating Eq. (5.A.2) and setting $t = t_p$

$$\begin{aligned}
\omega_p(t) &= \omega_0 \left(1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}(t_p - t_0)\right)^{-3/8} \\
&= \omega_0 \left(1 + \frac{8}{3}\frac{\dot{\omega}_0}{\omega_0}L(1 - \cos\mu) + \frac{8}{3}\frac{\dot{\omega}_0}{\omega_0}(t - t_0)\right)^{-3/8} \\
&\approx \omega_0 \left(1 + \frac{8}{3}\frac{\dot{\omega}_0}{\omega_0}L(1 - \cos\mu)\right)^{-3/8} \equiv \omega_p,
\end{aligned} \tag{5.A.8}$$

In the above, we can safely ignore the last term in the second line by the reasoning that the frequency does not evolve over the observation time. Notice that the pulsar term frequency is always less than the earth term frequency as we are observing the dynamics

of the SMBMB in the past when the orbital separation was larger. Determining the pulsar phase in this approximation is a bit trickier. Re-writing Eq. (5.A.1), we get

$$\begin{aligned} \int_{\Phi(t_{p,0})}^{\Phi(t_p)} d\Phi &= \int_{t_{p,0}}^{t_p} dt' \omega(t') \\ \Phi_p(t) - \Phi_{p,0} &= \omega_p(t - L(1-\cos\mu)) + \omega_p L(1-\cos\mu) \\ \therefore \Phi_p(t) &= \Phi_{p,0} + \omega_p t, \end{aligned} \tag{5.A.9}$$

where we have used the fact that $\omega(t) = \omega_p$ in the region of integration and we have adopted a notation in which $\Phi_p(t) \equiv \Phi(t_p)$. To determine the initial phase at $t = -L(1-\cos\mu)$ we use Eq. (5.A.1) to obtain

$$\Phi_{p,0} = \Phi(t = -L(1-\cos\mu)) = \Phi_0 + \frac{1}{32\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega_p^{-5/3}\right) \tag{5.A.10}$$

Although the above expressions for $\Phi_e(t)$ and $\Phi_p(t)$ are approximations, they hold true for nearly all values of $\mathcal{M}$ and ω_0 that we would expect in nature.

Appendix 5.B Auxiliary Pulsar Mode Jump

In E13, we discussed the difficulty posed by including the pulsar distance as a search parameter, showing that a very small change to the pulsar distance (≤ 1 pc) can result in a phase shift in the GW waveform of order 2π. In that work we sidestepped this problem by breaking the pulsar term into a "phase" term and an "evolution" term. The phase term corresponds to very small jumps in the pulsar distance that will change the constant phase of the pulsar term, whereas the evolution term corresponds to large jumps in the pulsar distance that will change the frequency evolution. We used separate parameters to jump in the phase and evolution. More explicitly, we introduce a pulsar phase for each pulsar that is used in the phase term and *also* include the pulsar distance that is only used in the evolution term. While this method allows for good mixing and acceptance rates, it adds an extra $N_{\rm psr}$ parameters to the search.

Here we will describe a new method that does not require any additional parameters. This jump technique is summarized as follows:

1. Perform initial jump (either correlated or uncorrelated as described above).

2. Construct pulsar phase of Eq. (5.A.10) using the new parameters. This phase is likely several radians from the pre-jump pulsar phase due to the pulsar distance jump.

3. We desire a small Gaussian jump in the pulsar initial phase. To accomplish this we will slightly modify the pulsar distance such that

$$\Phi_{p,0}(L^1 + \delta L) = \Phi^0_{p,0} + \delta\phi, \tag{5.B.1}$$

where the 1 and 0 superscripts denote post and pre-jump values, respectively, δL is a small pulsar distance offset, and $\delta\phi$ is a small Gaussian phase jump. We can re-write the above expression

$$\Phi^1_{p,0} + \left.\frac{d\Phi_{p,0}}{dL}\right|_{L=L^1} \delta L = \Phi^0_{p,0} + \delta\phi, \tag{5.B.2}$$

where $\Phi^1_{p,0} = \Phi_{p,0}(L^1)$ and we have simply used a Taylor expansion. Making use of Eq. (5.A.10) we solve for δL

$$\delta L = \frac{\Phi^1_{p,0} - \Phi^0_{p,0} + \delta\phi}{\omega_p(1 - \cos\mu^1)}. \tag{5.B.3}$$

4. Now let $L_{\rm new} = L^1 + \delta L$.

Essentially what we have done is to turn a pulsar distance jump into a pulsar phase jump. So in essence we are not breaking detailed balance as we are simply using the pulsar distance as an *auxiliary* parameter and initial pulsar phase as the actual *search* parameter. This auxiliary jump is called after every jump proposal in the cycle to ensure reasonable acceptance rates.

Chapter 6

First-order Likelihood Approximation for Stochastic GW Background Detection

> "*What I came to realize is that fear, that's the worst of it. That's the real enemy. So, get up, get out in the real world and you kick that bastard as hard as you can right in the teeth.*"
>
> — Walter White, *Breaking Bad: Better Call Saul*

This chapter is based on:
An Efficient Approximation to the Likelihood Function For Gravitational Wave Stochastic Background Detection Using Pulsar Timing Data
J. A. Ellis, X. Siemens, R. van Haasteren
ApJ (2013), 769, 63

6.1 Introduction

All the SMBBH mergers that have taken place throughout the history of our universe produce a stochastic background of gravitational waves (Lommen & Backer 2001; Jaffe & Backer 2003; Wyithe & Loeb 2003; Volonteri et al. 2003; Enoki et al. 2004; Sesana et al. 2008; Sesana 2013b; McWilliams et al. 2012), as well as individual periodic signals that may be detectable as above the confusion noise (Sesana et al. 2009; Sesana & Vecchio 2010; Roedig & Sesana 2012; Ravi et al. 2012; Mingarelli et al. 2012), and bursts (van Haasteren & Levin 2010; Cordes & Jenet 2012). A number of techniques have been

implemented to search pulsar timing data for the stochastic background (Detweiler 1979; Stinebring et al. 1990; Lommen 2002; Jenet et al. 2005, 2006; Anholm et al. 2009; van Haasteren et al. 2009a,b; Yardley et al. 2011; van Haasteren et al. 2011; Cordes & Shannon 2012; Demorest et al. 2013), as well as periodic signals (Jenet et al. 2004; Yardley et al. 2010; Corbin & Cornish 2010; Lee et al. 2011; Ellis et al. 2012b; Babak & Sesana 2012; Ellis et al. 2012c; Petiteau et al. 2013), and bursts (Finn & Lommen 2010).

For stochastic background searches, evaluations of the full likelihood are computationally challenging. PTAs are currently timing up to a few tens of pulsars, with several thousand points each. In addition, the likelihood function depends not only on the relatively small number of parameters that characterize GW stochastic background, but also on several intrinsic red and white noise parameters for each pulsar. A number of techniques have already been introduced to reduce the computational burden of such searches (van Haasteren 2013; Lentati et al. 2013b; Taylor et al. 2012), and we will discuss these results later in the chapter.

Although the stochastic background produces random changes in the times-of-arrival (TOAs) of an individual pulsar, the cross-correlation of its effects on two pulsars only[1] depends on the angular separation between pulsars (Hellings & Downs 1983) for an isotropic background. In this chapter we introduce an efficient approximation to the likelihood by using an expansion to first order in the amplitude of the cross-correlation terms introduced by Anholm et al. (2009). This technique has already used to analyze the first International Pulsar Timing Array Mock Data Challenge (Ellis et al. 2012a). The approximation affords us a computational savings quadratic in the number of pulsars in the pulsar timing array, a factor of a one to three orders of magnitude, depending on the size of the PTA.

This chapter is organized as follows. In Section 6.2 we write the likelihood function for the parameters of the stochastic background as well as intrinsic noise parameters of the pulsars, and introduce the first order approximation in the amplitude of the

[1]In general, the effect of this cross correlation will depend on the individual pulsar terms. However, as was shown in Anholm et al. (2009), this contribution will become negligible for pulsar distances and GW frequencies of interest.

cross-correlations, in Section 6.3, we show the effectiveness of our approximation using simulated gravitational wave backgrounds, and that the level of bias introduced by our approximation is negligible for astrophysically reasonable stochastic background amplitudes. We conclude in Section 6.4 with a summary of our results, compare our results to other work to increase the computational efficiency of stochastic background searches (van Haasteren 2013; Lentati et al. 2013b; Taylor et al. 2012), and introduce a technique that can be used to search for a combination of continuous wave signals and stochastic backgrounds, a possibility suggested by recent work (Ravi et al. 2012), which will be the basis for future work.

6.2 The Likelihood Function

In this chapter, we make use of the linear transformation approach to the likelihood function. We will also show that, in a frequentist sense, the maximum of the expectation value of the likelihood function is an unbiased estimator of the noise parameters in the low-signal regime.

Since we have assumed that our noise n is Gaussian and stationary, for a pulsar timing array with M pulsars we can write the probability distribution as the multi-variate Gaussian

$$p(\mathbf{n}|\vec{\theta}) = \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma}_n)}} \exp\left(-\frac{1}{2}\mathbf{n}^T\mathbf{\Sigma}_n^{-1}\mathbf{n}\right), \tag{6.2.1}$$

where

$$\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix} \tag{6.2.2}$$

is a vector of the noise time-series, $n_\alpha(t)$ for all pulsars, $\mathbf{\Sigma}_n$ is the *pre*-fit noise covariance matrix and $\vec{\theta}$ is a set of parameters that characterize the noise. Henceforth, a greek subscript will denote the pulsar number. As we noted above, we do not actually measure

$\mathbf{n}$, we measure the timing residuals $\mathbf{r} = \mathbf{R}\mathbf{n}$ where

$$\mathbf{R} = \begin{bmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_M \end{bmatrix}. \tag{6.2.3}$$

Extending Eq. (2.4.5) to multiple pulsars, the likelihood in terms of the timing residual data is simply

$$p(\mathbf{r}|\vec{\theta}) = \frac{1}{\sqrt{\det(2\pi\mathbf{\Sigma})}} \exp\left(-\frac{1}{2}\mathbf{r}^T\mathbf{\Sigma}^{-1}\mathbf{r}\right). \tag{6.2.4}$$

The inverse of $\mathbf{\Sigma}$ does not formally exist since we have removed degrees of freedom by fitting out the timing model. In practice, we can make use of a singular value decomposition to compute the determinant and pseudoinverse to evaluate the likelihood. Viewed in this way, the likelihood function for the residuals is simply a change of coordinates where $\mathbf{R}$ is a linear (but not invertible) map from $\mathbf{n} \to \mathbf{r} = \mathbf{R}\mathbf{n}$.

The covariance matrix for the timing residuals is the block matrix,

$$\mathbf{\Sigma} = \begin{bmatrix} P_1 & S_{12} & \dots & S_{1M} \\ S_{21} & P_2 & \dots & S_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ S_{M1} & S_{M2} & \dots & P_M \end{bmatrix}, \tag{6.2.5}$$

where

$$P_\alpha = \langle r_\alpha r_\alpha^T \rangle, \tag{6.2.6}$$

$$S_{\alpha\beta} = \langle r_\alpha r_\beta^T \rangle\big|_{\alpha\neq\beta}, \tag{6.2.7}$$

are the auto-covariance and cross-covariance matrices, respectively, for each set of residuals. It is very important to note that we work *exclusively* in the post-fit variables. As above we use the post-fit residuals, $r_\alpha = R_\alpha n_\alpha$ and the post-fit auto- and cross-correlation matrices, $P_\alpha = R_\alpha P_\alpha^{\rm prefit} R_\alpha^T$ and $S_{\alpha\beta} = R_\alpha S_{\alpha\beta}^{\rm prefit} R_\beta^T$. Henceforth, we will drop any mention of pre-fit or post-fit as we will only work with post-fit variables.

It is worth pointing out that this treatment is somewhat different from previous Bayesian analyses (van Haasteren et al. 2009a; van Haasteren & Levin 2010; van Haasteren

et al. 2011) (VHML). We use a *conditional* probability distribution function (pdf) whereas VHML used a *marginalized* pdf. In other words, we fix the best fit parameter offsets, $\delta\boldsymbol{\xi}_{\rm best}$ through our use of the projection matrix R, whereas VHML marginalizes over the parameter offsets $\delta\boldsymbol{\xi}$ (See Appendix 6.A for more details).

We would like to use the likelihood to determine the spectral index, $\gamma_{\rm gw}$, and amplitude, $A_{\rm gw}$, of the stochastic background from our data. The GW parameters are the same for all pulsars. In addition, each pulsar will have intrinsic noise parameters as well. The intrinsic pulsar timing noise is normally parametrized with four parameters: an amplitude A_α and spectral index γ_α for a power law red noise process, and EFAC and EQUAD parameters, $\mathcal{F}_\alpha$ and $\mathcal{Q}_\alpha$, for white noise processes. In general the EFAC parameter is a multiplicative factor representing any systematic effects in the uncertainty in each TOA based on the cross correlation of the folded pulse profile with a template (Taylor et al. 1992). The EQUAD parameter is an extra white noise parameter that is added to the TOA error in quadrature and could represent the expected pulse phase jitter (Cordes & Shannon 2010) and other white noise processes that are un-accounted for. Therefore, we write our auto-covariance as a sum of a common GWB term and a pulsar dependent term

$$P_\alpha = N_\alpha + S_{a\alpha}, \tag{6.2.8}$$

where N_α is the intrinsic noise auto-covariance matrix and $S_{a\alpha}$ is the common GWB auto-covariance matrix for pulsar α. It is convenient to work in a block matrix notation where

$$\boldsymbol{\Sigma} = \mathbf{N} + \mathbf{S}_a + \mathbf{S}_c = \mathbf{P} + \mathbf{S}_c, \tag{6.2.9}$$

where $\mathbf{P}$ is a block diagonal matrix with diagonals P_α and $\mathbf{S}_c$ is block matrix with off diagonals $S_{\alpha\beta}$, and zero block matrices on the diagonal.

We will now quickly show that, in a frequentist sense, the maximum of the expectation value of the likelihood function is an unbiased estimator of our signal parameters $\vec{\theta} = \{A_{\rm gw}, \gamma_{\rm gw}, A_\alpha, \gamma_\alpha, \mathcal{F}_\alpha, \mathcal{Q}_\alpha\}$. We write the log likelihood function as

$$\ln \mathcal{L} = -\frac{1}{2}\left[\mathrm{Tr}\, \ln \boldsymbol{\Sigma} + \mathbf{r}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}\right], \tag{6.2.10}$$

where we have used the fact that $\ln \det(A) = \mathrm{Tr} \ln(A)$ for a general matrix, A. To show that the maximum of the expectation value of this likelihood function is an unbiased

estimator of the signal parameters, $\vec{\theta}$, we wish to show that it is maximized, on average, for signal parameters $\vec{\theta} = \vec{\theta}_{\rm true}$. Taking the expectation value we obtain

$$\langle \ln \mathcal{L} \rangle = -\frac{1}{2} \mathrm{Tr} \left[\ln \mathbf{\Sigma} + \mathbf{X} \mathbf{\Sigma}^{-1} \right], \tag{6.2.11}$$

where $\mathbf{X} = \langle \mathbf{r}\mathbf{r}^T \rangle$ is the covariance matrix of the data. Defining $\partial_i = \partial / \partial \theta_i$ we obtain

$$\partial_i \langle \ln \mathcal{L} \rangle = -\frac{1}{2} \mathrm{Tr} \left[\mathbf{\Sigma}^{-1} \partial_i \mathbf{\Sigma} - \mathbf{X} \mathbf{\Sigma}^{-1} \partial_i \mathbf{\Sigma} \mathbf{\Sigma}^{-1} \right]. \tag{6.2.12}$$

Assuming that our noise model is correct, we have $\mathbf{X} = \mathbf{\Sigma}$ and

$$\partial_i \langle \ln \mathcal{L} \rangle = -\frac{1}{2} \mathrm{Tr} \left[\mathbf{\Sigma}^{-1} \partial_i \mathbf{\Sigma} - \partial_i \mathbf{\Sigma} \mathbf{\Sigma}^{-1} \right] = 0, \tag{6.2.13}$$

where we have used the fact that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for general matrices, A and B. Therefore, the maximum of the expectation value of the likelihood function is an unbiased estimator of our model parameters $\vec{\theta}$.

6.2.1 Likelihood with first order approximation

In practice the matrix $\mathbf{\Sigma}$ is quite large and therefore, computationally prohibitive to invert. Since many multi-frequency residual datasets now have on the order of 10^3 points, for many modern PTAs the matrix $\mathbf{\Sigma}$ will be of order $10^4 \times 10^4$. We would like to avoid inverting the full covariance matrix if at all possible. First let us rewrite the cross-covariance as $\mathbf{S}_{c,\alpha\beta} = \zeta_{\alpha\beta} \mathbf{S}_{\alpha\beta}$, where $\mathbf{S}_{\alpha\beta}$ is the temporal cross covariance between pulsar α and pulsar β. The coefficients represent the spatial correlations and are given by the Hellings and Downs coefficients

$$\begin{aligned} \zeta_{\alpha\beta} = \frac{3}{2} \frac{1 - \cos \xi_{\alpha\beta}}{2} \ln \left(\frac{1 - \cos \xi_{\alpha\beta}}{2} \right) \\ - \frac{1}{4} \frac{1 - \cos \xi_{\alpha\beta}}{2} + \frac{1}{2} + \frac{1}{2} \delta_{\alpha\beta}, \end{aligned} \tag{6.2.14}$$

where $\xi_{\alpha\beta}$ is the angular separation of pulsars α and β, and $\delta_{\alpha\beta}$ is the Kronecker delta. We denote $\mathbf{P} = \delta_{\alpha\beta} \mathbf{P}_{\alpha\beta}$ as the auto-covariance matrix of pulsar α describing the noise and auto-covariance of the GWB. We then use the following notation to form matrices from indexed quantities: $\mathbf{P} = \{P_{\alpha\beta}\}$. Now, we perform the expansion of $\mathbf{\Sigma}^{-1}$ in terms of

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \left(\mathbf{P} + \{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\}\right)^{-1} = \left(\mathbb{I} + \mathbf{P}^{-1}\{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\}\right)^{-1}\mathbf{P}^{-1} \\ &\approx \mathbf{P} - \left\{\sum_{\beta,\mu}\zeta_{\beta\mu}\mathbf{P}^{-1}_{\alpha\beta}\mathbf{S}_{\beta\mu}\mathbf{P}^{-1}_{\mu\nu}\right\} \\ &+ \left\{\sum_{\beta,\mu,\nu}\zeta_{\beta\mu}\zeta_{\mu\nu}\mathbf{P}^{-1}_{\alpha\beta}\mathbf{S}_{\beta\mu}\mathbf{P}^{-1}_{\mu\mu}\mathbf{S}_{\mu\nu}\mathbf{P}^{-1}_{\nu\sigma}\right\} + \mathcal{O}(\zeta^3).\end{aligned} \tag{6.2.15}$$

It is also possible to expand the determinant term in a similar fashion

$$\begin{aligned}\ln\det\boldsymbol{\Sigma} &= \mathrm{Tr}\ln\boldsymbol{\Sigma} = \mathrm{Tr}\ln(\mathbf{P} + \{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\}) \\ &= \mathrm{Tr}\left[\ln\mathbf{P} + \ln(\mathbb{I} + \mathbf{P}^{-1}\{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\})\right] \\ &\approx \mathrm{Tr}\left[\ln\mathbf{P} + \mathbf{P}^{-1}\{\zeta_{\alpha\beta}\mathbf{S}_{\alpha\beta}\}\right. \\ &\left. - \left\{\sum_{\beta,\mu,\nu}\zeta_{\beta\mu}\zeta_{\mu\nu}\mathbf{P}^{-1}_{\alpha\beta}\mathbf{S}_{\beta\mu}\mathbf{P}^{-1}_{\mu\mu}\mathbf{S}_{\mu\nu}\mathbf{P}^{-1}_{\nu\sigma}\right\}\right] + \mathcal{O}(\zeta^3).\end{aligned} \tag{6.2.16}$$

Here, the order $\mathcal{O}(\zeta)$ term is zero because $\mathbf{P}$ is block diagonal and $\{\mathbf{S}_{\alpha\beta}\}$ is block traceless and the trace of the product of a diagonal matrix and traceless matrix vanishes. If we ignore all terms of ζ^2 and higher order and return to our original notation then we see that

$$\boldsymbol{\Sigma}^{-1} \approx \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1} + \mathcal{O}(\zeta^2) \tag{6.2.17}$$

$$\ln\det\boldsymbol{\Sigma} \approx \mathrm{Tr}\ln\mathbf{P} + \mathcal{O}(\zeta^2). \tag{6.2.18}$$

This derivation may give us the sense that this expansion may hold true for all GWB amplitudes; however, this is not true as we will now show. Although we have written this approximation in terms of an expansion in the Hellings and Downs coefficients, it is also useful to think of it as an expansion in the amplitude of the GWB. Indeed, that it how it was conceived of in Anholm et al. (2009). We have not performed a true first order expansion however, since the inverse of the auto-correlations matrix $\mathbf{P}^{-1} = (\mathbf{N} + A^2_{\mathrm{gw}}\mathbf{A}_a)^{-1}$, where $A^2_{\mathrm{gw}}\mathbf{A}_a = \mathbf{S}_a$, contains terms of infinite order in the amplitude. We can essentially think of the $\mathcal{O}(\zeta)$ terms in Equations 6.2.17 and 6.2.18 as the corrections to the amplitude parameter when we have a spatially correlated signal. Thus, we have truncated these correction terms at $\mathcal{O}(A^2_{\mathrm{gw}})$ and we would not expect this approximation

to hold as $A_{\rm gw}$ becomes large with respect to the intrinsic noise in the pulsar as we will show in Section 6.3. With these approximations, it is now possible to write the approximate log-likelihood

$$\begin{aligned}\ln \mathcal{L} &= -\frac{1}{2}\left[\mathrm{Tr}\ln \mathbf{P} + \mathbf{r}^T\mathbf{P}^{-1}\mathbf{r} - \mathbf{r}^T\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}\mathbf{r}\right] \\ &= -\frac{1}{2}\sum_{\alpha=1}^{M}\Bigg[\mathrm{Tr}\ln P_\alpha + r_\alpha^T P_\alpha^{-1} r_\alpha \\ &\quad - \sum_{\beta\neq\alpha}^{M} r_\alpha^T P_\alpha^{-1} S_{\alpha\beta} P_\beta^{-1} r_\beta\Bigg].\end{aligned} \tag{6.2.19}$$

In the second line we have explicitly written out the sum over pulsars and pulsar pairs in order to highlight the fact that we only need to invert the individual auto-covariance matrices as opposed to the inverse of the full block covariance matrix, thereby, significantly reducing the computational cost of a single likelihood evaluation. Consider a PTA with M pulsars with N TOAs each. For a full likelihood evaluation we must perform one Cholesky inversion of the full covariance matrix which scales like $\sim \alpha(MN)^3$ and $\sim M^2$ matrix multiplications which scale like $\sim \beta N^3$. However, one evaluation of the first order likelihood requires M Cholesky inversions which scale like $\sim \alpha N^3$ and M matrix multiplications which, again, scale like $\sim \beta N^3$. Though benchmarking tests we have found that $\beta \sim 10\alpha$ and thus the matrix multiplications will dominate both likelihood calls for a reasonable sized PTAs ($M \lesssim 100$) resulting in a computation speedup factor of $\sim (\alpha/\beta)M^2$.

It is possible to analytically show that the maximum of the expectation value of this approximate likelihood is an unbiased estimator in the same manner as above. First we take the expectation value of the log-likelihood

$$\langle \ln \mathcal{L}\rangle = -\frac{1}{2}\,\mathrm{Tr}\left[\ln \mathbf{P} + \mathbf{X}\mathbf{P}^{-1} - \mathbf{X}\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}\right] \tag{6.2.20}$$

and then take a derivative with respect to a model parameter

$$\begin{aligned}\partial_i\langle \ln \mathcal{L}\rangle = -\frac{1}{2}\,\mathrm{Tr}\Big[&\mathbf{P}^{-1}\partial_i\mathbf{P} - \mathbf{X}\mathbf{P}^{-1}\partial_i\mathbf{P}\mathbf{P}^{-1} \\ &+ \mathbf{X}\mathbf{P}^{-1}\partial_i\mathbf{P}\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1} - \mathbf{X}\mathbf{P}^{-1}\partial_i\mathbf{S}_c\mathbf{P}^{-1} \\ &+ \mathbf{X}\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}\partial_i\mathbf{P}\mathbf{P}^{-1}\Big].\end{aligned} \tag{6.2.21}$$

Here we will work in the small signal regime where $A^2_{\rm gw}$ is small compared to the amplitude of the intrinsic noise. Assuming that we have modeled the covariance matrix correctly, we have $\mathbf{X} = \mathbf{\Sigma}$. Writing out the explicit amplitude dependence we assume

$$\mathbf{P} = \mathbf{N} + A^2_{\rm gw}\mathbf{A} \Rightarrow \mathbf{P}^{-1} \approx \mathbf{N}^{-1} - A^2_{\rm gw}\mathbf{N}^{-1}\mathbf{A}\mathbf{N}^{-1} \tag{6.2.22}$$

$$\mathbf{\Sigma} = \mathbf{N} + A^2_{\rm gw}\mathbf{A} + A^2_{\rm gw}\mathbf{C}, \tag{6.2.23}$$

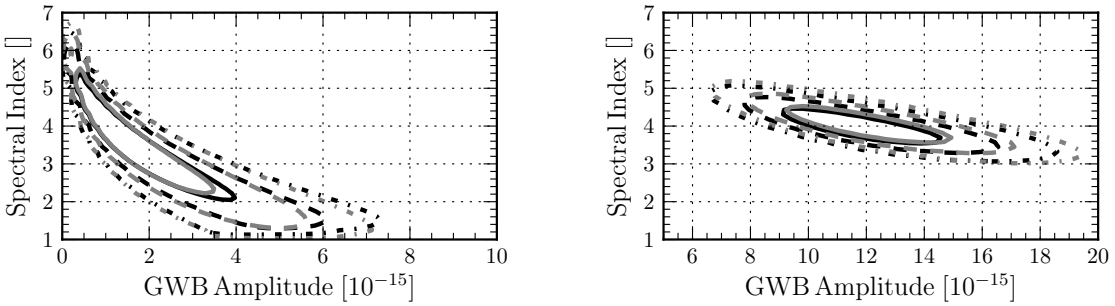
where $\mathbf{N}$, $\mathbf{A}$, and $\mathbf{C}$ are the auto-covariance of the noise, the auto-covariance of the GWB and the cross-covariance of the GWB, respectively. Then, to first order in $A^2_{\rm gw}$ we have

$$\begin{aligned}\partial_i \langle \ln \mathcal{L} \rangle = -\frac{1}{2} \mathrm{Tr} \Big[&\mathbf{N}^{-1} \partial_i (A^2_{\rm gw}\mathbf{A}) \\ &- \mathbf{N}^{-1} \partial_i (A^2_{\rm gw}\mathbf{A}) - \partial_i (A^2_{\rm gw}\mathbf{C}) \mathbf{N}^{-1} \Big] = 0,\end{aligned} \tag{6.2.24}$$

where the first two terms cancel and the third term is the trace of the product of a diagonal matrix and a traceless matrix. Thus, to first order in $A^2_{\rm gw}$, the maximum of the expectation value of this approximate likelihood is an unbiased estimator of the our signal parameters θ in the weak signal limit. In other words, we have shown analytically that for reasonably small stochastic background amplitudes, as is expected, the parameters that we infer from this likelihood function will be unbiased on average. This result will be verified in the next section through the use of simulations.

6.3 Simulations

Here we will compare our first order likelihood approximation to the full likelihood of VHML and perform mock searches of simulated data with and without an injected stochastic GWB in order to demonstrate its efficacy. We will also perform monte-carlo simulations to test the consistency of our likelihood function. These simulations are solely meant as a proof of principle and do not claim to reproduce all features of real PTA data (irregular sampling, jumps, time varying DM corrections, etc.). However, our analysis method makes no assumptions about sampling by operating in the time domain and takes all timing model parameters into account via the projection matrices introduced in Section 6.2. The application of this method to real NANOGrav and IPTA datasets will be the subject of future work. For all simulations in the present work we use TEMPO2 and



(a) (b)

Figure 24 : Comparison of full likelihood (gray) of van Haasteren et al. (2009a) and the first order likelihood (black). (a): 10 pulsars $A = 1 \times 10^{-15}$, (b): 10 pulsars $A = 1 \times 10^{-14}$

and the `fake`, `GWbkgrd`, `general2` and `designmatrix` plugins to generate the residuals and the corresponding design matrices. All simulated white noise is solely radiometer noise at the level of 100 ns unless otherwise noted.

6.3.1 Mock searches

First we will perform a simple test to compare the first order likelihood of this work and the full likelihood of VHML. Here we use a PTA with 10 pulsars observed at a cadence of 20 TOAs per year for 5 years where we have fixed the EFAC parameter to be one (all white noise is encompassed in error bars as simulated) and assume that there is no intrinsic red noise, resulting in a search over two parameter; the amplitude of the stochastic GWB, A, and the power spectral index, γ. For both cases a grid search was carried out with 100 points in each dimension and $A \in (0, 1 \times 10^{-14})$ for an injected value of $A = 1 \times 10^{-15}$ and $A \in (0, 2 \times 10^{-14})$ for an injected value of $A = 1 \times 10^{-14}$, all the while we have $\gamma \in [1, 7]$. The results are presented in Figure 24 where the contours denote the one, two and three sigma credible regions, the gray contours are from the VHML likelihood function and the black contours are from the first order likelihood. In Figure 24(a) we have injected a stochastic GWB with $A = 1 \times 10^{-15}$ and $\gamma = 13/3$. First we notice that the injected value ('×' marker) is well within the 1-sigma credible regions for both likelihood functions. We also see that the confidence contours are nearly identical, with the first order likelihood preferring slightly larger amplitudes and smaller spectral indices.

This simulation indicates that the first order likelihood is a very good approximation to the full likelihood when our signal is relatively small, showing no discernible bias and faithfully reproducing nearly identical credible regions.

In Figure 24(b) we have injected a stochastic GWB with $A = 1 \times 10^{-14}$ and $\gamma = 13/3$. Again, the injected value lies within the 1-sigma credible region, however; now we do notice a difference between two credible regions from the full and first order likelihoods. The first order likelihood is biased towards lower amplitudes and lower spectral indices. In fact we can almost see where the first order approximation begins to break down. Notice that the contours are nearly identical for lower amplitudes and deviate more with increasing amplitude. This behavior is not surprising in that we know that this likelihood is only unbiased to first order in the amplitude as shown in Section 6.2. In fact, it is impressive that this approximation performs this well with only a small bias in the large signal limit (even with timing residuals lower than 100 ns in many pulsars, the signal-to-noise-level of the data simulated here is well above any reasonable estimates for future PTA sensitivities.). This bias will be discussed further in Section 6.3.3.

The simulations used in the work have been quite ideal and do not contain any systematic effects such as clock errors which can manifest as a correlated noise source with uniform correlation coefficients (Yardley et al. 2011), errors in solar system ephemerides, which can manifest as dipole signals in the residuals, or new physics such as non-gr polarization modes (Lee et al. 2008; Chamberlin & Siemens 2012) or massive gravitons (Lee et al. 2010) which would change the shape of the Hellings and Downs curve. We have, for the most part, also assumed that the intrinsic pulsar noise can be assumed to be white gaussian noise with no discernible red noise. While previous work suggests that there will be red noise present in many MSPs (Shannon & Cordes 2010), analyses of the present timing data (van Haasteren et al. 2011; Perrodin et al. 2013b; Ellis et al. 2014) suggest that the data is white noise dominated and there is little to no evidence for red noise. However further study of the model selection problem taking in to account the aforementioned effects is crucial to present detection efforts and will be the subject of a future paper.

6.3.2 The detection problem

We now turn to the question of detection. In a Bayesian analysis we would like to compute the odds that there is a GWB present in our data. Not surprisingly, the tool normally used to this end is the Odds ratio of Bayes factors. Consider two models that we will label M_1 and M_2, then the Odds ratio is defined as

$$\mathcal{O} = \mathcal{B}(M_1, M_2|\mathbf{r})\frac{p(M_1)}{p(M_2)}, \tag{6.3.1}$$

where

$$\mathcal{B}(M_1, M_2|\mathbf{r}) = \frac{\int d\vec{\theta}_1\, p(\mathbf{r}|\vec{\theta}_1, M_1)p(\vec{\theta}_1)}{\int d\vec{\theta}_2\, p(\mathbf{r}|\vec{\theta}_2, M_2)p(\vec{\theta}_2)} \tag{6.3.2}$$

is the Bayes factor (i.e the ratio of the marginalized likelihood functions over parameters $\vec{\theta}_1$ and $\vec{\theta}_2$ corresponding to models M_1 and M_2 respectively), $\mathbf{r}$ is our data and $p(M_1)$ and $p(M_2)$ are the a priori probabilities on models M_1 and M_2 respectively. Note that the Bayes factor is the data dependent part of the odds ratio where the a priori probabilities of the models is somewhat subjective, and as such, we will only consider Bayes factors when discussing detection in the this work [2]. For our purposes, we would like to compare at least three different models when weighing the odds of a stochastic GWB in our data:

1. $M_{\rm gw}$: A power law stochastic GWB with spatial correlations described by the Hellings and Downs coefficients $\zeta_{\alpha\beta}$, amplitude $A_{\rm gw}$ and power spectral index $\gamma_{\rm gw}$, individual power law red noise processes for each pulsar with amplitude A_α and power spectral index γ_α and white noise for each pulsar characterized by an EFAC parameter $\mathcal{F}_\alpha$ and EQUAD parameter $\mathcal{Q}_\alpha$.

2. $M_{\rm corr}$: A common red noise process among pulsars (as suggested in Shannon & Cordes (2010)) with no spatial correlations and individual intrinsic red and white components as in model $M_{\rm gw}$.

3. $M_{\rm null}$: Only intrinsic red and white noise processes with no common red or white noise components among pulsars.

[2] It is possible to use astrophysical information such as the expected level of the stochastic background compared to our noise or the expectation number of single sources to construct the a priori probabilities. Here we will quantify our ignorance by considering equal a priori probabilities of all tested models.

Comparing models $M_{\rm gw}$ and $M_{\rm null}$ will tell us whether or not there is evidence for any common red noise in our data but it will not necessarily tell us that this common noise is due to the stochastic GWB or some other common red noise source. Hence, a large Bayes factor $\mathcal{B}(M_{\rm gw}, M_{\rm null}|\mathbf{r})$ is necessary but not sufficient for detection. However, the comparison of models $M_{\rm gw}$ and $M_{\rm corr}$ can really give us information about the nature of the common red noise signal. As the two aforementioned models are identical except for the spatial correlations, a large Bayes factor $\mathcal{B}(M_{\rm gw}, M_{\rm corr}|\mathbf{r})$ will give us the odds that there is a common red noise process described spatial correlations $\zeta_{\alpha\beta}$. Since these *spatial* correlations are the signature of a stochastic GWB, the condition that this Bayes factor be large is both the necessary and sufficient condition for detection. In fact, this Bayes factor is closely related to signal-to-noise ratios in previous detection schemes (Jenet et al. 2005; Anholm et al. 2009; Yardley et al. 2011; Chamberlin et al. 2013) that measure the significance of the cross correlations.

This first order likelihood approximation has already been tested on the open and closed (Ellis et al. 2012a) IPTA Mock Data Challenge, where all challenges consisted of 130 data points per pulsar with 36 pulsars. For the closed data challenge, we have computed the Bayes factors mentioned in the previous section. In Ellis et al. (2012a) we have shown that we do indeed see very strong evidence for both a common red noise signal and a red noise signal with spatial correlations described by the Hellings and Downs coefficients. However, as we mentioned above, although in this case, the evidence for both models $M_{\rm gw}$ and $M_{\rm corr}$ is very high, as we expect, the Bayes factor $\mathcal{B}(M_{\rm gw}, M_{\rm null})$ is much larger than $\mathcal{B}(M_{\rm gw}, M_{\rm corr})$. For this reason, we expect that in analysis of real PTA data we will begin to see strong evidence for common red noise before we are able to see strong evidence for the expected cross correlations. In other words, as we gain more sensitivity, the first two terms in Eq. (6.2.19) will dominate the likelihood function and the third term will only play a significant role as our sensitivity increases further. A full analysis of this feature along with projected sensitivity curves based on future pulsar timing campaigns and hardware upgrades will be explored in future work.

6.3.3 The Empirical Distribution Function

Here we will test the consistency and unbiasedness of our model through injections. Simply put, it is a type of hypothesis testing similar to the Kolmogorov-Smirnov test. In this test the null-hypothesis, our analysis method is internally consistent, is accepted when for $x\%$ of realizations, the true injected parameter lies within the inner $x\%$ of the marginalized posterior distribution. A similar test was done recently in van Haasteren & Levin (2013) in *one* dimension through the use of the empirical distribution function (EDF). Here we will review this method and generalize it to two dimensional marginalized posterior distributions. We define the inner high-probability region (HPR) of the two-dimensional marginalized posterior distribution as

$$\begin{aligned} \int_W p(\theta_1, \theta_2) d\theta_1 d\theta_2 &= a \\ W &= \{\theta_1, \theta_2 \in \mathbb{R} : p(\theta_1, \theta_2) > L_a\}, \end{aligned} \tag{6.3.3}$$

where L_a is some value > 0 unique to each a that corresponds to a curve of equal probability in the two-dimensional parameter space. In practice we lay down a grid in this two-dimensional parameter space and perform our search over the two parameters of interest (for the stochastic background we search over A and γ, the dimensionless strain amplitude and power spectral index of the GWB). We then define a set of points $\{A_i, \gamma_i\} \in \mathcal{S}_a : p(A_i, \gamma_i) > L_a$, that is to say we find all points in our grid that correspond to posterior values that lie inside our contour curve L_a. To determine if the injected values of $\{A_{\rm true}, \gamma_{\rm true}\}$ lie within the HPR we simply check to see if the injected values are consistent with the set $\mathcal{S}_a$. To do this we first define the complementary set to be $\bar{\mathcal{S}}_a$ such that points that are in this set are outside or the HPR. Now we define two chi-squared functions in the parameter space

$$\chi_a(A_i, \gamma_i)^2 = \left(\frac{A_i - A_{\rm true}}{A_{\rm true}}\right)^2 + \left(\frac{\gamma_i - \gamma_{\rm true}}{\gamma_{\rm true}}\right)^2 \tag{6.3.4}$$

$$\bar{\chi}_a(A_j, \gamma_j)^2 = \left(\frac{A_j - A_{\rm true}}{A_{\rm true}}\right)^2 + \left(\frac{\gamma_j - \gamma_{\rm true}}{\gamma_{\rm true}}\right)^2, \tag{6.3.5}$$

where $\{A_i, \gamma_i\}$ and $\{A_j, \gamma_j\}$ are elements of the sets $\mathcal{S}_a$ and $\bar{\mathcal{S}}_a$, respectively. Finally, we define the empirical distribution function (EDF) as

$$F_k(a) = \frac{1}{k} \sum_{n=1}^{k} \Theta(\min \bar{\chi}_a^2 - \min \chi_a^2), \tag{6.3.6}$$

where $\Theta(x)$ is the Heaviside function. The term inside the sum indicates an event when the injected values are "closer" (in the chi-squared sense) to one of the elements of $\mathcal{S}_a$ than to any of the elements of $\bar{\mathcal{S}}_a$, therefore we can say that the values $\{A_{\rm true}, \gamma_{\rm true}\}$ join the set $\mathcal{S}_a$ and lie within the HPR defined in Eq. (6.3.3). Now that we have defined our EDF, the rest of the analysis mimics van Haasteren & Levin (2013).

For this analysis we simulated 1000 datasets for 6 different scenarios. In all cases we chose the white noise level to be 100 ns while we chose GWB amplitudes of 1×10^{-15}, 2×10^{-15}, and 3×10^{-15} for PTAs with both 10 and 15 pulsars with a 5 year baseline. Figure 25 shows the EDF for the six models outlined above. The thick lines denote a 10

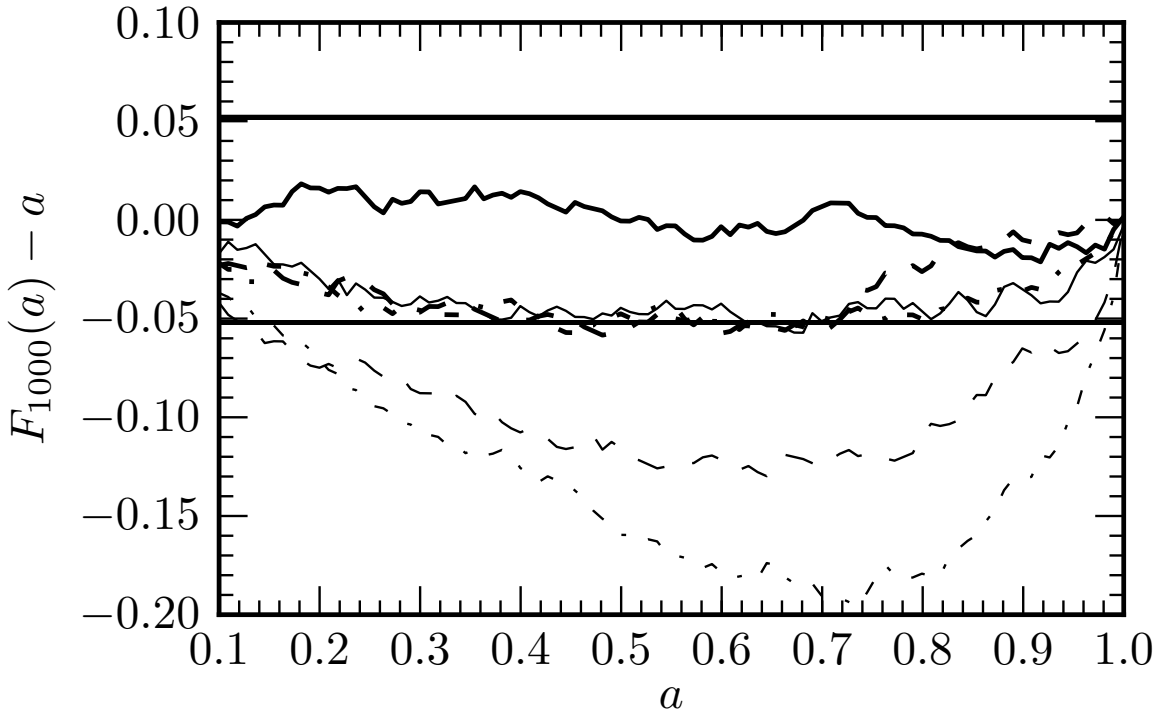


Figure 25 : Empirical distribution function for 6 scenarios. The thick lines denote a 10 pulsar PTA and the thin lines denote a 15 pulsar PTA and the solid, dashed and dotted lines denote injected stochastic GWB amplitudes of 1×10^{-15}, 2×10^{-15}, and 3×10^{-15}, respectively. The solid lines at ± 0.052 represent the value at which we should reject the null-hypothesis that our analysis method is consistent and unbiased.

pulsar PTA and the thin lines denote a 15 pulsar PTA and the solid, dashed and dotted lines denote injected stochastic GWB amplitudes of 1×10^{-15}, 2×10^{-15}, and 3×10^{-15}, respectively. The solid lines at ± 0.052 represent the value at which we should reject the null-hypothesis that our analysis method is consistent and unbiased. Firstly, we note

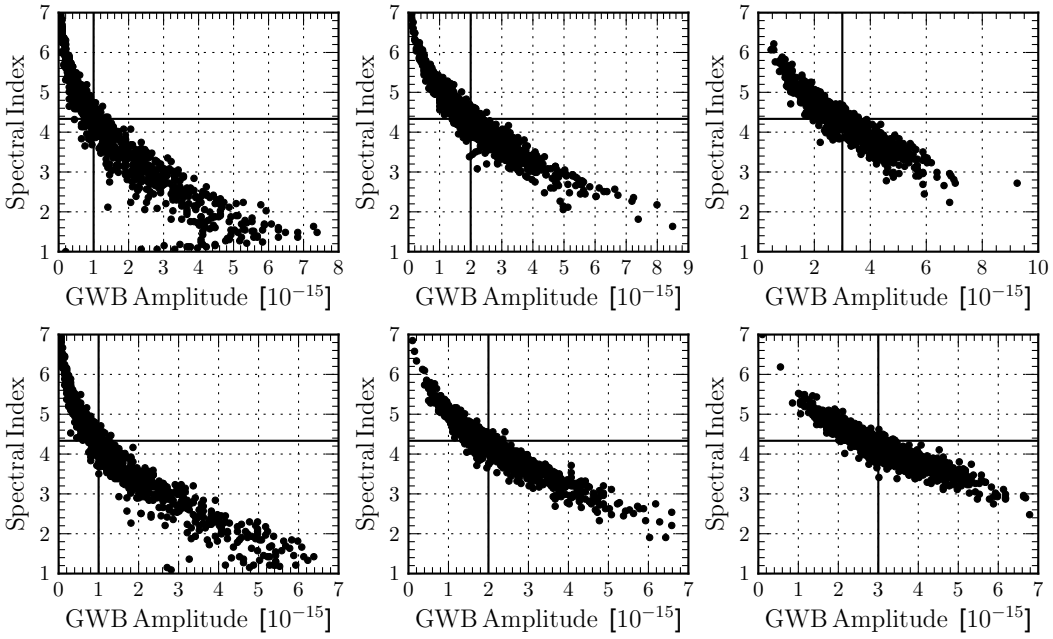


Figure 26 : Here we show the scatter of the maximum likelihood values of the GWB amplitude and spectral index from the Monte-Carlo simulations. From left to right the injected amplitudes are 1×10^{-15}, 2×10^{-15}, and 3×10^{-15} with spectral index 13/3 for a 10 pulsar PTA (top row) and 15 pulsar PTA (bottom row). We can see that nearly all of these distributions display minimal bias.

that for both the 10 and 15 pulsar PTA, our analysis method is consistent for an injected amplitude of $A = 1\times10^{15}$. We obtain similar results in the 10 pulsar case for amplitudes of $A = 2\times10^{-15}$ and $A = 3\times10^{-15}$. Here we do see that our method is indeed slightly biased for these larger amplitudes but the degree of bias is almost negligible. However, for these same amplitudes in the 15 pulsar case there is a significant bias. Even though there is a bias present in these scenarios, the EDF does not give information about how this bias presents itself in the two dimensional parameter space. In Figure 26 we show the two-dimensional scatter plot of the maximum likelihood parameters from our Monte-Carlo simulations. It is clear that the bias in our two-dimensional parameter space of interest is *practically* very small. In fact the means of the distributions for A and γ for the 10 pulsar case are $(1.6, 2.25, 3.14)\times10^{-15}$ and $(4.17, 4.24, 4.23)$, respectively and for the 15 pulsar case we obtain $(1.56, 2.29, 3.22)\times10^{-15}$ and $(4.11, 4.12, 4.13)$, respectively. In the first row of Figure 26 we show the 10 pulsar case with increasing GWB amplitude and the second row we show the same for the 15 pulsar case. In the cases where there is a bias

present, the likelihood function prefers slightly lower spectral indices and slightly larger amplitudes. However, from our experience with the MDC this bias can also present itself by preferring a slightly higher spectral index and lower amplitude. It should be noted that even the smallest of the amplitudes tested here are near the upper range of the expected level of the stochastic GWB (Sesana 2013b) and that the white noise rms of the pulsars is slightly unrealistic in our current PTA regime. In fact we expect to have maybe five or six pulsars that time at or below the 100 ns level while we have many others that have much larger white noise rms. Thus we can conclude that even though our likelihood is somewhat biased at larger amplitudes (as is expected), for *realistic* astrophysically likely stochastic GWBs this method is effectively consistent and unbiased. In fact, in terms of setting upper limits on the stochastic GWB amplitude, this method is practically identical to using the full likelihood, while much more computationally efficient.

6.4 Discussion and Conclusions

Here will will briefly discuss future prospects of conducting a simultaneous search for continuous GWs and the stochastic GWB. We will also compare our work to other recent efforts to speed up PTA GW data analysis and discuss the importance of our first-order likelihood method.

6.4.1 Simultaneous Detection of Continuous GWs and a Stochastic GWB

One very important feature of the first order likelihood method is that it can also be applied to searches for continuous GWs. This will allow us to simultaneously search for a correlated stochastic background and resolve individual sources that are bright enough to stand out above such a background. In standard continuous GW searches using PTAs (Babak & Sesana 2012; Ellis et al. 2012c; Petiteau et al. 2013) the assumption is made that any detectable single source will be bright enough such that the noise (e.g stochastic GWB) can be approximated as a gaussian process that is uncorrelated among pulsars. However, recent work (Ravi et al. 2012) has shown that we are likely to see a few single sources per frequency bin that will stand out from the typical isotropic stochastic background, thus in order to resolve the weakest of these it is crucial to simultaneously search

for a correlated stochastic background as well as the continuous source. We can then write down a combined likelihood function assuming a deterministic source of functional form $\mathbf{s}(\vec{\lambda})$

$$p(\mathbf{r}|\vec{\theta},\vec{\lambda}) = \frac{1}{\sqrt{\det 2\pi\mathbf{\Sigma}}}\exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{s})^T\mathbf{\Sigma}^{-1}(\mathbf{r}-\mathbf{s})\right), \tag{6.4.1}$$

where our noise (including the stochastic background) parameters are $\vec{\theta}$ and our single source parameters are $\vec{\lambda}$. Using our first order likelihood approach we can approximate Eq. (6.4.1) as

$$\begin{aligned}\ln p(\mathbf{r}|\vec{\theta},\vec{\lambda}) =&\approx -\frac{1}{2}\left[\mathrm{Tr}\ln\mathbf{P} + (\mathbf{r}-\mathbf{s})^T\mathbf{P}^{-1}(\mathbf{r}-\mathbf{s}) - (\mathbf{r}-\mathbf{s})^T\mathbf{P}^{-1}\mathbf{S}_c\mathbf{P}^{-1}(\mathbf{r}-\mathbf{s})\right]\\ &= -\frac{1}{2}\sum_{\alpha=1}^{M}\left[\mathrm{Tr}\ln P_\alpha + (r_\alpha-s_\alpha)^T P_\alpha^{-1}(r_\alpha-s_\alpha) - \sum_{\beta\neq\alpha}^{M}(r_\alpha-s_\alpha)^T P_\alpha^{-1}S_{\alpha\beta}P_\beta^{-1}(r_\beta-s_\beta)\right].\end{aligned} \tag{6.4.2}$$

As in the stochastic background case, this again will speed up computations because we only have to invert the *individual* auto-covariance matrices as opposed to the *full* data covariance matrix. Although there have been proposed methods to speed up the computation of the stochastic likelihood function (van Haasteren 2013), this is not applicable to continuous sources because it relies on essentially applying a low pass filter to the data. However, since we expect continuous sources across the entire frequency band (with higher frequency sources possibly standing out above the background) we must keep all frequency information. Therefore our first order likelihood approximation is a viable option when looking to significantly speed up computation time while losing minimal information about potential GW signals.

As always, to claim a detection we must do some sort of model comparison, be it a Neyman-Pearson test for Frequentist statistics or an odds ratio or Bayes factor for Bayesian statistics. For example if we want to assess the likelihood of that a continuous GW is in our data we want to compute the following Bayes factor

$$\mathcal{B} = \frac{\mathcal{Z}_{\mathrm{CW}}}{\mathcal{Z}_{\mathrm{noise}}} = \frac{\int\int d\vec{\lambda}d\vec{\theta}p(\mathbf{r}|\vec{\theta},\vec{\lambda})p(\vec{\lambda})p(\vec{\theta})}{\int d\vec{\theta}p(\mathbf{r}|\vec{\theta})p(\vec{\theta})}, \tag{6.4.3}$$

where $\mathcal{Z}_{\mathrm{CW}}$ and $\mathcal{Z}_{\mathrm{noise}}$ are the evidence for the gravitational CW and noise models, respectively. However, notice that $\vec{\theta}$ depends on our stochastic GWB parameters as we treat all stochastic processes as "noise" in this analysis. If we do not include the GWB parameters

in the model then we could mistake a low frequency GWB for a single continuous source, thus including the GWB stochastic background in both models is crucial to detection and eventually characterization of a single GW source. We should also mention that the biases mentioned in section 6.3.3 are not as important if we simply wish to let the noise parameters vary along with the single source parameters since these noise parameters will be marginalized over in the end. An exploration of these combined searches will be the subject of a future paper.

6.4.2 Comparison with Other Work

Recently there have been three studies devoted to making the analysis of PTA data more computationally efficient. First, van Haasteren (2013, hereafter vH13) have developed a method dubbed Acceleration By Compression (ABC) to speed up this analysis. The main point of that work is to write the data in a compressed basis, keeping the minimum number of basis vectors to maximize the ability to characterize a correlated red signal. vH13 also makes use of an interpolation scheme to compute the covariance matrix which further improves the efficiency of the algorithm at the cost of large memory usage. The aforementioned method has proved to be very efficient and accurate in setting upper limits on the stochastic GWB and characterizing injected signals, however, since it relies on a reduced basis that essentially "throws away" high frequency information it is impossible to obtain a reliable Bayes Factor when comparing models that allow for varying white noise components. Since our first-order likelihood function makes use of all the information in the data we can indeed compute reliable Bayes factors and make confident statements about detection. We note however that the first-order likelihood of this work and the ABC method of vH13 are complementary. The two methods can in principle be combined for even greater efficiency.

Most recently there have been two analyses of the IPTA MDC that aim to make the PTA data analysis more efficient. First, Lentati et al. (2013b) have developed a novel model-independent method for the estimation of the spectral properties of an isotropic stochastic GWB. It makes use of a frequency domain approach and is extremely efficient and results in computational speedups of two to three orders of magnitude over the full

likelihood implementation. It has also been extensively tested on the MDC datasets and has proved to be very accurate in characterizing the stochastic GWB. Our first order likelihood method is indeed complementary to the aforementioned work as it provides a way to efficiently evaluate the likelihood function in a full time domain analysis which will be vital for cross-checks of real-life detection candidates.

Finally, Taylor et al. (2012) have implemented the full VHML likelihood function and have made it more efficient through the use of optimized linear algebra libraries with multithreading and parallelization resulting in significant speedups in the likelihood evaluation. However, all of these methods could just as well be applied to the first-order likelihood which would still be more efficient than the full likelihood by a factor proportional to the number of pulsars in the array.

This work and recent work have shown that there has indeed been significant progress on making the likelihood evaluation more efficient for pulsar timing arrays. All of these methods are complementary and will provide important cross checks for future stochastic GWB detection candidates.

6.4.3 Summary

In this chapter we have introduced a novel way to speed up the computation of the likelihood function for PTAs when searching for a stochastic GWB. This was accomplished by expanding the likelihood function to first order in the Hellings and Downs correlation coefficients expected for a stochastic GWB leading to a computational speedup on the order of the square of the number of pulsars in the PTA. For typical PTAs this results in a speed-up of a few hundred to about a thousand. We have briefly discussed the implementation of this technique on the first IPTA Mock Data Challenge and showed that this algorithm performs well in extracting the injected GWB parameters and making a significant detection through various Bayes factors. Though this is indeed an approximation to the full likelihood function we have shown through extensive simulations that the bias introduced in the estimation of GWB parameters is minimal and negligible in many cases. This was accomplished through an analytical computation of the expectation value

of the maximum likelihood, direct comparisons of the full and first-order likelihood functions on simulated data sets and through a statistical Monte-Carlo approach based on the Empirical Distribution Function. Although this work has focused solely on the detection and characterization of a stochastic GWB, this likelihood function can also be used to estimate the intrinsic red and white noise parameters of individual pulsars simultaneously with the GWB parameters.

Appendix 6.A Relationship to VHML likelihood

Making use of Eq. (3.2.7), the likelihood function for the noise can be written as

$$p(\mathbf{n}|\vec{\theta}) = p(\mathbf{r}|\vec{\theta}, \delta\boldsymbol{\xi}_{\rm best}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_n)}} \times \exp\left(-\frac{1}{2}(\mathbf{r} - \mathbf{M}\delta\boldsymbol{\xi}_{\rm best})^T\boldsymbol{\Sigma}_n^{-1}(\mathbf{r} - \mathbf{M}\delta\boldsymbol{\xi}_{\rm best})\right). \tag{6.A.1}$$

This can be thought of as a *conditional* pdf, where the values of $\delta\boldsymbol{\xi}_{\rm best}$ are fixed. In van Haasteren & Levin (2013) it was shown that the marginalized likelihood can be written as

$$p(\mathbf{r}|\vec{\theta}) = \int d\delta\boldsymbol{\xi}\, p(\mathbf{r}|\vec{\theta}, \delta\boldsymbol{\xi}) = \frac{\exp\left[-\frac{1}{2}\mathbf{r}^T\mathbf{G}^T\left(\mathbf{G}^T\boldsymbol{\Sigma}_n\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{r}\right]}{\sqrt{\det 2\pi\mathbf{G}^T\boldsymbol{\Sigma}_n\mathbf{G}}}, \tag{6.A.2}$$

where $\mathbf{G}$ is the matrix constructed from the final $(N - N_{\rm fit})$ columns of the matrix $\mathbf{U}$ in the singular value decomposition of the design matrix, $\mathbf{M} = \mathbf{USV}^T$.

We will now explore the G matrix and the R matrix obtained from the marginalized and conditional pdfs, respectively. As mentioned above, R can be thought of as an oblique projection operator that projects the pre-fit residuals into the post-fit residual space, whereas G^T can be thought of a projection operator that projects our data onto the null space of M, that is, it projects the data into a subspace orthogonal to the timing model fit. Since R is not generally symmetric and therefore is an oblique projection operator, it does not have such a simple mathematical interpretation. However, we can recast our problem in terms of "weighted" residuals then we have the following transformations: $r \to Wr$, $M \to WM$, and $R \to W^{-1}RW$, where W is the weighting matrix defined above. In this case minimizing the chi-squared becomes an unweighted least squares problem and we obtain the exact same estimates of $\delta\boldsymbol{\xi}_{\rm best}$ and likelihood function as before. In this case

R is symmetric and can be thought of as an orthogonal projection operator that projects our weighted data onto the null space of the weighted timing model (WM). However, in order to compute the likelihood we still have to invert the covariance matrix $\Sigma_r = R\Sigma_n R^T$ which is singular. To do this we rely on the pseudo-inverse. The pseudo-inverse of Σ_r is easiest defined in terms of its eigen-decomposition $\Sigma_r = EDE^T$, with E the matrix of eigenvectors of Σ_r, and D the diagonal matrix with $D_{ii} = \lambda_i$ the eigenvalues of Σ_r. It so happens that for a symmetric positive semi-definite matrices like these, the eigen-decomposition is also the singular value decomposition (SVD). The pseudo-inverse of Σ_r is then

$$\overline{\Sigma_r^{-1}} = E\overline{D^{-1}}D^T, \tag{6.A.3}$$

where the overbar indicates that we are taking a pseudo-inverse and $\overline{D^{-1}}_{ii} = 1/\lambda_i$ for $\lambda > 0$ and $\overline{D^{-1}}_{ii} = 0$ otherwise. Note that when all the error bars are the same (i.e. $W = \sigma^{-1}\mathbb{I}$ with σ constant), the matrix $G^T\Sigma_n G$ has the same eigenvalues as the non-singular part of $R\Sigma_n R^T$ and we have

$$\overline{(R\Sigma_n R^T)^{-1}} = G(G^T\Sigma_n G)^{-1}G^T. \tag{6.A.4}$$

Thus we have obtained a very interesting result that in the case of uniform uncertainties, the conditional pdf making use of a pseudo-inverse is equivalent to the marginalized pdf making use of the projection matrix G^T. However, in general this is not true and the two methods are indeed different. Although, in many cases the uncertainties are similar on a majority of the TOAs, thus the two methods will not differ much in practice.

Chapter 7

Discussion and Conclusions

"*We ain't found shit!*"

— Trooper, *Spaceballs*

7.1 Summary

This dissertation has presented a comprehensive set of tools for the detection and characterization of continuous GWs, a stochastic GW background, and intrinsic noise processes present in the pulsar timing residuals. Throughout this dissertation, we have stressed that a more Bayesian approach where all parameters including timing model, noise, and GW parameters be allowed to vary simultaneously in the search is the most robust. This is not to say that frequentist methods such as those derived in Chapter 3 are not useful. Since the frequentist statistics are generally much less computationally expensive they serve as excellent tools to provide valuable information that can then be fed into the larger Bayesian analysis. Such frequentist tools can also serve as a proxy to a full analysis when doing a very large set of simulations to assess the overall sensitivity of PTAs.

The work presented in this dissertation has had a significant impact in the field of GW detection using pulsar timing.

- The noise modeling presented in Chapter 2 has allowed us to study the noise in pulsar timing data in more detail than in the past. Furthermore, since we include the (linearized) timing model parameter and noise parameters simultaneously in our analysis, we take into account any correlations between timing model parameters

and the noise. This has been taken a step further in full Bayesian timing packages that use the full non-linear timing model and full noise model to construct probability distributions of all timing and noise parameters (Lentati et al. 2013a; Vigeland & Vallisneri 2013). Most recently, and for the first time, the 9-year NANOGrav data release will use results from this noise analysis when determining the timing model parameters and uncertainties, leading to more realistic uncertainties on the timing model parameters.

- The $\mathcal{F}$-statistic approach discussed in Chapter 3 has become a standard tool in the continuous wave detection toolbox. This technique has been applied to the 5-year NANOGrav data release (Chapter 5) and will be part of the upcoming IPTA continuous wave analysis. Recently, the idea of maximizing over GW parameters has been taken a step further in Taylor et al. (2013) where the pulsar induced phase is numerically maximized/marginalized, which greatly reduces computational time and the size of the parameter space.

- The Bayesian analysis pipeline presented in Chapter 4 is now the basis for the current Bayesian continuous pipelines. More importantly, this work has led to the development of a fully functional and quite general MPI-enabled parallel tempering MCMC[1] that is now implemented in several Bayesian analysis pipelines throughout NANOGrav and the EPTA.

- The continuous GW analysis presented in Chapter 5 is the first full NANOGrav collaboration paper on the topic of continuous GWs. Although there is no evidence for any continuous GWs in this data set, we have presented the most constraining upper limit on the strength of such GWs to date. This work is also the first that has carried out a full Bayesian analysis including pulsar timing parameters, GW parameters and noise parameters simultaneously. Lastly, the methods used for this analysis will for the basis for the upcoming IPTA continuous GW analysis.

- The first-order likelihood expansion technique of Chapter 6 was one of the first robust techniques to significantly reduce the computational burden of full PTA

[1] `https://github.com/jellis18/PAL2`

Bayesian analyses. Since its development, other techniques (van Haasteren 2013; Lentati et al. 2013b) have proven more efficient; however, when all of these methods are combined, we achieve the greatest computational savings.

7.2 Prospects for Detection of the Stochastic GW Background

A question that is of particular interest is "When do PTAs expect a GW detection?". This is, of course a very difficult question to answer as it depends on a slew of uncertainties such as the overall population and rate of SMBHB mergers, the physics of SMBMB mergers (it is not entirely GW driven at large orbital separations), the rate and quality of newly discovered pulsars, the timing stability of our currently timed pulsars, and the availability of observing facilities. While a clear and concise answer to this question is impossible, we can make an estimate using reasonable assumptions about the above uncertainties. Here we will focus on the NANOGrav PTA and only the stochastic GW background. A complete and more detailed analysis will be published in a future paper, here we simply summarize our results.

We simulate a NANOGrav timing program through 2025 assuming that for each year past 2014 we add 4 pulsars to the program (two at the Green Bank Telescope and two at the Arecibo Radio Telescope) with RMS residuals equal to the median of the currently timed pulsars at both telescopes (272 ns at Arecibo and 323 ns at GBT). Up until 2014 we use the real measured TOA uncertainties for each pulsar. We simulated many realizations of this PTA with different values for the amplitude of the stochastic background (assuming circular GW driven binaries as in Chapter 1) and compute the upper limit on the amplitude that would be measured in the absence of any GW background and the amplitude at which we attain 50 and 90 percent detection probability. As our detection statistic we use the optimal cross correlation statistic of Anholm et al. (2009) that has been modified to take into account the timing model via the marginalization approach of Chapter 2.

Figure 27 shows the results of one set of simulations as described above. The red, green, and blue curves represent the 95% upper limit in the absence of any GW background, and the amplitude at which we attain 50% and 90% detection probability. The

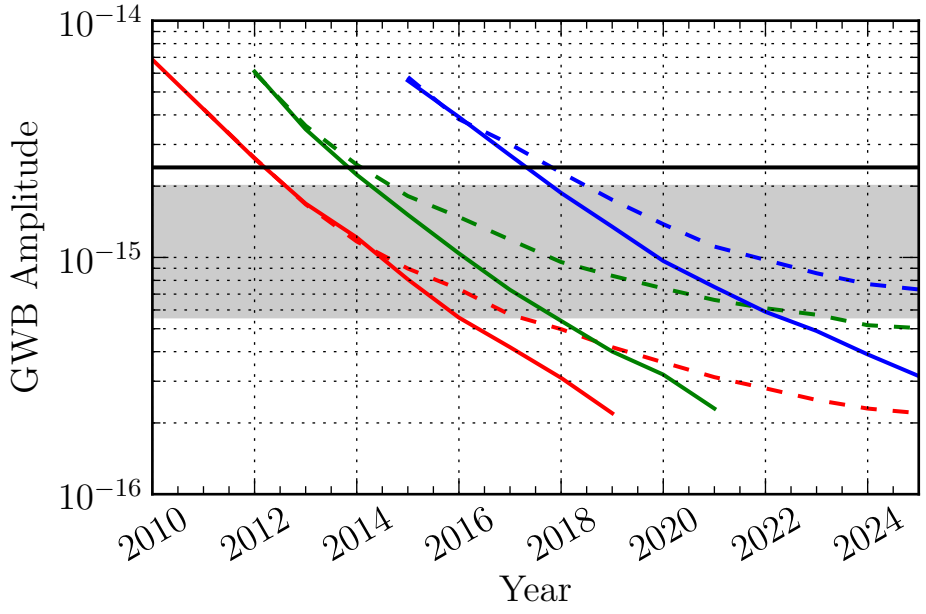


Figure 27 : GW background amplitude vs. year. The red, green, and blue curves represent the 95% upper limit in the absence of any GW background, and the amplitude at which we attain 50% and 90% detection probability, respectively. The solid and dashed lines represent 0 ns and 10 ns (at five years) of spin-noise (i.e., intrinsic red noise) (Shannon & Cordes 2010) in the data. The GW background is assumed to be from circular GW driven SMBHBs with characteristic strain spectrum derived in Chapter 1. The gray shaded region is the one-sigma uncertainty on the amplitude of the stochastic GW background from Sesana (2013b) and the horizontal black line shows the best published upper limit on the amplitude of the stochastic GW background (Shannon et al. 2013).

the solid and dashed lines represent 0 ns and 10 ns (at five years) of spin-noise (i.e., intrinsic red noise) (Shannon & Cordes 2010) in the data. This estimate of red noise is likely realistic as we see nearly no evidence of red noise in the 5-year NANOGrav data release. The gray shaded region is the one-sigma uncertainty on the amplitude of the stochastic GW background from Sesana (2013b) and the horizontal black line shows the best published upper limit on the amplitude of the stochastic GW background (Shannon et al. 2013). We see from the figure that in 2010 with 5 years of timing data, the upper limit from the simulations is nearly $A < 7 \times 10^{-15}$ which is excellent agreement with the results of Demorest et al. (2013). Furthermore, we see that in the most optimistic case of zero spin-noise we will have either made a detection with 90% confidence or will have ruled out the entire one-sigma uncertainty region on the GW amplitude by the year 2022. If there is significant red spin-noise then this level of sensitivity will be delayed by ~ 2 years and will grow much more slowly afterward due to the fact that our sensitivity will now only increase with the addition of new pulsars to the array in agreement with the analytical results of Siemens et al. (2013).

The simulations above address uncertainties in the timing stability of currently timed pulsars; however, it does not address the uncertainty in the overall stochastic GW background signal itself. Recently, a significant amount of work has demonstrated that the environment (i.e., stars and gas) of the SMBHs will play a large role in their evolution at large orbital separations (Sesana 2013a; McWilliams et al. 2012; Ravi et al. 2014, and references therein) and will generally lead to a decrease in signal power at low frequencies compared to the purely circular GW driven case. To address this we have performed a set of simulations using the same observing strategy described above, but now the GW background has a power spectrum that is indicative of these dynamic environmental effects. Figure 28 shows the corresponding upper limits and minimum detectable amplitudes for three different evolution scenarios for the SMBHBs. The first model assumes circular GW driven binaries as above (black dotted line in the left side of the figure), the second model assumes that stellar hardening dominates the binary evolution at large orbital separations but does not cause any eccentricities (solid red line in left side of figure), and the third model again assumes evolution driven by stellar hardening but that it now causes

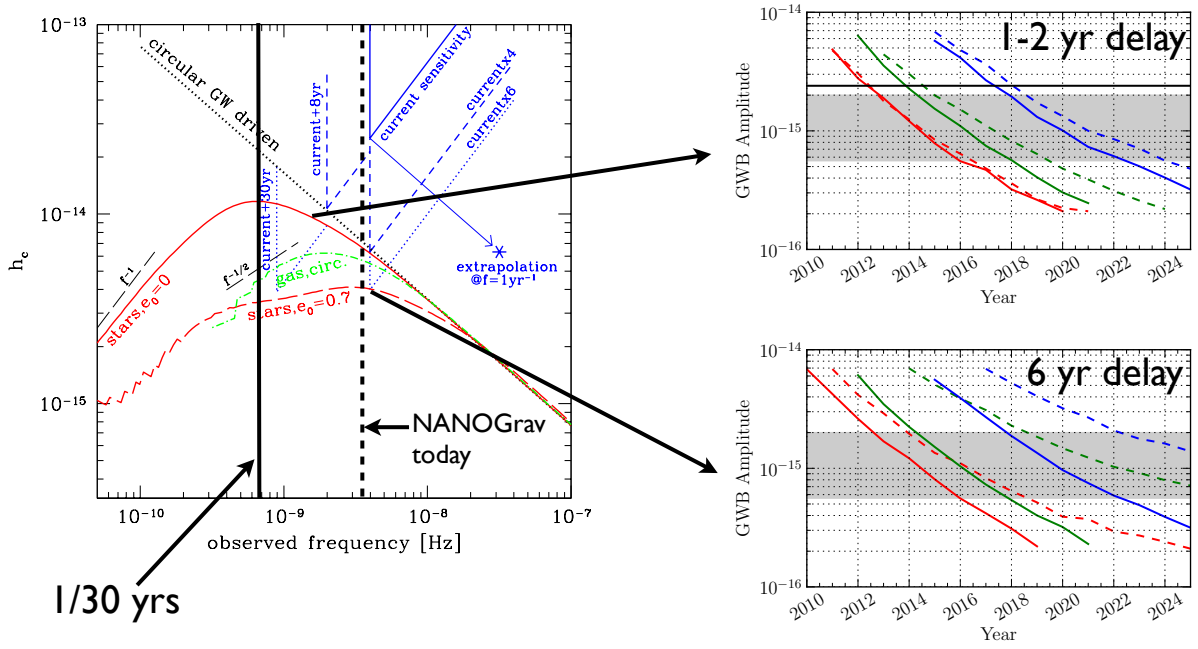


Figure 28 : **Left**: Characteristic strain amplitude vs. GW frequency for a range of different SMBHB evolution scenarios (Sesana 2013a). Particular attention should be focused on the dotted black line that shows the standard circular GW driven case and the solid and dashed red lines showing an evolutionary scenario that is dominated by stellar hardening at large orbital separations. The solid and dashed curves show initial eccentricities of the SMBHBs of 0 and 0.7, respectively. **Right**: GWB amplitude vs. time. The simulations are identical to those described above but now using different GW background distributions. The top right shows the difference between circular GW driven SMBHBs (solid lines) and stellar driven SMBHBs with 0 eccentricity (dashed lines). The bottom right plot shows the difference between circular GW driven SMBHBs (solid lines) and stellar driven SMBHBs with 0.7 initial eccentricity (dashed lines).

all binaries to enter the PTA band with eccentricities of 0.7 (dashed red line in the left side of the figure). The right side of the figure shows the results in which we see that a stellar hardening with 0 eccentricity only delays our circular GW driven estimates by ~ 2 years, whereas a stellar driven evolution resulting in high eccentricities significantly changes our sensitivity, resulting in delays of ~ 6 years relative to the circular GW driven case. Even though there are large uncertainties in the physical effects and efficiency of stellar hardening, assuming that *all* binaries have eccentricity equal to 0.7 when they enter the GW dominated regime is pessimistic and rather unlikely.

We have now addressed some of the uncertainties in the timing stability of pulsars and the uncertainties in SMBHB evolution, lastly, we turn to uncertainties in the availability of observing facilities. Here we investigate different scenarios in which we lose access to one of our two telescopes, either the GBT or Arecibo. In the following simulations we assume that if we lose access to the GBT then all pulsars timed at the GBT will cease being timed completely and we will only add two pulsars per year at Arecibo. This is mainly due to the limited declination range of the Arecibo Radio telescope ($-1^\circ 20' < \delta < +38^\circ 02'$). In the case where we lose access to the Arecibo telescope we assume that we could move all pulsars timed at Arecibo to the GBT at the cost of reducing the observing cadence by a factor of two for all pulsars. In this case, however; we still assume the addition of 4 pulsars per year, two with the GBT and two others from our IPTA collaborators. Figure 29 again shows our 95% upper limits and minimum detectable amplitudes at 50% and 90% detection probability for both telescopes (solid lines), only GBT (dashed-dotted), and only Arecibo (dashed). Here we assume that we would lose access to the GBT after 2015 and would lose access to the Arecibo telescope after 2016. We see that since we are still able to add 4 pulsars per year in the GBT-only case, we do not lose as much sensitivity as in the only Arecibo case, where we can only add two pulsars per year. In either case; however, we are still able to either detect the GW background or rule out a significant area of parameter space by 2025.

In summary, we have shown that although there is a large amount of uncertainty in the long-term timing stability of pulsars, the exact astrophysics that drive SMBHBs to merger, and the availability of observing facilities, there is a very good chance that we

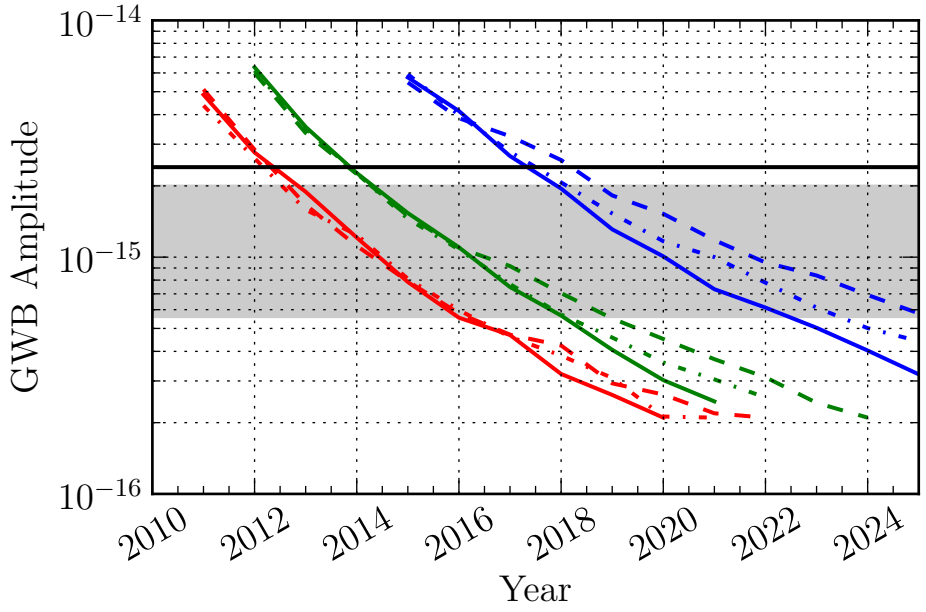


Figure 29 : GW background amplitude vs. year. The observing scenario is described in the text above. The solid lines represent our usual observing scenario using both telescope and the dash-dotted and dashed lines represent the scenarios where we lose access to the GBT and Arecibo, respectively.

will either make a detection or will be able to make significant astrophysical statements by ruling out a large area of parameter space in a ten year time frame.

7.3 Further Work

In order to expand on the work presented in this thesis, several further lines of research are either planned or ongoing.

Robust noise modeling: As was mentioned in Chapter 2, we currently have a quite complex noise model that includes separate parameters for different back-ends/frequency combinations. We also have the ability to model red noise and DM variations in several different ways. There are many free parameters and different models to choose from. We could just run with the most complicated noise model and see which parameters have support from the data but it would be optimal if we could easily compare different models via the Bayes factor. This noise modeling will not only help with GW detection efforts but will also yield more reliable timing model parameter fits and uncertainties. This type of analysis is currently underway

and we are striving to make the process more automated.

Combined GW searches: Currently all published GW searches deal specifically with one kind of signal, either continuous, stochastic or burst. Furthermore, all published continuous wave and burst analyses assume that the stochastic GW background will be weak enough to treat the data from each pulsar as independent. A much more robust and sensitive search pipeline would include all kinds of GW signals simultaneously. This will make the complexity of such an analysis more difficult because of the increased dimensionality and the likely correlations between different kinds of signals if weak. Such methods are currently being developed and are functional in existing data analysis pipelines[2,3].

More efficient sampling techniques: As the number of pulsars in PTAs grow and noise models and GW models become more complicated, the dimensionality of our parameter space becomes very large. It is very challenging to efficiently sample such large parameter spaces and even more difficult to effectively evaluate the Bayesian evidence which will be required for a confident detection statements. Currently we make use of parallel tempering and several other convergence aids that greatly improve our ability to explore the space; however, we still require faster convergence if we ever hope to include the full non-linear timing model and GWs in the same model (this is the "holy grail" of PTA data analysis). Techniques such as Guided Hamiltonian sampling (see e.g. Lentati et al. 2013b, and references therin) or Reverse-Jump MCMC (Green 1995) may prove useful in this regard.

More complicated GW signal models: While current data analysis techniques are capable of making a confident GW detection, many simplifying assumptions in the modeling of GWs make these analysis methods sub-optimal when considering these more complicated models. Specifically, we need to relax two such assumptions; circular orbits and single sources. Recent work has shown that SMBHBs may have

[2] `https://github.com/jellis18/PAL2`

[3] `https://github.com/vhaasteren/piccard`

a non-negligible eccentricity (Roedig & Sesana 2012) and that they may be distributed anisotropically (Ravi et al. 2012). This can be accomplished by modifying existing codes to include eccentricity using both the earth and the pulsar term. Inclusion of the pulsar term is crucial for parameter estimation studies as it will allow one to measure the evolution of the eccentricity over thousands of years, which in principle could tell us about the stellar and gas content of the SMBHB system in the past, leading to an increased understanding of galaxy mergers and SMBHB system dynamics.

Multimessenger Astronomy: While nearly all current efforts within the PTA community are focused on the *detection* of GWs, it is also critical to determine what kind astrophysical information we can extract from GW observations with PTAs *after* the detection has been made. One could perform simulations of advanced PTA datasets with the best possible GW models to answer questions such as: Is it possible to use these data to distinguish general relativity from other theories of gravity? How accurately can we measure intrinsic black hole parameters such as spin and mass? Can we determine the sky location of SMBHBs accurately enough to do a targeted electromagnetic (EM) followup? Furthermore, several studies (Sesana 2013a; McWilliams et al. 2012; Ravi et al. 2014, and references therein) have suggested that the characteristic strain spectrum of the stochastic GW background will deviate from a power-law at low frequencies. A study could be performed to determine if it is possible to extract any astrophysical information from this spectral break using PTA data. In a similar manner one could test whether or not it is possible to definitively determine the origin of the stochastic GW background through observations.

7.4 Closing Remarks

As PTAs continue to add more pulsars and increase timing precision the likelihood of a GW detection increases. This is a very important time in the history of GW detection using PTAs as our sensitivity is allowing us to place astrophysically meaningful constraints

on specific models of structure formation and black hole host relations. Furthermore, it is an exciting time for a data analyst in general as several new data sets are about to be released with the best timing precision yet attained. With increasing computational power and more realistic models, we are able to probe deeper than ever before into pulsar timing data in order to give the most accurate and reliable estimates of pulsar timing parameters and noise values allowing us to do more important astrophysics. Lastly, as the IPTA collaboration grows more mature we will have a less restricted flow of ideas allowing us to use combined datasets and several independent data analysis methods to make ever increasing *confident* statements about the presence of GWs in our data.

Bibliography

Allen, B., & Romano, J. D. 1999, Phys. Rev. D, 59, 102001

Andrieu, C., & Thoms, J. 2008, Statistics and Computing, 18, 343

Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, Phys. Rev. D, 79, 084030

Babak, S., & Sesana, A. 2012, Phys. Rev. D, 85, 044034

BICEP2 Collaboration et al. 2014, ArXiv e-prints

Braak, C. 2006, Statistics and Computing, 16, 239

Burke-Spolaor, S. 2013, Classical and Quantum Gravity, 30, 224013

Burt, B. J., Lommen, A. N., & Finn, L. S. 2011, ApJ, 730, 17

Caprini, C., Durrer, R., & Siemens, X. 2010, Phys. Rev. D, D82, 063511

Chamberlin, S., Creighton, J., Ellis, J., Price, L., Romano, J., & Siemens, X. 2013, in preparation

Chamberlin, S. J., & Siemens, X. 2012, Phys. Rev. D, 85, 082001

Champion, D. J., et al. 2010, ApJL, 720, L201

Coles, W., Hobbs, G., Champion, D. J., Manchester, R. N., & Verbiest, J. P. W. 2011, MNRAS, 418, 561

Corbin, V., & Cornish, N. J. 2010, arXiv:1008.1782

Cordes, J., & Jenet, F. 2012, Astrophys.J., 752, 54

Cordes, J., & Shannon, R. 2012, Astrophys.J., 750, 89

Cordes, J. M., & Lazio, T. J. W. 2002, ArXiv Astrophysics e-prints

—. 2003, ArXiv Astrophysics e-prints

Cordes, J. M., & Shannon, R. M. 2010

—. 2012, ApJ, 750, 89

Cornish, N. J., & Crowder, J. 2005, Phys. Rev. D, 72, 043005

Cornish, N. J., & Porter, E. K. 2007, Classical and Quantum Gravity, 24, 5729

Cutler, C. 1998, Phys. Rev. D, 57, 7089

Cutler, C., & Schutz, B. F. 2005, Phys. Rev. D, 72, 063006

Danzmann, K., & Rdiger, A. 2003, Classical and Quantum Gravity, 20, S1

Demorest, P. B. 2007, PhD thesis, University of California, Berkeley

Demorest, P. B., et al. 2013, ApJ, 762, 94

Detweiler, S. 1979, ApJ, 234, 1100

DuPlain, R., Ransom, S., Demorest, P., Brandt, P., Ford, J., & Shelton, A. L. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7019, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, MNRAS, 372, 1549

Ellis, J., Siemens, X., & Chamberlin, S. 2012a, arXiv:1210:5274

Ellis, J. A. 2013, Classical and Quantum Gravity, 30, 224004

Ellis, J. A., Jenet, F. A., & McLaughlin, M. A. 2012b, ApJ, 753, 96

Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012c, ApJ, 756, 175

Ellis, J. A., Siemens, X., & van Haasteren, R. 2013, ApJ, 769, 63

Ellis, J. A., Vallisneri, M., van Haasteren, R., & Siemens, X. 2014, in preparation

Enoki, M., Inoue, K. T., Nagashima, M., & Sugiyama, N. 2004, ApJ, 615, 19

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Finn, L. S., & Lommen, A. N. 2010, ApJ, 718, 1400

Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300

Gelman, A., Roberts, G., Gilks, W., Bernardo, J., Berger, J., Dawid, A., & Smith, A. 1996, Bayesian Statistics

Gelman, A., & Rubin, D. B. 1992, Statistical Science, 7, pp. 457

Green, P. J. 1995, Biometrika, 82, 711

Gültekin, K., et al. 2009a, ApJ, 698, 198

—. 2009b, ApJ, 698, 198

Guo, Q., et al. 2011, MNRAS, 413, 101

Haario, H., Saksman, E., & Tamminen, J. 2001, Bernoulli, 7, 223

—. 2005, Computational Statistics, 20, 265

Harry, I. W., Allen, B., & Sathyaprakash, B. S. 2009, Phys. Rev. D, 80, 104014

Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, 217, 709

Hobbs, G., et al. 2010, Classical and Quantum Gravity, 27, 084013

—. 2012, MNRAS, 427, 2780

Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655

Hotan, A. W., van Straten, W., & Manchester, R. N. 2004, Publications of the Astronomical Society of Australia, 21, 302

Hulse, R. A., & Taylor, J. H. 1975, ApJL, 195, L51

Iguchi, S., Okuda, T., & Sudou, H. 2010, ApJL, 724, L166

Jaffe, A. H., & Backer, D. C. 2003, ApJ, 583, 616

Janssen, G. H., Stappers, B. W., Kramer, M., Purver, M., Jessner, A., & Cognard, I. 2008, in American Institute of Physics Conference Series, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, ed. C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi, 633–635

Jaranowski, P., & Królak, A. 2000, Phys. Rev. D, 61, 062001

—. 2005, Living Reviews in Relativity, 8, 3

Jaranowski, P., Królak, A., & Schutz, B. F. 1998, Phys. Rev. D, 58, 063001

Jeffreys, H. 1961, Theory of Probability, 3rd edn. (Oxford, England: Oxford)

Jenet, F., et al. 2009, arXiv:0909.1058

Jenet, F. A., Hobbs, G. B., Lee, K. J., & Manchester, R. N. 2005, ApJL, 625, L123

Jenet, F. A., et al. 2006, ApJ, 653, 1571

Jenet, F. A., Lommen, A., Larson, S. L., & Wen, L. 2004, ApJ, 606, 799

Ju, W., Greene, J. E., Rafikov, R. R., Bickerton, S. J., & Badenes, C. 2013, ApJ, 777, 44

Kaspi, V. M., Taylor, J. H., & Ryba, M. F. 1994, ApJ, 428, 713

Keith, M. J., et al. 2013, MNRAS, 429, 2161

Kocsis, B., & Sesana, A. 2011, MNRAS, 411, 1467

Lauer, T. R., Tremaine, S., Richstone, D., & Faber, S. M. 2007, ApJ, 670, 249

Lazio, T. J. W. 2013, Classical and Quantum Gravity, 30, 224011

Lee, K., Jenet, F. A., Price, R. H., Wex, N., & Kramer, M. 2010, ApJ, 722, 1589

Lee, K. J., Jenet, F. A., & Price, R. H. 2008, ApJ, 685, 1304

Lee, K. J., Wex, N., Kramer, M., Stappers, B. W., Bassa, C. G., Janssen, G. H., Karuppusamy, R., & Smits, R. 2011, MNRAS, 414, 3251

Lentati, L., Alexander, P., Hobson, M. P., Feroz, F., van Haasteren, R., Lee, K., & Shannon, R. M. 2013a, arXiv:1310.2120

Lentati, L., Alexander, P., Hobson, M. P., Taylor, S., Gair, J., Balan, S. T., & van Haasteren, R. 2013b, Phys. Rev. D, 87, 104021

Littenberg, T. B. 2011, Phys. Rev. D, 84, 063009

Littenberg, T. B., & Cornish, N. J. 2009, Phys. Rev. D, 80, 063007

—. 2010, Phys. Rev. D, 82, 103007

Lommen, A. N. 2002, in Neutron Stars, Pulsars, and Supernova Remnants, ed. W. Becker, H. Lesch, & J. Trümper, 114–+

Lommen, A. N., & Backer, D. C. 2001, ApJ, 562, 297

Lorimer, D., & Kramer, M. 2005, Cambridge Observing Handbooks for Research Astronomers, Vol. 4, Handbook of Pulsar Astronomy, 1st edn. (Cambridge, U.K.; New York, U.S.A: Cambridge University Press)

Lorimer, D. R. 2008, Living Reviews in Relativity, 11, 8

Lynch, R. S., & Green Bank North Celestial Cap Survey Collaborations. 2013, in IAU Symposium, Vol. 291, IAU Symposium, ed. J. van Leeuwen, 41–46

Lyne, A., Hobbs, G., Kramer, M., Stairs, I., & Stappers, B. 2010, Science, 329, 408

Manchester, R. N. 2008, in American Institute of Physics Conference Series, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, ed. C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi, 584–592

Manchester, R. N., et al. 2013, Publications of the Astronomical Society of Australia, 30, 17

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993

McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2012, ArXiv e-prints

McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2012

Messenger, C., Prix, R., & Papa, M. A. 2009, Phys. Rev. D, 79, 104017

Mingarelli, C., Grover, K., Sidery, T., Smith, R., & Vecchio, A. 2012, Phys.Rev.Lett., 109, 081104

Olmez, S., Mandic, V., & Siemens, X. 2010, Phys.Rev., D81, 104028

Perrodin, D., et al. 2013a, ArXiv e-prints

—. 2013b, in preparation

Petiteau, A., Babak, S., Sesana, A., & de Araújo, M. 2013, Phys. Rev. D, 87, 064036

Phinney, E. S. 2001, ArXiv Astrophysics e-prints

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C (2nd ed.): the art of scientific computing (New York, NY, USA: Cambridge University Press)

Prix, R. 2007, Classical and Quantum Gravity, 24, 481

Ravi, V., Wyithe, J. S. B., Hobbs, G., Shannon, R. M., Manchester, R. N., Yardley, D. R. B., & Keith, M. J. 2012

Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. 2014, ArXiv e-prints

Roedig, C., & Sesana, A. 2012, Journal of Physics Conference Series, 363, 012035

Romani, R. 1989, in NATO ASI Series, Series C, Vol. 262, Timing Neutron Stars, ed. H. Ögelman & E. van den Heuvel (Dordrecht, Netherlands; Boston, U.S.A.: Kluwer), 113–117

Sathyaprakash, B. S., & Schutz, B. F. 2009, Living Reviews in Relativity, 12, 2

Sazhin, M. V. 1978, Soviet Ast., 22, 36

Sesana, A. 2013a, arXiv:1307.4086

—. 2013b, MNRAS, 433, L1

Sesana, A., Roedig, C., Reynolds, M. T., & Dotti, M. 2012, MNRAS, 420, 860

Sesana, A., & Vecchio, A. 2010, Phys. Rev. D, 81, 104008

Sesana, A., Vecchio, A., & Colacino, C. N. 2008, MNRAS, 390, 192

Sesana, A., Vecchio, A., & Volonteri, M. 2009, MNRAS, 394, 2255

Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607

Shannon, R. M., et al. 2013, Science, 342, 334

Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, Classical and Quantum Gravity, 30, 224015

Simon, J., Lommen, A., Stappers, B., Finn, L., Jenet, F., & Polin, A. 2013, in preparation

Stairs, I. H. 2003, Living Reviews in Relativity, 6, 5

Starobinsky, A. A. 1979, JETP Lett., 30, 682

Stinebring, D. R., Ryba, M. F., Taylor, J. H., & Romani, R. W. 1990, Physical Review Letters, 65, 285

Tanaka, T., Menou, K., & Haiman, Z. 2012, MNRAS, 420, 705

Tanaka, T. L., & Haiman, Z. 2013, Classical and Quantum Gravity, 30, 224012

Taylor, J., Wolszczan, A., Damour, T., & Weisberg, J. 1992, Nature, 355, 132

Taylor, J. H., & Weisberg, J. M. 1982, ApJ, 253, 908

Taylor, S., Ellis, J., & Gair, J. 2013, In Preparation

Taylor, S. R., & Gair, J. R. 2012, Phys. Rev. D, 86, 023502

Taylor, S. R., Gair, J. R., & Lentati, L. 2012

the LIGO Scientific Collaboration et al. 2013, ArXiv e-prints

Vallisneri, M. 2008, Phys. Rev. D, 77, 042001

—. 2012, Phys. Rev. D, 86, 082001

Valtonen, M. J., et al. 2008, 452, 851

van der Sluys, M., Mandel, I., Raymond, V., Kalogera, V., Röver, C., & Christensen, N. 2009, Classical and Quantum Gravity, 26, 204010

van der Sluys, M., Raymond, V., Mandel, I., Röver, C., Christensen, N., Kalogera, V., Meyer, R., & Vecchio, A. 2008, Classical and Quantum Gravity, 25, 184011

van Haasteren, R. 2013, MNRAS, 429, 55

van Haasteren, R., & Levin, Y. 2010, MNRAS, 401, 2372

—. 2013, MNRAS, 428, 1147

van Haasteren, R., et al. 2011, MNRAS, 414, 3117

van Haasteren, R., Levin, Y., McDonald, P., & Lu, T. 2009a, MNRAS, 395, 1005

—. 2009b, MNRAS, 395, 1005

Veitch, J., et al. 2012, Phys. Rev. D, 85, 104045

Verbiest, J. P. W., et al. 2009, MNRAS, 400, 951

Verbiest, J. P. W., Weisberg, J. M., Chael, A. A., Lee, K. J., & Lorimer, D. R. 2012, ApJ, 755, 39

Vigeland, S. J., & Vallisneri, M. 2013, arXiv:1310.2606

Volonteri, M., Haardt, F., & Madau, P. 2003, ApJ, 582, 559

Wahlquist, H. 1987, General Relativity and Gravitation, 19, 1101

Wald, R. M. 1984, General Relativity, 1st edn. (Chicago: University Of Chicago Press)

Waldman, S. 2011

Wen, Z. L., Jenet, F. A., Yardley, D., Hobbs, G. B., & Manchester, R. N. 2011, ApJ, 730, 29

Wen, Z. L., Liu, F. S., & Han, J. L. 2009, ApJ, 692, 511

Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 590, 691

Yardley, D. R. B., et al. 2011, MNRAS, 414, 1777

—. 2010, MNRAS, 407, 669

Justin A. Ellis

EDUCATION

7/2011–present **Ph.D Student in Physics**
Center for Gravitation, Cosmology and Astrophysics, UWM, Milwaukee WI

8/2009–6/2011 **Ph.D Student in Physics**
West Virginia University, Morgantown WV

1/2007–6/2009 **B.S. in Physics (Mathematics, Astronomy minor)**
West Virginia University, Morgantown WV

8/2004–12/2006 **B.S. Student in Mathematics**
Shepherd University, Shepherdstown WV

RESEARCH EXPERIENCE

7/2011–present **Graduate Research Assistant**
Center for Gravitation, Cosmology and Astrophysics, UWM, Milwaukee WI
Advisor: Dr. Xavier Siemens

8/2009–6/2011 **Graduate Research Assistant**
Dept. of Physics, West Virginia University, Morgantown WV
Advisor: Dr. Maura McLaughlin

6/2008–6/2009 **Undergaduate Research Assistant**
Dept. of Physics, West Virginia University, Morgantown WV
Advisor: Dr. Earl Scime

8/2005–12/2006 **Undergaduate Research Assistant**
Dept. of Environmental Sciences, Shepherd University, Shepherdstown WV
Advisor: Dr. Jason Best

RESEARCH INTERESTS

Detection of gravitational waves through pulsar timing, ground or space based interferometers; Bayesian inference; data analysis techniques; compact objects, including neutron stars, black holes, and white dwarfs; simulation of gravitational radiation from varying sources to gain relevant astrophysical insight; alternative theories of gravity; cosmology, dark matter, dark energy; early universe high energy physics.

TEACHING EXPERIENCE

8/2009–8/2010 Physics 111 (calculus based introductory physics) Lab Instructor
Dept. of Physics, West Virginia University

1/2010–8/2010 Physics 111 (calculus based introductory physics) Tutor
Dept. of Physics, West Virginia University

1/2008–6/2008 Physics 111 (calculus based introductory physics) Lab Instructor
Dept. of Physics, West Virginia University

8/2007–12/2007 Physics 101 (algebra based introductory physics) Lab Instructor
Dept. of Physics, West Virginia University

AWARDS & HONORS

2014	Einstein Fellowship (JPL/Caltech)
2013	Distinguished Dissertation Fellowship (UWM)
2013	NASA Wisconsin Space Grant Consortium Fellowship
2012	Blue Apple Award (22nd Midwest Relativity Meeting)
2012	NASA Wisconsin Space Grant Consortium Fellowship
2011	Best Graduate Student Oral Presentation (West Virginia Academy of Science)
2009	Outstanding Physics Senior Award (WVU)
2008	Reddy Scholarship for Academic Excellence
2008	Eberly College of Arts and Sciences Award for Academic Excellence
2004–2008	West Virginia Promise Scholarship

ORGANIZATIONS

2010–present	Member of IPTA (International Pulsar Timing Array)
2010–present	Member of AAS (American Astronomical Society)
2010–present	Member of ASAP (Arecibo Science Advocacy Partnership)
2009–present	Member of NANOGrav (North American Nano-hertz Observatory for Gravitational Waves)
2008–present	Member of the American Physical Society
2008–present	Member of Phi Sigma Theta National Honor Society
2007–present	Member of Golden Key National Honor Society

SELECTED CONFERENCES AND SEMINARS

International Pulsar Timing Array (IPTA) Collaboration Meeting, Banff, Canada, June 16–27 2014, "The IPTA continuous wave search project." (invited talk).

American Physical Society (APS) April Meeting, Savannah GA, April 5–8, 2014, "NANOGrav Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits" (talk)

American Astronomical Society (AAS) 223rd Meeting, Washington DC, January 6–10 2014, "I get by with a little help from my friends: Enhancing PTA sensitivity to GWs using EM counterparts" (talk).

CaJAGWR Seminar, Caltech, Pasadena CA, October 8 2013, "Gravitational Wave Detection with Pulsar Timing Arrays" (talk)

Jet Propulsion Laboratory Seminar, Pasadena CA, October 7 2013, "Gravitational Wave Detection with Pulsar Timing Arrays" (talk)

Wisconsin Space Grant Consortium Conference, Milwaukee WI, August 14 2013, "Searching for Gravitational Waves using Pulsar Timing Arrays" (talk).

10th Edoardo Amldi Conference on Gravitational Waves, Warsaw Poland, July 10 2013, "Continuous Gravitational Wave Search Methods and Results from PTAs" (talk).

Cambridge University Seminar, Cambridge UK, July 1 2013, "Searching for Gravitational Waves using Pulsar Timing Arrays" (invited talk).

International Pulsar Timing Array (IPTA) Collaboration Meeting, Krabi Beach Thailand, June 24–28 2013, "Single Source Detection and Upper Limits with IPTA data" (invited talk).

Physical Applications of Millisecond Pulsars, Aspen CO, January 19–25 2013, "When Will We Detect GWs?" (talk).

American Astronomical Society (AAS) 221st Meeting, Long Beach CA, January 6–11 2013, "Gravitational Wave Searches with Pulsar Timing Data" (talk).

22nd Midwest Relativity Meeting , Chicago, Illinois, September 28–29, 2012, "Gravitational Wave Searches with Pulsar Timing Arrays." (talk).

International Astronomy Union (IAU) Meeting, Beijing China, August 19–31, 2012, "Gravitational Wave Searches in Pulsar Timing Data" (talk).

International Pulsar Timing Array (IPTA) Collaboration Meeting, Kiama Australia, June 25–29, 2012, "An Overview of Single-Source Detection Algorithms" (invited talk).

American Astronomical Society (AAS) 219th Meeting, Austin Texas, January 8–12 2012, "Bayesian Methods for Covariance Estimation of Pulsar Timing Residuals" (poster).

21st Midwest Relativity Meeting , Urbana, Illinois, November 4–5 2011, "Detection Methods for Continuous Gravitational Waves Using Pulsar Timing Arrays" (talk).

9th Edoardo Amaldi Conference on Gravitational Waves, Cardiff, Wales, July 10–15 2011, "Detection Methods for Continuous GWs Using PTAs" (poster).

International Pulsar Timing Array (IPTA) Collaboration Meeting, Morgantown/Snowshoe, West Virginia, June 6–17 2011, "Detection of Continuous Gravitational Waves with Pulsar Timing Arrays" (talk).

American Astronomical Society (AAS) 217th Meeting, Seattle, Washington, January 9–12 2011,"Continuous Gravitational Wave Searches in Pulsar Timing Data" (poster).

International Pulsar Timing Array (IPTA) Collaboration Meeting, Leiden, Netherlands, June 21 – July 2 2010, "The impact of a stochastic gravitational-wave background on pulsar timing parameters" (poster).

50th Annual Meeting of the Division of Plasma Physics (DPP), Dallas Texas, November 17–21 2008, "A magneto-optic probe for magnetic fluctuation measurements" (poster).

PUBLICATIONS

S. Chamberlin, J. Creighton, **J. A. Ellis**, L. Price, J. Romano, and X. Siemens, "A Practical Application of the Optimal Statistic for Stochastic Gravitational Wave Background Searches in Pulsar Timing Data," *in preparation* (2014).

Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, J. M. Cordes, P. B. Demorest, X. Deng, T. Dolch, **J. A. Ellis**, R. D. Ferdman, L. S. Finn, N. Garver-Daniels, F. Jenet, G. Jones, V. M. Kaspi, M. Koop, M. Lam, T. J. W. Lazio, A. N. Lommen, D. R. Lorimer, J. Luo, R. S. Lynch, D. R. Madison, M. McLaughlin, S. T. McWilliams, D. J. Nice, N. Palliyaguru, T. T. Pennucci, S. M. Ransom, A. Sesana, X. Siemens, I. H. Stairs, D. R. Stinebring, K. Stovall, J. Swiggum, M. Vallisneri, R. van Haasteren, Y. Wang, and W. W. Zhu, "NANOGrav Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits," *arXiv:1404.1267* (2014).

X. Siemens, **J. A. Ellis**, F. Jenet, and J. D. Romano, "The stochastic background: scaling laws and time to detection for pulsar timing arrays, " *Classical and Quantum Gravity* **30** (2013) no. 22, 224015.

J. A. Ellis, "A Bayesian analysis pipeline for continuous GW sources in the PTA band," *Classical and Quantum Gravity* **30** (2013) no. 22, 224004.

J. A. Ellis, X. Siemens, and R. van Haasteren, "An Efficient Approximation to the Likelihood for Gravitational Wave Stochastic Background Detection Using Pulsar Timing Data," *ApJ* **769** (May, 2013) 63.

P. B. Demorest, R. D. Ferdman, M. E. Gonzalez, D. Nice, S. Ransom, I. H. Stairs, Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, J. M. Cordes, **J. A. Ellis**, L. S. Finn, P. Freire, S. Giampanis, F. Jenet, V. M. Kaspi, J. Lazio, A. N. Lommen, M. McLaughlin, N. Palliyaguru, D. Perrodin, R. M. Shannon, X. Siemens, D. Stinebring, J. Swiggum, and W. W. Zhu, "Limits on the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves," *ApJ* **762** (Jan., 2013) 94.

J. A. Ellis, X. Siemens, and S. Chamberlin, "Results of the First IPTA Closed Mock Data Challenge," `arXiv:1210.5274` (Oct., 2012) .

J. A. Ellis, X. Siemens, and J. D. E. Creighton, "Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays," *ApJ* **756** (Sept., 2012) 175.

J. Ellis, F. A. Jenet, and M. A. McLaughlin, "Practical Methods for Continuous Gravitational Wave Detection Using Pulsar Timing Data," *ApJ* **753** (July, 2012) 96.

J. A. Ellis, M. A. McLaughlin, and J. P. W. Verbiest, "The impact of a stochastic gravitational-wave background on pulsar timing parameters," *MNRAS* **417** (Nov., 2011) 2318–2329.

W. Przybysz, **J. A. Ellis**, S. Thakur, A. Hansen, R. Hardin, S. Sears, and E. Scime, "A magneto-optic probe for magnetic fluctuation measurements.," *Rev Sci Instrum* **80** (2009) no. 10, 103502.