LOW-FREQUENCY GRAVITATIONAL WAVE SEARCHES AND DATA ANALYSIS WITH HAMILTONIAN SAMPLING

by

Gabriel E. Freedman

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Physics

at

The University of Wisconsin-Milwaukee

May 2025

ABSTRACT

LOW-FREQUENCY GRAVITATIONAL WAVE SEARCHES AND DATA ANALYSIS WITH HAMILTONIAN SAMPLING

by

Gabriel E. Freedman

The University of Wisconsin-Milwaukee, 2025
Under the Supervision of Professor Sarah Vigeland, PhD

The pulsar timing array (PTA) community has found evidence for a correlated stochastic signal following the Hellings-Downs pattern indicative of an isotropic stochastic gravitational wave background (GWB), opening up a new area of gravitational wave astrophysics in the low-frequency regime. The most likely source of such a background is a population of supermassive black hole binaries, and particularly loud individual sources could be detected in future datasets. Searching for these single continuous gravitational wave (CW) sources adds additional computational complexity to an already time-intensive analysis. This increases the already large number of parameters needed to be sampled concurrently and introduces strong covariance into the model, namely between binaries emitting at low frequencies and the GWB.

In this dissertation we discuss the development of data analysis methods aimed at addressing the computational roadblocks facing PTA science. We focus on the implementation of the Hamiltonian Monte Carlo (HMC) sampling algorithm, which uses sample proposals based on the gradient of the model likelihood to more efficiently explore the high-dimensional covariant parameter spaces compared to the random-walk techniques currently employed. The HMC method was originally introduced for the studies of quantum chromodynamics but this marks the first time it is broadly adapted for methods of nanohertz GW detection.

We present an end-to-end pipeline for performing joint Bayesian searches for both a GWB and CW sources using Hamiltonian sampling, and produce a collection of benchmarking and consistency tests to display its improvements over current methods. In particular, we show that our work will scale more favorably towards future PTA datasets in terms of computational cost saved, as we continue to add more pulsars to our array and increase our data volume. Lastly we shift our focus to yet another area of GW astrophysics, the millihertz frequency regime that will be analyzed by the Laser Interferometer Space Antenna (LISA). We begin to explore whether a Hamiltonian sampling pipeline can aid in the inference of extreme mass ratio inspiral sources, an as of yet unsolved problem for LISA data science.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GW	Gravitational Wave
LIGO ...	Laser Interferometer Gravitational Wave Observatory
PTA	Pulsar Timing Array
GWB ...	Gravitational Wave Background
LISA ...	Laser Interferometer Space Antenna
TT	Transverse Traceless
SMBH ..	Supermassive Black Hole
SMBHB ..	Supermassive Black Hole Binary
CW	Continuous Wave
PSD	Power Spectral Density
MSP ...	Millisecond Pulsar
TOA ...	Time of arrival
SSB	Solar System Barycentric
NANOGrav	North American Nanohertz Observatory for Gravitational Waves
EPTA ...	European Pulsar Timing Array
PPTA ...	Parkes Pulsar Timing Array
InPTA ..	Indian Pulsar Timing Array
CPTA ...	Chinese Pulsar Timing Array
MPTA ..	MeerKAT Pulsar Timing Array
IPTA ...	International Pulsar Timing Array
ORF ...	Overlap Reduction Function
HD	Hellings Downs

MCMC ..	Markov Chain Monte Carlo
MH	Metropolis Hastings
HMC ...	Hamiltonian Monte Carlo
NUTS ...	No U-Turn Sampler
EMRI ...	Extreme Mass Ratio Inspiral
LDC ...	LISA Data Challenge
TDI	Time Delay Interferometry
MBH ...	Massive Black Hole
CO	Compact Object
AK	Analytic Kludge
AAK ...	Augmented Analytic Kludge
NK	Numerical Kludge

ACKNOWLEDGMENTS

First and foremost I would like to thank my parents. They have forever been steadfast in their support of my many endeavors, and have remained incredibly engaged and interested in my life and research. They are always eager to listen and learn about my work, even though they may admit they stopped understanding what I do about a decade ago.

There are many people I would like to acknowledge for their academic supervision and support. I would like to thank Dr. Wolfgang Choyke, my first research advisor when I was an undergraduate at the University of Pittsburgh. He imparted on me the discipline and meticulousness required to be a successful scientist. I also appreciated our countless hours of conversations on anything but physics. I would also like to thank Dr. Arthur Kosowsky. While I was employed full-time out of college Arthur allowed me to work with his group in the evenings on a research project, something that would turn out to be a huge boon for my graduate applications. Next, and most importantly, I want to thank my graduate advisor Dr. Sarah Vigeland for all of her support, mentorship, and encouragement. When I was a young graduate student sheepishly looking for an advisor to take me in, she went out of her way to find me a project to work on and join her group, and for that I am forever grateful. I also want to especially thank Dr. Aaron Johnson, a former UWM postdoc. Aaron had by far the largest influence on my programming skills and how I approach data analysis, and I would not be where I am today without his guidance.

Graduate school, and life in general, is made better by the people with whom you spend it. I am fortunate to have met and befriended so many wonderful people at UWM, in the CGCA, and as a member of the NANOGrav collaboration. I want to thank Shash-

wat for always being there to discuss research ideas or ramble about life, for being a wonderful travel companion, and for being a great friend. Aliyah and Tom are two other close friends from my first days in graduate school, and I am thankful for our bar trivia nights, family dinners, sports excursions, and all other activities where we could all hang out together. I want to thank Amanda, Lulu, and Ronan for their friendship and making our office space a fun and enjoyable place to work, and for helping me stay organized when I was overwhelmed with administrative duties. On that note I also want to acknowledge everyone who was a part of Coffeeshop Astrophysics and helped grow this amazing public outreach group during my tenure in charge. Throughout my PhD I also spent a fair bit of time at the University of Illinois at Urbana-Champaign, and I want to thank the graduate students there for making me feel right at home and part of their own social groups. I also want to give a quick shout out to Roast Coffee Company in Milwaukee and all of the great people that work there. I likely wrote nearly a third of my dissertation at that coffee shop, including this sentence.

Lastly, I am immensely grateful for my wonderful and loving partner, Jess. Every time I have felt lost, she has been there to help me find my path again. I appreciate her unwavering support as I found my way into graduate school and worked tirelessly through it, just as I have tried to reciprocate the support as she worked through her own physics PhD program in Illinois. I am so proud of everything she has accomplished, and am so excited to see us both succeed.

CHAPTER 1

Introduction

In 1916 Einstein introduced his theory of general relativity (Einstein, 1916), relating mass-energy to the curvature of spacetime and completely altering how we describe gravity. One important prediction of his theory, among many, was that there should exist small ripples in spacetime, called gravitational waves (GWs), that radiate away from accelerating masses. For decades this prediction went completely untested, as the GWs were expected to be so minuscule that it was well beyond the level of experiments that could feasibly be engineered. The first evidence of their existence arose from observing the orbital decay of a binary pulsar (Taylor & Weisberg, 1982). Nearly a century after Einstein published his theory, the Laser Interferometer Gravitational Wave Observatory (LIGO) made the first direct detection of GWs (Abbott et al., 2016), using two interferometers located in Washington and Louisiana to measure GW emission originating from a binary system of merging stellar-mass black holes. The last decade has seen an explosion of interest in and development of GW astronomy, combined with dozens of new detections from LIGO including merging binary neutron star systems.

Recently we crossed another milestone in the field, with pulsar timing array (PTA) experiments reporting the first evidence of a low-frequency stochastic gravitational wave background (GWB) signal (Agazie et al., 2023a; EPTA Collaboration et al., 2023; Reardon et al., 2023a). Direct evidence of GWs now exists across multiple frequency bands. Looking to the future, upcoming missions such as the Laser Interferometer Space Antenna (LISA) will elucidate yet another portion of the GW spectrum, the millihertz frequency regime. With all of the experiments together we can study black hole systems of all varieties, from the stellar-mass sources for LIGO all the way up to supermassive black holes,

of masses ranging from millions to billions that of our sun, that are detectable by PTAs. The complete picture of GW astronomy continues to come into focus.

1.1 GENERAL RELATIVITY AND GRAVITATIONAL WAVES

In this section we briefly review the core concepts from the theory of general relativity and gravitational waves that ultimately develop into the foundation on which this entire dissertation is built. For a more formal and in-depth description of these subjects, refer to Wald (1984) for topics in general relativity and Creighton & Anderson (2011) for GWs. For the remainder of this dissertation we will work in geometerized units where $G = c = 1$.

The crux of general relativity is that gravity can be described as a geometric property of the structure of spacetime, represented by a four-dimensional manifold with curvature that is connected to the effects of a gravitational field. More specifically, it relates the spacetime curvature to the energy-momentum present within that spacetime. This relation is governed by the Einstein field equations

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1.1.1}$$

where $G_{\mu\nu}$ is the Einstein Tensor, $g_{\mu\nu}$ is the metric, $T_{\mu\nu}$ is the stress-energy tensor describing continuous matter distributions and fields[1], and both $R_{\mu\nu}$ and R are quantities contracted from the Riemann tensor $R^{\rho}_{\mu\lambda\nu}$, which describes the curvature of the manifold. The term $R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu}$ is the Ricci tensor, and $R = g^{\mu\nu} R_{\mu\nu}$ is the Ricci scalar. The spacetime metric, or metric tensor, can in general be represented as a 4×4 symmetric matrix defining properties including time, distance, and curvature. For example, one can consider the spatially flat Minkowski spacetime, denoted by $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ upon

[1]The T_{00} term represents the energy density, the T_{0k} terms denote the kth component of linear momentum density, and the T_{kl} terms represent the kth momentum flux across a surface perpendicular to l.

which is based the theory of special relativity. Eq. (1.1.1) comprises a set of coupled, non-linear, second order partial differential equations for the components of the metric $g_{\mu\nu}$, a complicated if not impossible analytical task for any general metric, but under certain assumptions can be greatly simplified and provide illuminating results.

1.1.1 Linearized Gravity

The first step towards understanding GWs and their production is to study small perturbations, denoted by $h_{\mu\nu}$, around a flat spacetime metric and the expansion of the Einstein field equations. Explicitly, we are considering the metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1. \tag{1.1.2}$$

For all current GW detectors, the assumption that these linear perturbations are small is a valid one. We substitute Eq. (1.1.2) in Eq. (1.1.1) and are able to keep terms up to linear order in $h_{\mu\nu}$, leading to a theory of linearized gravity. For purely mathematical convenience, the typical next step in developing this solution is to rewrite the metric perturbation in a form known as the trace-reversed metric perturbation,

$$\begin{aligned} \bar{h}_{\mu\nu} &= h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \\ h_{\mu\nu} &= \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h}, \end{aligned} \tag{1.1.3}$$

where $h = \eta^{\mu\nu}h_{\mu\nu}$ and $\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu}$. Making this change reduces the complexity of the Einstein equations, and for any solution $\bar{h}_{\mu\nu}$ we can promptly convert back to $h_{\mu\nu}$. In general $h_{\mu\nu}$, as a symmetric 4×4 matrix, will have 10 independent components. Similar to studies of electromagnetism, we can make an appropriate gauge choice to reduce the total

number of degrees of freedom. Choosing the Lorenz gauge, under the condition that the metric perturbation is invariant under a coordinate transformation $x^\mu \to x'^\mu = x^\mu + \xi^\mu(x)$, gives

$$\partial^\nu \bar{h}_{\mu\nu} = 0. \tag{1.1.4}$$

Imposing this gauge freedom gives four additional conditions on $\bar{h}_{\mu\nu}$, bringing the total number of independent components down to 6. The linearized Einstein field equations are then written as

$$\Box \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}, \tag{1.1.5}$$

where the operator $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the flat-space d'Alembertian. Eq. (1.1.5) is simply the classical wave equation with an added source term. This demonstrates that the solutions of the linearized Einstein equations can be described as some class of waves, otherwise known as gravitational waves.

1.1.2 Gravitational Waves and their Generation

We now seek solutions of Eq. (1.1.5) to describe GWs and their production. To do so, we consider two different regimes: far outside the radiating source, and near it. In the far-field or vacuum regime, the source term of the equation vanishes, and we are left solving

$$\Box \bar{h}_{\mu\nu} = 0. \tag{1.1.6}$$

We have additional gauge freedom here, and use the standard transverse traceless gauge (TT) from the GW literature. This gauge further reduces the number of indepen-

dent components of $\bar{h}_{\mu\nu}$ from 6 down to 2, according to the following conditions

$$h^{0\mu} = 0, \qquad h^i_i = 0, \qquad \partial^j h_{ij} = 0. \tag{1.1.7}$$

We point out these conditions imply that $h^{TT}_{\mu\nu} = \bar{h}^{TT}_{\mu\nu}$, where for the remainder of this dissertation we will use the TT superscript to denote quantities given in the TT gauge. Plane waves are solutions to Eq. (1.1.6). Applying the gauge conditions and choosing a coordinate system such that the wave is propagating in the z-direction, yields the solutions

$$h^{\mathrm{TT}}_{\mu\nu}(t,z) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos[\omega(t-z)], \tag{1.1.8}$$

where we have introduced the subscripts $+$ and $\times$ to denote the two distinct GW polarizations, the plus- and cross-mode polarizations, allowed by general relativity. The two modes are named based on the shape of the tidal deformations they induce on a ring of test masses. Figure 1.1 displays the effect of the two different modes for a ring perpendicular to the direction of GW propagation.

When we next look into the near-field solutions, we will want them to similarly be given in the TT gauge. To do so for some arbitrary solution $h_{\mu\nu}$, we can define a transverse projection operator for a GW propagating in the direction $\hat{\mathbf{n}}$,

$$P_{ij} = \delta_{ij} - n_i n_j, \tag{1.1.9}$$

and from it construct a full TT-gauge projector,

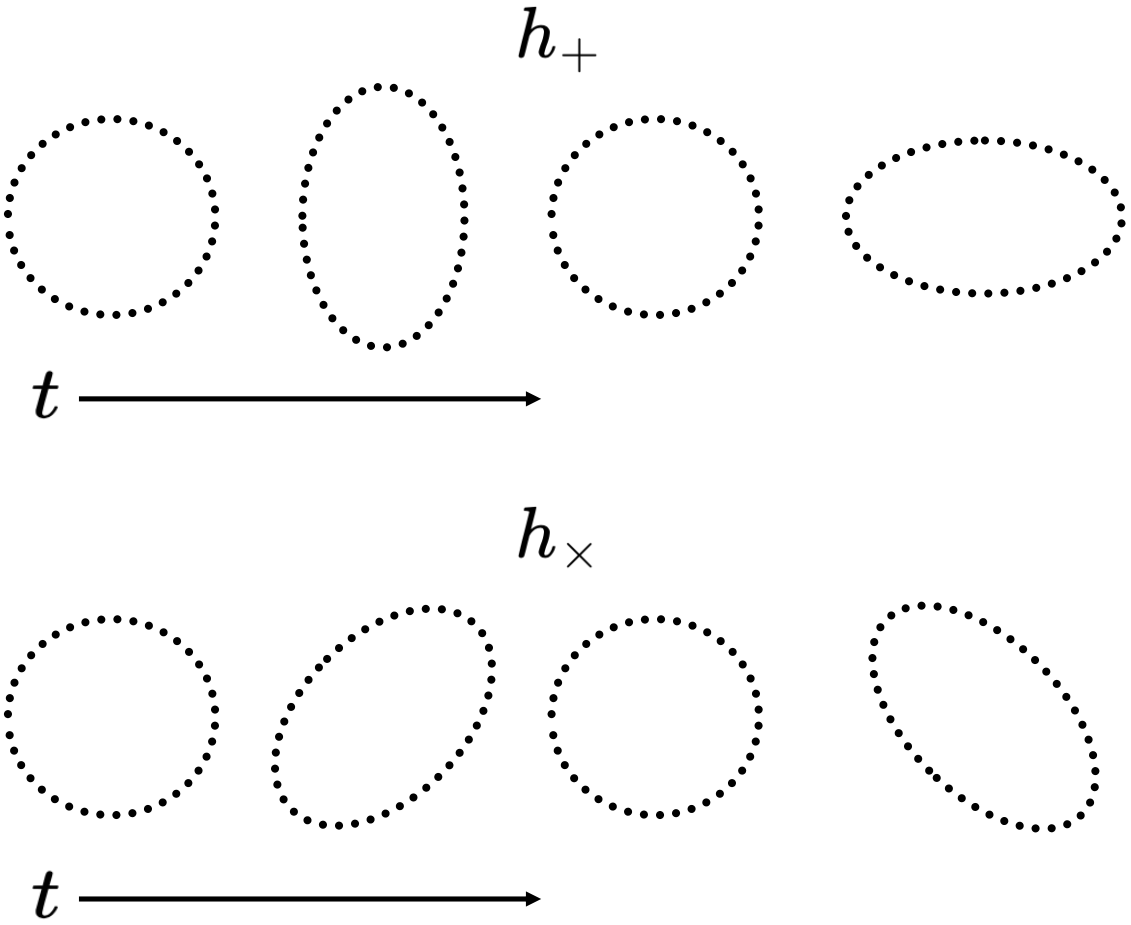

Figure 1.1: Illustration of the two independent polarization modes of a GW acting on a ring of particles. The particles lie in the $x - y$ plane and the wave is propagating in the $\hat{z}$ direction. Time moves from left to right, with the rings displayed at $t = P/4$ intervals for an incident GW with period P.

$$\Lambda_{ij,kl} = P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}. \tag{1.1.10}$$

With the far-field solution complete, we now want to connect them to solutions nearby the source where the GWs are produced. The solutions of the wave equation with a source can be solved using the retarded Green's function,

$$G\left(\mathbf{x}, \mathbf{x}', t, t'\right) = -\frac{\delta(t - |\mathbf{x} - \mathbf{x}'|)}{4\pi|\mathbf{x} - \mathbf{x}'|}, \tag{1.1.11}$$

which in turn yields

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = 4\int d^3x' \frac{T_{\mu\nu}\left(t - |\mathbf{x} - \mathbf{x}'|, \mathbf{x}\right)}{|\mathbf{x} - \mathbf{x}'|}. \tag{1.1.12}$$

By applying Eq. (1.1.10) outside of the source, we can project this solution into the TT gauge. Next we assume that we are observing the GW at some distance far from the source, at much greater distances than the GW wavelength, meaning that $|\mathbf{x} - \mathbf{x}'| \simeq r$ is effectively constant. Then we can Taylor expand the integrand in Eq. (1.1.12), keeping only the leading term, which gives the metric perturbation as

$$h_{ij}^{\mathrm{TT}}(t, \mathbf{x}) = \frac{4}{r}\Lambda_{ij,kl}\int d^3x' T_{ij}(t - r, \mathbf{x}). \tag{1.1.13}$$

Lastly, we derive an equation for the integral of the spatial components T_{ij} of the stress-energy tensor. Using the conservation law $\partial_\mu T^{\mu\nu} = 0$ and the divergence theorem we arrive, after some algebra, at the famous quadrupole formula,

$$h_{ij}^{\mathrm{TT}}(t, \mathbf{x}) = \frac{2}{r}\ddot{I}_{kl}^{TT}(t - r), \tag{1.1.14}$$

where I_{ij} is the quadrupole moment given by

$$I_{ij}(t) = \int d^3x' T_{00}(t, \mathbf{x}) x'^i x'^j. \tag{1.1.15}$$

It is also useful to define the reduced[2] quadrupolar moment as

$$\mathcal{I}_{ij}(t) = \int d^3x' T_{00}(t, \mathbf{x}) \left(x'^i x'^j - \frac{1}{3} r^2 \delta_{ij} \right). \tag{1.1.16}$$

As one final aside, it can be useful to calculate the total power radiated by the GW source, also called the GW luminosity. This is found, in the quadrupole approximation, by taking a time derivative of the metric perturbation and integrating over the sphere, resulting in

$$L_{\mathrm{GW}} = \frac{1}{5} \left\langle \dddot{\mathcal{I}}_{ij} \dddot{\mathcal{I}}_{ij} \right\rangle, \tag{1.1.17}$$

where the angle brackets $\langle \cdot \rangle$ denote averaging over multiple wave cycles.

We now have in hand formulas to calculate the production of GWs from any source, so long as we have some knowledge or make some assumptions about its mass distribution. In order for GWs to be produced at all, those objects must be accelerating. Solutions to Eq. (1.1.14) will again be written in terms of the $+$ and $\times$ polarization modes. The components of h_{ij}^{TT}, analogous to the amplitudes of classical waves, are referred to as strain, representing the ratio by which lengths between two test masses are stretched or compressed. GW strain is dimensionless, think $h = \delta L / L$ for some distance L or $h = \delta T / T$ for some light travel time T. This is at the crux of all GW experiments: to measure the strain in some statistically significant way so as to discern the properties of its source.

[2]Reduced here meaning traceless, i.e., $\mathcal{I}_{ij} = I_{ij} - \frac{1}{3} \delta^{ij} I_{kk}$

1.2 THE ARRAY OF GRAVITATIONAL WAVE SOURCES

What are the primary sources of GWs? The short answer, at least in the detectable sense, are a wide variety of compact objects. The specific answer to this question, however, varies depending on the frequency regime in which you are probing. The collection of current and upcoming GW experiments cover an expected frequency range of roughly 12 orders of magnitude ($\sim 10^{-9}$ Hz $- 10^{3}$ Hz), from which we can divide roughly into three categories: "high"-, "middle"-, and "low"-frequency regimes. Figure 1.2 displays the expected gravitational wave strain for sources in these three regimes vs. the corresponding GW frequencies. At the high-frequency end ($\sim 10^{0} - 10^{3}$ Hz) we find the signals of inspiralling and merging stellar-mass black holes and neutron stars. This is the domain of ground-based detectors such as LIGO, Virgo, and KAGRA (LIGO Scientific Collaboration et al., 2015; Acernese et al., 2015; Akutsu et al., 2021). The middle-frequency band ($\sim 10^{-5} - 10^{-1}$ Hz) will be explored and studied by future space-based laser interferometers, most notably the Laser Interferometer Space Antenna (LISA; Amaro-Seoane et al. 2017), where we will be able to study the population of white-dwarf binaries in our galaxy, observe the mergers of massive black hole binaries, and study the capture of smaller compact objects around these massive black holes. The LISA mission, and one of its potential GW sources, is the focus of Chapter 5. At the low-frequency end ($\sim 10^{-9} - 10^{-7}$ Hz) we reach the territory of pulsar timing arrays (Sazhin, 1978; Detweiler, 1979; Foster & Backer, 1990). Here the overwhelming majority of GW signals originate from binary systems of supermassive black holes. The remainder of this section will further elucidate this primary class of sources and derive the expected GW strain we could observe from both a single binary as well as a cosmic collection of binaries as a stochastic gravitational wave background.

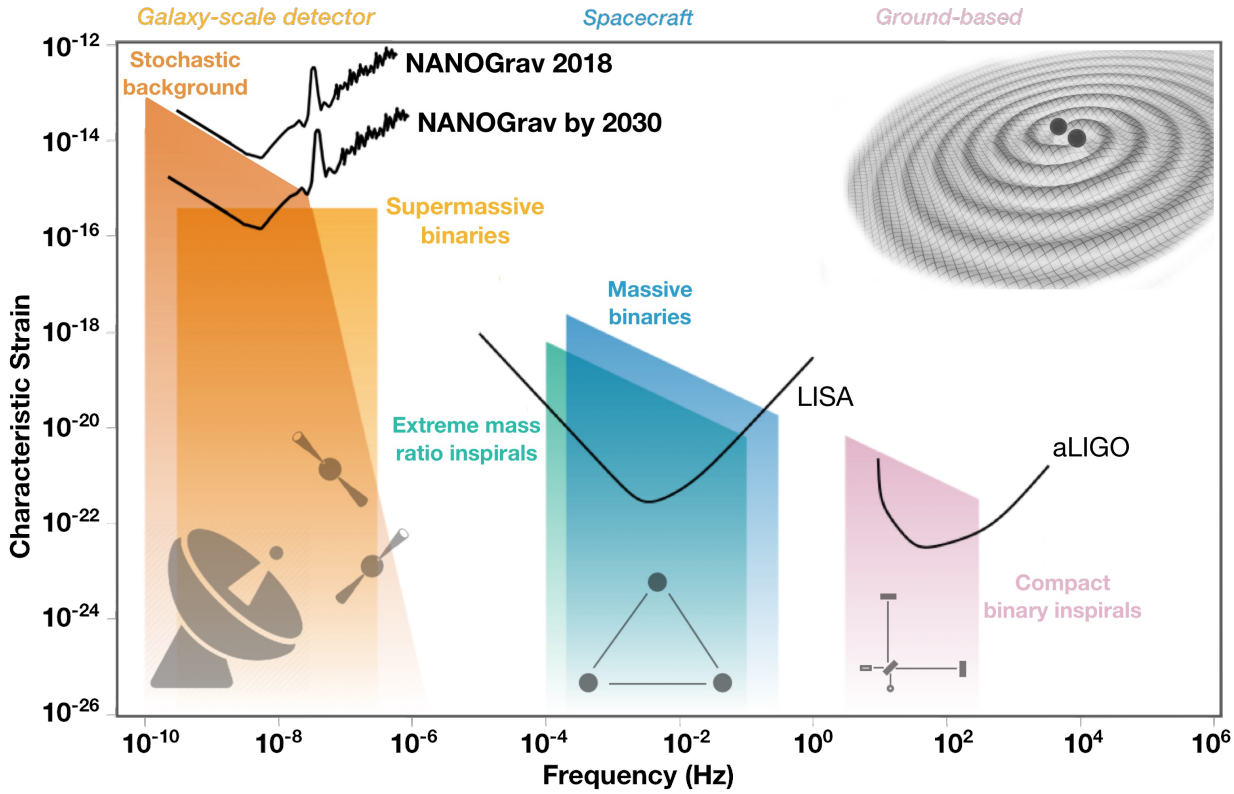

Figure 1.2: Figure credit: S. Taylor and C. Mingarelli, produced by GWplotter and adapted from a figure in Mingarelli & Mingarelli (2018). The GW spectrum is displayed as a characteristic strain vs. frequency plot. The distinct orange, blue, and pink regions denote the regimes probed by pulsar timing arrays, space-based interferometers, and ground-based interferometers, respectively. The black lines represent sensitivity curves for the detectors in each region.

1.2.1 Supermassive Black Hole Binaries

Nearly every large galaxy is expected to house a supermassive black hole (SMBH) at its center (Richstone et al., 1998). When two such galaxies merge their component black holes eventually coalesce into a binary system, forming a supermassive black hole binary (SMBHB; Ostriker & Hausman 1977; White 1980; Lacey & Cole 1993; Milosavljević & Merritt 2003). The evolution of the SMBHB orbit is a complex process influenced by numerous factors including the structure of the galactic environment, the distribution of surrounding stars and gas, the properties of the individual black holes, and the processes that define the evolution change over the course of the binary's millions of years march towards coalescence (Begelman et al., 1980; Yu, 2002; Escala et al., 2005; Dotti et al., 2007; Haiman et al., 2009). Initially it is dominated by dynamical friction, stripping energy from the binary until it reaches an orbital separation on the order of $\sim$1 pc. At sub-parsec ($\sim 0.1 - 0.01$ pc) separation, GW emission will take over as the prominent factor giving the binary its final push to an eventual merger event. What physical processes drive the evolution of the SMBHB below 1 pc separation to the point of GW emission is still unknown. This is an open question known colloquially as the "final parsec problem".

Assuming the binaries overcome this final parsec problem and reach a stage in their evolutionary track where they are dominated by the radiation of GWs, we have a tantalizing opportunity to study their properties and interactions not only through their GW emission, but through broader multi-messenger astrophysics as well. The detection of GWs from SMBHBs can be united with electromagnetic observations to help place constraints on the binary population, gain information on galaxy mergers and formation, and study properties of the surrounding astrophysical environments (Ravi et al., 2015; Sesana, 2013; Quinlan, 1996).

1.2.2 Continuous Gravitational Waves from a single SMBHB

First we will examine the GW emission from a single SMBHB source. We begin by considering an SMBHB system with black holes of masses m_1 and m_2 and orbital separation a. We work in the center-of-mass frame of the binary with total mass $M = m_1 + m_2$ and reduced mass $\mu = m_1 m_2/(m_1 + m_2)$. Initially we will consider the binary to be in uniform circular motion with a constant source-frame orbital frequency ω_s. We also will place the orbit of the binary to lie in the x-y plane with the observer placed directly on the z-axis. Later we will relax the above assumptions to account for both GW frequency evolution as well as an SMBHB orbit that is inclined with respect to the observer.

In this reference frame and with the above specifications the quadrupolar tensor takes the matrix form

$$I_{ij} = \frac{\mu a^2}{2} \begin{pmatrix} 1 + \cos(2\omega_s t) & \sin(2\omega_s t) & 0 \\ \sin(2\omega_s t) & 1 - \cos(2\omega_s t) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.2.1}$$

After taking two time derivatives we arrive at

$$\ddot{I}_{ij} = 2\mu a^2 \omega_s^2 \begin{pmatrix} -\cos(2\omega_s t) & -\sin(2\omega_s t) & 0 \\ -\sin(2\omega_s t) & \cos(2\omega_s t) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.2.2}$$

Because we are working in the case of uniform circular motion, any terms containing $\dot{a}$ or $\dot{\omega}_s$ necessarily vanish. When we later move to a quasi-circular regime ($\dot{\omega}_s \ll \omega_s^2$) to describe the frequency evolution of the binary, the above time derivatives can again safely be ignored. This can now be related via Eq. (1.1.14) to give our initial expressions for the GW strain functions $h_+(t)$ and $h_\times(t)$,

$$h_+(t) = \frac{-4\mu a^2\omega_s^2}{r}\cos\left(2\omega_s t_{\rm ret}\right), \\ h_\times(t) = \frac{-4\mu a^2\omega_s^2}{r}\sin\left(2\omega_s t_{\rm ret}\right), \tag{1.2.3}$$

where we have introduced the shorthand notation for the retarded time $t_{\rm ret} = t - r$.

It is now time for a modest amount of foreshadowing. We note that in practice it is very difficult if not impossible in many situations to accurately measure either the orbital separation or individual mass components of the SMBHB. Therefore we wish to move towards representations of $h_+(t)$ and $h_\times(t)$ that are parameterized such to be more conducive to any eventual data analysis pipeline. First we will utilize Kepler's Third Law to write the orbital separation a in terms of the total mass M and the orbital frequency ω_s as

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

We introduce a quantity called the chirp mass $M_c = (m_1 m_2)^{3/5}(m_1+m_2)^{-1/5} = \mu^{3/5}M^{2/5}$ as a measure of the effective mass of the binary. Substituting this expression and Eq. (1.2.4) into Eq. (1.2.3) we obtain

$$h_+(t) = \frac{-4M_c^{5/3}\omega_s^{2/3}}{r}\cos\left(2\omega_s t_{\rm ret}\right), \\ h_\times(t) = \frac{-4M_c^{5/3}\omega_s^{2/3}}{r}\sin\left(2\omega_s t_{\rm ret}\right). \tag{1.2.5}$$

At this stage we must generalize the above expressions. We begin by moving the binary orbit from the strictly circular to the quasi-circular regime and allow for a nonzero (albeit still quite small) orbital frequency evolution. An SMBHB source will radiate energy over time through GW emission and will experience a corresponding decrement of

its orbital separation. As the binary orbit shrinks, its orbital frequency must grow according to Eq. (1.2.4). In order to calculate this change in frequency over time $\dot{\omega}_s$ we first need an expression for the power radiated from the SMBHB by GW emission. This can be found by taking another time derivative of Eq. (1.2.2) and utilizing Eq. (1.1.17) to find

$$L_{\rm GW} = -\frac{dE_{\rm GW}}{dt} = -\frac{32}{5}\mu^2 a^4 \omega_s^6. \tag{1.2.6}$$

We can then equate this energy loss to the change in orbital energy $E_{\rm orb} = -M\mu/2a$ of the binary resulting from the increase in orbital frequency

$$\begin{aligned} -\frac{dE_{\rm GW}}{dt} &= \frac{dE_{\rm orb}}{dt}, \\ -\frac{32}{5}\mu^2 a^4 \omega_s^6 &= \frac{1}{2}\frac{M\mu}{a^2}\dot{a}, \\ \implies \dot{\omega}_s &= \frac{96}{5} M_c^{5/3} \omega_s^{11/3}, \end{aligned} \tag{1.2.7}$$

where we have used $\dot{a} = -(2/3)M^{1/3}\omega_s^{-5/3}\dot{\omega}_s$, a consequence of Eq. (1.2.4). It now becomes more apparent why the chirp mass is a useful quantity to define and use. To leading order, it determines the evolution of the SMBHB orbit.

Taking t_0 to be some reference time and ω_0 to be the initial orbital frequency at $t = t_0$, we can integrate the above expression to find a function for the orbital frequency at a given time t,

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

We can integrate again to solve for the orbital phase $\Phi(t)$, noting that for a circular orbit $d\Phi/dt = \omega_s$. Doing so gives

$$
\Phi(t) = \Phi_0 + \frac{1}{32} M_c^{5/3}\left(\omega_0^{-5/3} - \omega_s(t)^{5/3}\right), \tag{1.2.9}
$$

where Φ_0 denotes the orbital phase evaluated at $t = t_0$. Returning to our GW strain functions, we can substitute ω_s by its full expression $\omega_s(t)$, and replace $\omega_s t_{\rm ret}$ by $\Phi(t_{\rm ret})$. This allows us to write the strain as arbitrary functions of time t,

$$
\begin{aligned}
h_+(t) &= \frac{-4 M_c^{5/3} \omega_s(t_{\rm ret})^{2/3}}{r} \cos\left(2\Phi(t_{\rm ret})\right), \\
h_\times(t) &= \frac{-4 M_c^{5/3} \omega_s(t_{\rm ret})^{2/3}}{r} \sin\left(2\Phi(t_{\rm ret})\right).
\end{aligned}
\tag{1.2.10}
$$

One further generalization is to allow the observer to stray from the z-axis and exist at some inclination angle $\imath$ with respect to the binary. To accommodate this change we first rotate the quadrupole tensor, accomplished by computing the product $R_x^{ki}(\imath)\ddot{I}_{ij}R_x^{jl}(\imath)^{-1}$ for the matrix $R_x(\theta)$ defining a rotation about the x-axis. Projecting to the TT gauge, and in general following the same steps outlined above, we arrive at the updated functions of the GW strain that are valid for observers at any arbitrary inclination,

$$
\begin{aligned}
h_+(t) &= \frac{-2M_c^{5/3}\omega_s(t_{\rm ret})^{2/3}}{r}\left(1+\cos^2(\imath)\right)\cos\left(2\Phi(t_{\rm ret})\right), \\
h_\times(t) &= \frac{-4M_c^{5/3}\omega_s(t_{\rm ret})^{2/3}}{r}\cos(\imath)\sin\left(2\Phi(t_{\rm ret})\right).
\end{aligned}
\tag{1.2.11}
$$

Lastly we note that the above equations for the strain are given in the rest-frame of the source, whereas we seek a solution in the observer-frame. For local[3] sources they are essentially equivalent, but for SMBHBs with non-negligible redshift there are additional changes needed to Eq. (1.2.11) to account for cosmological effects. This is done by making the following adjustments to the binary's chirp mass and frequency,

$$
\mathcal{M} = (1+z)M_c, \qquad \omega = \frac{\omega_s}{1+z},
\tag{1.2.12}
$$

and by substituting the distance r by the luminosity distance $d_L = (1+z)r$. Applying these changes yields the final expressions for the GW strain,

$$
\begin{aligned}
h_+(t) &= \frac{-2\mathcal{M}^{5/3}\omega(t_{\rm ret})^{2/3}}{d_L}\left(1+\cos^2(\imath)\right)\cos\left(2\Phi(t_{\rm ret})\right), \\
h_\times(t) &= \frac{-4\mathcal{M}^{5/3}\omega(t_{\rm ret})^{2/3}}{d_L}\cos(\imath)\sin\left(2\Phi(t_{\rm ret})\right).
\end{aligned}
\tag{1.2.13}
$$

Even though we can model the full frequency evolution of the source, the evolution rate is generally very small over the time the emitted GWs are in the nHz band. These signals are therefore given the moniker of "continuous waves" (CWs).

[3] i.e., $(1+z) \approx 1$ for redshift z

1.2.3 Stochastic Gravitational Wave Backgrounds

The low-frequency GW regime is permeated by a cosmic population of SMBHBs, all behaving roughly as outlined in the previous section. The common analogy used is that of a crowded bar or cafe. As an observer we can try to hone in on the individual voices of specific patrons, but we will also always hear the background hum of their collective conversations. The same can be said about detecting these GW signals. Even in the cases where we do not have the capabilities to resolve single binaries, we can still aim to detect the incoherent sum of the full signal population. This is what is referred to as a stochastic gravitational wave background (GWB). The stochasticity represents the fact that the GW signal we are now looking for isn't something we can write down explicit analytic expressions for like in Eq. (1.2.13), but rather is something we treat as a random process with some statistical properties that we can infer.

Here we will review the characteristic GW strain spectrum for a stochastic GWB. Initially we will make no assumptions about the source of the GWB, but later we will take the general results and apply it to a superposition of SMBHBs in circular orbits. We direct the reader to Phinney (2001) for a more in-depth analysis and a complete derivation. We start by considering the plane wave solution for the metric perturbation given in Eq. (1.1.8). For a stochastic GWB signal we can generalize this expression as an expansion over plane waves from some arbitrary direction $\hat{\mathbf{n}}$ pointing from the observer to a source whose location we can parameterize with (θ, ϕ) denoting its polar and azimuthal angles, respectively. We define our coordinate system using the following orthonormal basis vectors,

$$\begin{aligned}\hat{\mathbf{n}} &= \sin\theta\cos\phi\hat{\mathbf{x}} + \sin\theta\sin\phi\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}},\\ \hat{\mathbf{u}} &= \cos\theta\cos\phi\hat{\mathbf{x}} + \cos\theta\sin\phi\hat{\mathbf{y}} - \sin\theta\hat{\mathbf{z}},\\ \hat{\mathbf{v}} &= -\sin\phi\hat{\mathbf{x}} + \cos\phi\hat{\mathbf{y}},\end{aligned} \tag{1.2.14}$$

from which we can construct the GW polarization basis tensors

$$e^{+}_{ij}(\hat{\mathbf{n}}) = \hat{u}_i\hat{u}_j - \hat{v}_i\hat{v}_j, \qquad e^{\times}_{ij}(\hat{\mathbf{n}}) = \hat{u}_i\hat{v}_j - \hat{v}_i\hat{u}_j. \tag{1.2.15}$$

The metric perturbation for the GWB can be written as

$$h_{ij}(t,\mathbf{x}) = \sum_A \int_{-\infty}^{\infty} df \int d^2\hat{\mathbf{n}}\, \tilde{h}_A(f,\hat{\mathbf{n}}) e^A_{ij}(\hat{\mathbf{n}}) e^{-2\pi i f(t-\hat{\mathbf{n}}\cdot\mathbf{x})}, \tag{1.2.16}$$

where $A \in [+,\times]$ again indexes the two allowed GW polarization modes and the $\tilde{h}_A(f,\hat{\mathbf{n}})$, calculated as the Fourier transform of $h_A(t,\hat{\mathbf{n}})$, are complex random variables defining the GWB. Since we seek to characterize its statistical properties, the quantities of interest are the moments and correlators of $h_{ij}(t,\mathbf{x})$, and by extension $\tilde{h}_A(f,\hat{\mathbf{n}})$. We now make a set of assumptions that will define one of the correlators. The assumptions about the GWB are as follows:

- *It is isotropic to first approximation*. A population of background signals of cosmological origin (take the cosmic microwave background, for example) can generally be considered spatially isotropic to first order. Small anisotropies may, and likely do, exist, but for getting a grasp on the statistics of the background signal this assumption is a valid starting point. Note that this condition of isotropy would not be satisfied when considering non-cosmological backgrounds, such as those of galactic origin.

- *It is Gaussian distributed*. The Gaussianity of the GWB is a direct consequence of the central limit theorem. For a large number of sources we expect their superposition to be a Gaussian stochastic process.

- *It is stationary*. This condition states that the background is not changing appreciably over any observation campaign we are performing. Changes in the GWB, such as its spectrum being redshifted, happen on time scales on the order of the age of the universe whereas GW experiments extend at most over years or decades. Another consequence of stationarity is that is that the expectation $\langle h_A(f, \hat{\mathbf{n}})\rangle = 0$.

- *It is unpolarized*. We have defined the GWB as the sum of many different random individual signals. It is therefore reasonable to assume that this combined signal would be unpolarized.

With these assumptions in hand the statistical properties of the Fourier modes of the GWB can be summarized by the following expression

$$\langle h_A(f, \hat{\mathbf{n}}) h^*_{A'}(f', \hat{\mathbf{n}}')\rangle = \delta(f - f')\frac{\delta^{(2)}(\hat{\mathbf{n}}, \hat{\mathbf{n}}')}{4\pi}\frac{\delta_{AA'}}{2}S_h(f), \tag{1.2.17}$$

where the function $S_h(f)$ is the singe-sided power spectral density (PSD) of the GWB. The "one-sided" classifier stems from the fact that we only integrate over a physical range of positive frequencies. The units of the PSD are Hz^{-1}. The above expression uniquely characterizes the signal. From $S_h(f)$ we can define another important quantity called the characteristic strain,

$$h_c \equiv \sqrt{f S_h(f)}, \tag{1.2.18}$$

which is what is most often cited in the GW literature when discussing these stochastic signals. It is a dimensionless quantity, and when plotted against a detector's sensitivity

curve on a log-log scale it can be related to the signal-to-noise ratio by calculating the area between the two curves (Moore et al., 2015).

In order to gain some physical intuition of the signal and how it is distributed across frequency space, we can think of the signal in terms of its fractional energy density, defined as

$$\Omega_{\rm gwb}(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm gwb}}{d\log f}, \tag{1.2.19}$$

where the energy density is normalized to $\rho_c = 3H_0^2/8\pi$, the critical density needed for a closed universe, and $\rho_{\rm gwb}$ can be related to the metric perturbation by

$$\rho_{\rm GWB} = \frac{1}{32\pi}\left\langle \dot{h}_{ij}\dot{h}^{ij}\right\rangle. \tag{1.2.20}$$

Using Eqs. (1.2.20), (1.2.16), and (1.2.17) we can solve for the fractional energy density as a function of the characteristic strain of the GWB. It is then straightforward to show

$$\Omega_{\rm gwb}(f) = \frac{2\pi^2}{3H_0^2}f^2h_c^2(f). \tag{1.2.21}$$

The above expression is valid for stochastic GWB signals from any collection of sources that satisfy the given conditions. We can now take results from Sec. 1.2.2 to uncover the spectral shape of a GWB arising from inspiralling SMBHBs in circular orbits. This is accomplished by re-expressing the $d\rho_{\rm gwb}/d\log f$ term in Eq. (1.2.19) in a form that is representative of its component binaries. Taking the distribution of sources to be continuous, we can write this as an integral over all emitting sources

$$\begin{aligned}\frac{d\rho_{\rm gwb}}{d\log f} &= \int_0^\infty dz \frac{dn}{dz}\frac{1}{1+z}\frac{dE_{\rm gw}(f_s)}{d\log f_s}\bigg|_{f_s=f(1+z)}, \\ &= \int_0^\infty dz \frac{dn}{dz}\frac{1}{1+z} f_s \frac{dE_{\rm gw}(f_s)}{dt_s}\frac{dt_s}{df_s}\bigg|_{f_s=f(1+z)},\end{aligned} \tag{1.2.22}$$

where dn/dz is the number density of sources over redshift, f_s and t_s are the source frequency and time, respectively, and $dE_{\rm gw}/d\log f$ is the GW energy emitted by a binary in its source frame per unit of logarithmic frequency. We can now make a simple scaling argument to discern the properties of the GWB spectrum. Using Eq. (1.1.17) we know that $dE_{\rm gw}/dt_s \propto f_s^{10/3}$ and from Eq. (1.2.7) we know that $df_s/dt_s \propto f_s^{11/3}$. This results in the expression in Eq. (1.2.22) scaling as $f^{2/3}$, and comparing to Eqs. (1.2.19) and (1.2.21) we finally arrive at the characteristic strain spectrum for a population of SMBHBs,

$$h_c(f) \propto f^{-2/3}. \tag{1.2.23}$$

This simple power-law model for the GWB is used throughout the low-frequency GW literature and is at the basis of many data analysis pipelines studying such stochastic backgrounds. More generally, environmental effects around the binary may cause deviations from this expected spectral shape (Sampson et al., 2015a), changing the proportionality constant or perhaps altering the spectral slope away from a true power-law. Such effects are not discussed here, and throughout the rest of this dissertation we will stick to the above expression as our fiducial model.

1.3 DETECTING GRAVITATIONAL WAVES USING PULSAR TIMING

Although low-frequency GWs can be described rather succinctly from a theoretical perspective, developing actual experiments to detect such signals is an entirely different beast of a problem. The high-frequency GW regime can be captured by building interferometers with sizes on the order of a few kilometers. Experiments in the millihertz-frequency regime necessitates the complete suppression of terrestrial noise, which is accomplished by moving the laser interferometer to space. The interferometer arm lengths then can also be made considerably longer, on the order of a few million kilometers. Once we begin discussing low-frequency GWs, we are talking about waves with periods ranging from months to decades and wavelengths on the order of lightyears. For these GW sources we use a method called pulsar timing to effectively turn our galaxy into one giant GW detector.

The following section introduces pulsars and the notion of pulsar timing. We then derive the response of pulsar timing to emitted GWs, and end with a discussion on the correlation signature we expect to see for detecting a stochastic GWB signal.

1.3.1 Pulsars

In 1967, while working on a project to study quasars with a newly built radio telescope, Jocelyn Bell Burnell noticed in her data loud, regular pulses of emission originating from the same location on the sky (Hewish et al., 1968). This marked the first official discovery of "pulsating stars," or pulsars. Pulsars are highly-magnetized, rapidly spinning neutron stars, existing as the collapsed remnants of massive stars that underwent supernova explosions. They emit beams of radio emission from their magnetic poles that can be observed on Earth whenever the beam crosses our line of sight. This leads to the depiction of pulsars as "cosmic lighthouses", as we can observe this uptick in radio emission from

the beam pointing towards us with some periodicity. This relates the pulse period of the pulsar to its rotational period.

The pulse period of a pulsar will gradually change over time as the emitted radiation carries away rotational kinetic energy from the star. This rate of slowing, called the pulsar's spin-down $\dot{P}$, along with its pulse period P are two of the most important fundamental properties in classifying these objects. The largest population of pulsars are the "canonical pulsars", with spin periods on the order of seconds, spin-down rates roughly between $10^{-13} - 10^{-16}$, and large magnetic fields. Over time canonical pulsars will spin down to the point where they no longer can produce significant radio emission. In the cases where these "dead" pulsars also have a binary companion star, and if that star is massive enough to overflow its Roche lobe, they can begin accreting matter off of the companion and spin back up, decreasing their rotational periods and once again emitting radio beams. This second population of pulsars is known as millisecond pulsars (MSPs). They have periods on the order of milliseconds, spin-down rates between $10^{-19} - 10^{-21}$, and smaller magnetic fields. MSPs are incredibly stable with very precise pulse behavior.

1.3.2 Overview of Pulsar Timing

The stability of MSPs make them exceptional astrophysical timekeepers. The basis of pulsar timing is that one can observe radio pulses with some regular cadence, deduce the rotational parameters of a pulsar, construct some theoretical timing model of the source, then ultimately compare that model against the observed pulse times-of-arrival (TOAs). For a more in-depth description of the pulsar timing observational method, see Lorimer & Kramer (2012). Single pulse observations will vary randomly between rotations (Helfand et al., 1975), so instead thousands of pulses are averaged to form a stable pulse shape. From this average pulse profile we can construct a timing model for the pulsar to predict

future pulses. This model includes any known deterministic events that would advance or delay the pulse arrival time. It also includes correction terms to move the observations into a solar system barycentric (SSB) frame of reference. For the arrival time $t_{\rm tm}$ given by the deterministic timing model in the SSB frame we have in total

$$t_{\rm tm} = t_t - t_0 + \Delta_{\rm clock} - \Delta_{\rm DM} + \Delta_{R\odot} + \Delta_{E\odot} + \Delta_{S\odot} + \Delta_{\rm bin}, \tag{1.3.1}$$

where t_t is the arrival time measured at the observatory, t_0 a reference epoch, and the remainder the principal delays. The term $\Delta_{\rm clock}$ accounts for additional clock corrections. The term $\Delta_{\rm DM}$ represents the dispersion delay induced by the interstellar medium. The Roemer delay term $\Delta_{R\odot}$ comes from the time to cross from the Earth's orbit to the SSB. The Einstein delay term $\Delta_{E\odot}$ accounts for changes due to the gravitational field of the pulsar and time dilation from its motion relative to Earth. The Shapiro delay $\Delta_{S\odot}$ term is a general relativistic effect stemming from the radio emission traveling through the potential well of the Sun. Lastly the $\Delta_{\rm bin}$ term is included when the pulsar is in a binary system and contains the previous three delays now induced by the companion. This timing model can then be compared against observed TOAs creating a dataset of timing residuals,

$$\delta t = t_{\rm obs} - t_{\rm det}. \tag{1.3.2}$$

The timing residuals encode all other potential effects on the pulse arrival times including any that may be induced by GWs. It is the primary data product used in pulsar timing analyses, discussed more at length in Chapter 2.

As we will see later in this chapter, detecting GWs with pulsar timing necessitates we study the correlations between observations of multiple pulsars. Therefore it is benefi-

cial for low-frequency GW experiments to observe a large array of pulsars. Today there are six pulsar timing array (PTA) collaborations spread across the globe, each one observing some set of pulsars using a host of different radio telescopes: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav; Ransom et al. 2019), the European Pulsar Timing Array (EPTA; Kramer & Champion 2013), the Parkes Pulsar Timing Array (PPTA; Manchester et al. 2013), the Indian Pulsar Timing Array (InPTA; Tarafdar et al. 2022), the Chinese Pulsar Timing Array (CPTA; Xu et al. 2023), and the MeerKAT Pulsar Timing Array (MPTA; Miles et al. 2023). There is also the International Pulsar Timing Array (IPTA; Perera et al. 2019), a global consortium of the major PTA collaborations.

1.3.3 Pulsar Timing Response to GWs

Pulsar timing provides a set of residuals that can contain the effects due to the metric perturbations caused by GWs along the line of sight to the pulsar. Here we will review the response of timed pulses to a passing GW. This derivation is based on the one found in Maggiore (2018) and can be referenced for more explicit details. What we seek is the integral of the GW-induced redshift over time, which we will define as

$$z(t,\hat{\Omega}) \equiv -\frac{\Delta\nu}{\nu} = \frac{\nu_p - \nu_e}{\nu_p} = \frac{\Delta T}{T}, \tag{1.3.3}$$

where ν_e denotes the observed radio frequency at the Earth, ν_p denotes the emitted radio frequency at the pulsar, T is the pulsar's rotational period, and ΔT represents the change in observed period induced by the GW after one rotation. We want to describe the fractional change in the pulse arrival times between successive pulses. If we consider a GW propagating in the direction $\hat{\Omega}$[4] and a pulsar at some distance L from the SSB sending

[4] so $\hat{\Omega} = -\hat{\mathbf{n}}$ as defined in Eq. (1.2.14)

radio beams towards Earth in the $\hat{\mathbf{p}}$ direction, then we can write the difference between the time of pulse emission and observation as

$$t_{\rm obs} - t_{\rm em} = L + \frac{p^i p^j}{2} \int_{t_{em}}^{t_{em}+L} dt' h_{ij}^{TT}\left(t', (t_{em} + L - t')\hat{\mathbf{p}}\right). \tag{1.3.4}$$

That is, the difference in times is equal to the light travel time between the Earth and pulsar, as well as some additional delay stemming from the GW metric perturbation given in the TT gauge. Note that the distance L from the Earth to the pulsar is the distance in the TT gauge. In this frame the Earth and pulsar are at fixed spatial coordinates, so even as the GW is passing the coordinates $\mathbf{x}_e = 0$ and $\mathbf{x}_p = L\hat{\mathbf{p}}$ remain defined by the Earth and pulsar positions, respectively. By comparing the observed times between two consecutive pulses, say at t_{obs} and at $t'_{obs} = t_{obs} + T$, it can be shown that the change in pulse arrival time satisfies

$$\Delta T = \frac{p^i p^j}{2} \int_{t_{em}}^{t_{em}+L} dt' \left[h_{ij}^{TT}\left(t' + T, (t_{em} + L - t')\hat{\mathbf{p}}\right) - h_{ij}^{TT}\left(t', (t_{em} + L - t')\hat{\mathbf{p}}\right)\right]. \tag{1.3.5}$$

The pulse period T ($\mathcal{O}(10^{-3})$ s) is significantly smaller than the typical GW periods in this band ($\mathcal{O}(10^{8})$ s) so the first term in Eq. (1.3.5) can be Taylor-expanded[5] to first order in T about t'. Then we can insert the monochromatic plane wave solution from Eq. (1.1.8) and use Eq. (1.3.3) to get a general expression for the GW-induced redshift for a timed pulsar

$$z(t, \hat{\Omega}) = \frac{p^i p^j}{2(1 + \hat{\Omega}\cdot\hat{p})} \left[h_{ij}^{TT}(t_e, 0) - h_{ij}^{TT}(t_p, \mathbf{x}_p)\right], \tag{1.3.6}$$

[5]Note that this expansion only applies to the t' in the first argument of h_{ij}^{TT}, not the spatial argument $(t_{\rm em} + L - t')\hat{\mathbf{p}}$

where the times t_e and t_p represent the times at which the propagating GW passes the Earth and the pulsar, respectively, and $\mathbf{x}_p = L\hat{\mathbf{p}}$ denotes the pulsar position. The two times are related geometrically via $t_p = t_e - L(1 + \hat{\Omega} \cdot \hat{\mathbf{p}})$. Therefore there are two components to the induced redshift on the pulse arrival times: one stemming from the metric perturbation at the Earth and one from the metric perturbation at some earlier time at the pulsar. This is the genesis of what are called the "Earth term" and "pulsar term" of the GW signal in the PTA literature. Since the light travel time between from a pulsar to Earth is on the order of thousands of years, this allows us to potentially track GW sources across long periods of their evolution. Finally, we can integrate Eq. (1.3.6) over time to calculate the induced residuals, or the timing response, in a specific pulsar,

$$\begin{aligned} s(t, \hat{\Omega}) &= \int_0^t dt' \frac{\Delta\nu}{\nu} = \sum_A F^A(\hat{\Omega}) \int_0^t dt' \left[h_A(t_e) - h_A(t_p)\right] \\ &= \sum_A F^A(\hat{\Omega}) \left[s_A(t_e) - s_A(t_p)\right], \end{aligned} \tag{1.3.7}$$

where we have written the metric perturbation in terms of its two polarization modes and corresponding basis tensors $h_{ij}^{TT}(t, \hat{\Omega}) = \sum_A e_{ij}^A(\hat{\Omega}) h_A(t)$ and defined the PTA antenna pattern functions as

$$F^A(\hat{\Omega}) = \frac{1}{2} \frac{p^i p^j}{(1 + \hat{\Omega} \cdot \hat{p})} e_{ij}^A(\hat{\Omega}). \tag{1.3.8}$$

These functions describe how the GW signal will induce a larger signal response in some pulsars compared to others based on their locations on the sky relative to the GW source. To visualize this effect, we plot the two antenna pattern functions in Figure 1.3 for a hypothetical source at the location of the Virgo galaxy cluster, the closest massive galaxy

cluster to the Milky Way.

Given an expression for the GW strain of a particular source, we can use Eq. (1.3.7) to calculate what effect we should see in our pulsar timing experiments. To illustrate one example we will consider the single SMBHB in a circular orbit, whose expressions for the plus- and cross-polarization modes of the GW strain were derived earlier in Sec. 1.2.2. We will add in one additional variable, the polarization angle ψ of the GW, which will add another rotation to the strain components,

$$\begin{aligned} h_+(t,\psi) &= h_+(t)\cos 2\psi - h_\times(t)\sin 2\psi, \\ h_\times(t,\psi) &= h_\times(t)\cos 2\psi + h_+(t)\sin 2\psi. \end{aligned} \tag{1.3.9}$$

Under the assumption that the SMBHB frequency evolution is small, we can evaluate the integral in Eq. (1.3.7) and write down the two components of the signal response

$$\begin{aligned} s_+(t) &= \frac{2\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}\left[-\sin(2\Phi(t))(1+\cos^2(\imath))\cos(2\psi) - 2\cos(2\Phi(t))\cos(\imath)\sin(2\psi)\right], \\ s_\times(t) &= \frac{2\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}\left[2\cos(2\Phi(t))\cos(\imath)\cos(2\psi) - \sin(2\Phi(t))(1+\cos^2(\imath))\sin(2\psi)\right]. \end{aligned} \tag{1.3.10}$$

We plot examples of Eq. (1.3.10) in Figure 1.4 for three simulated pulsars placed randomly on the sky. The top plot shows if we only consider the Earth term contribution, and the bottom plot shows the full signal response of Eq. (1.3.7) including both the Earth and pulsar term contributions. Note that the pulsar term has a phase shift relative to Earth.

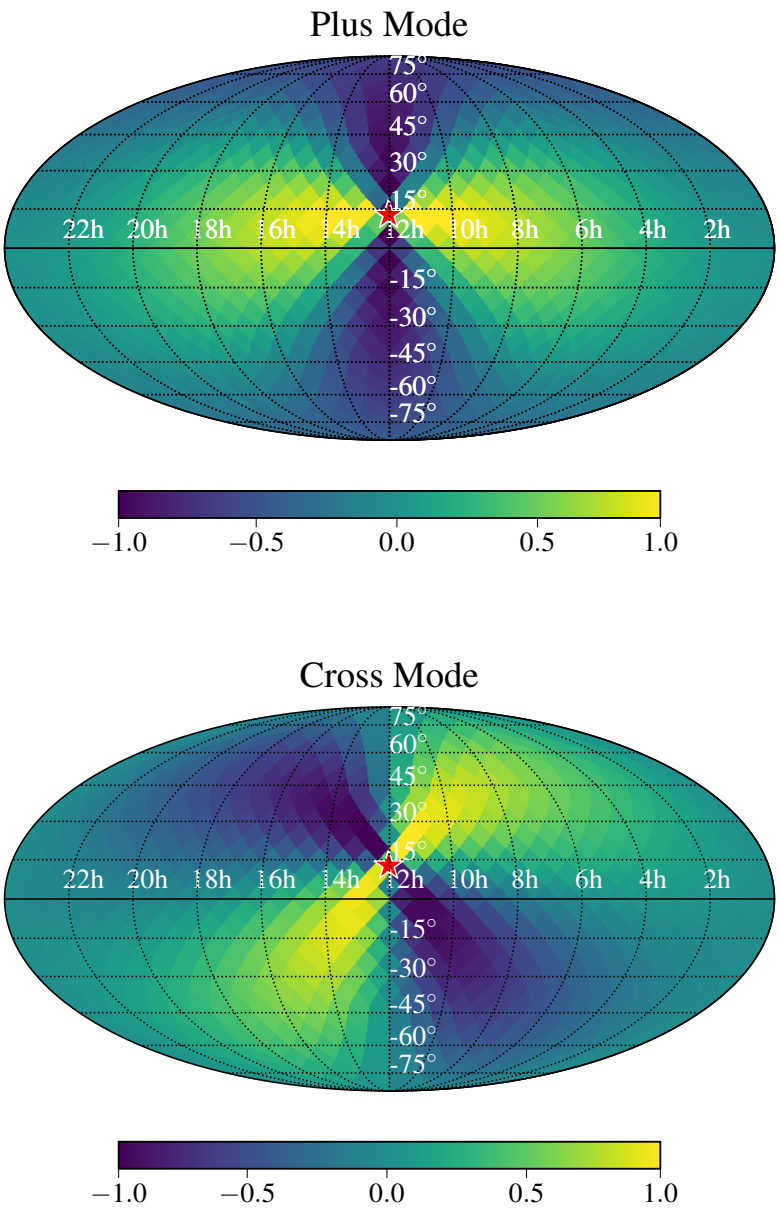

Figure 1.3: Antenna pattern response functions for $\hat{\Omega}$ in equatorial coordinates where $\hat{\Omega}$ represents the direction to the pulsar, calculated for $F^{+}(\hat{\Omega})$ (top) and $F^{\times}(\hat{\Omega})$ (bottom) for a source located at a RA of $12^{\mathrm{h}}27^{\mathrm{m}}$ and a Dec of $+12^{\circ}43'$. This happens to be the location of the Virgo galaxy cluster, denoted by a red star.

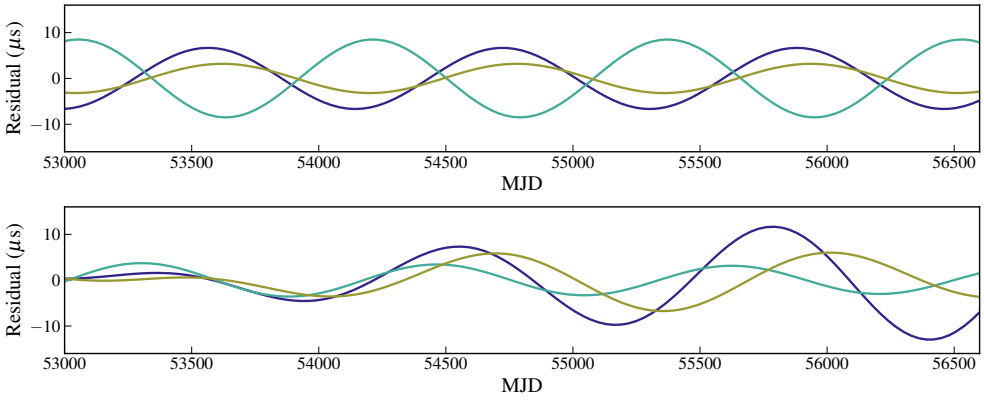

Figure 1.4: GW-induced residuals from a single SMBHB source in the timing of three different simulated pulsars placed randomly on the sky. The top plot displays only the Earth term, while the bottom plot shows both the Earth and pulsar terms. The SMBHB is placed at a RA of $12^{\mathrm{h}}27^{\mathrm{m}}$ and a Dec of $+12°43'$ at a distance $d_L = 15$ Mpc, with $\mathcal{M}_c = 5 \times 10^9 M_\odot$ and $f_{\mathrm{GW}} = 10$ nHz.

1.3.4 Timing Response to a Stochastic GWB

We are interested in the statistical properties of the signal response and relating it to the characteristic strain of the GWB. In particular we can look at the power induced on multiple different pulsars and calculate how they are correlated, therefore providing a way to differentiate the GWB from other potential stochastic signals. The following derivation closely follows the one given in Maggiore (2018), and for additional examples we point the reader to Taylor (2021) or Anholm et al. (2009). Recalling Eq. (1.2.16) we can calculate the induced redshift for pulsar a,

$$z_a(t,\hat{\Omega}) = \sum_A \int_{-\infty}^{\infty} df \int d^2\hat{\Omega} F_a^A(\hat{\Omega}) h_A(f,\hat{\Omega}) e^{-2\pi i f t} \left[1 - e^{2\pi i f L_a (1+\hat{\Omega}\cdot\hat{\mathbf{p}_a})}\right]. \tag{1.3.11}$$

Every pulsar will have its own unique response characterized by its position on the sky and distance from Earth. It can be shown that the correlation between two pulsars (a,

b), under the same set of conditions for the background as in Sec. 1.2.3, follows

$$\langle z_a(t,\hat{\Omega}) z_b^*(t',\hat{\Omega}')\rangle = \frac{1}{2}\sum_A \int_{-\infty}^{\infty} df \int \frac{d^2\hat{\Omega}}{4\pi} F_a^A(\hat{\Omega}) F_b^A(\hat{\Omega}) S_h(f) \kappa_{ab}(f,\hat{\Omega}), \tag{1.3.12}$$

where the exponential terms have been collected as

$$\kappa_{ab}(f,\hat{\Omega}) = \left[1 - e^{2\pi i f L_a(1+\hat{\Omega}\cdot\hat{\mathbf{p}}_a)}\right]\left[1 - e^{2\pi i f L_b(1+\hat{\Omega}\cdot\hat{\mathbf{p}}_b)}\right]. \tag{1.3.13}$$

The four terms in this expression correspond to an Earth term – Earth term correlation, a pulsar term – pulsar term correlation, and two Earth term – pulsar term correlations. We can neglect three of the four terms that would otherwise contain rapidly oscillating exponentials and keep only the Earth term – Earth term component. Only $\kappa_{ab}(f,\hat{\Omega}) = 1$ remains (Note: for the exact same pulsar this would be $\kappa_{aa}(f,\hat{\Omega}) = 2$). For a deeper discussion on this topic we refer the reader to Mingarelli et al. (2013). The spatial integral in Eq. (1.3.12) can now be solved analytically, to give the overlap reduction function (ORF) for the GWB. The final result is

$$\sum_A \int \frac{d^2\hat{\Omega}}{4\pi} F_a^A(\hat{\Omega}) F_b^A(\hat{\Omega}) = \xi_{ab}\ln\xi_{ab} - \frac{1}{6}\xi_{ab} + \frac{1}{3}, \tag{1.3.14}$$

where $\xi_{ab} = (1-\cos\theta_{ab})/2$ for the angular separation θ_{ab} between pairs of pulsars on the sky. This ORF is a famous result known as the Hellings-Downs (HD) curve (Hellings & Downs, 1983) and defines the expected spatial correlation pattern in pulsar timing observations under the influence of a stochastic GWB. It is often normalized to unity for identical pulsars $a = b$ in the literature and for distinct pulsar pairs $a \neq b$ the HD curve is commonly written as

$$\Gamma_{ab} = \frac{3}{2}\xi_{ab}\ln\xi_{ab} - \frac{\xi_{ab}}{4} + \frac{1}{2}. \tag{1.3.15}$$

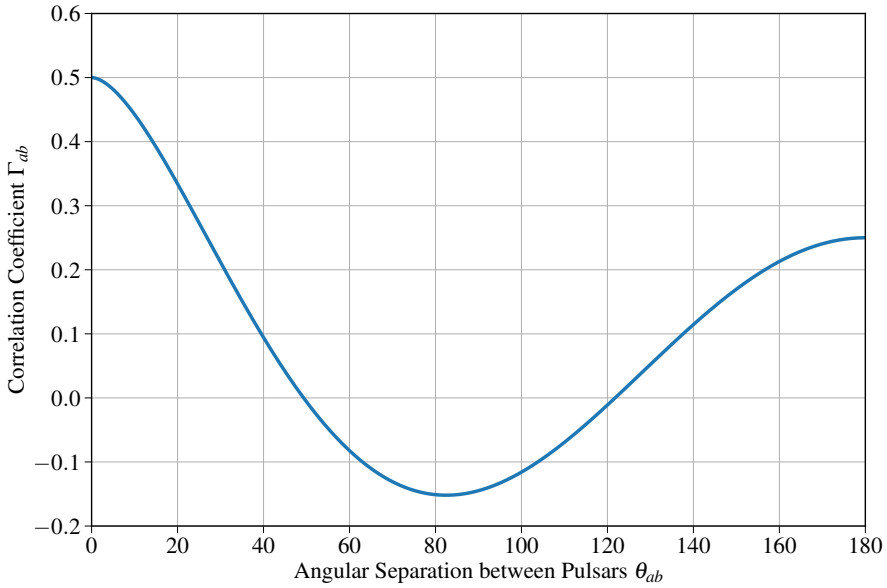

Figure 1.5: The Hellings-Downs curve plotting the correlation coefficient Γ_{ab} for two pulsars a and b as a function of their angular separation θ_{ab}. The function is normalized such that $\Gamma_{ab}(0) = 0.5$.

We plot this function in Figure 1.5. It displays the unique signature of a stochastic GWB and is the primary result that all modern data analysis pipelines in PTAs seek to find. There are a few important characteristics of the HD curve that we note below:

- The value of the HD curve at an angular separation of 0° is 0.5, which is also the maximum of the function. This represents two pulsars that are at nearly identical locations on the sky with Earth-term contributions to the residuals that would be completely correlated. The exact value of 0.5 is a choice of normalization, as explained above, such that the correlation between two identical pulsars would be 1 as expected.

- At an angular separation of 180° the correlation coefficient is exactly half of that at 0°. This is due to the factor of $(1 + \hat{\Omega} \cdot \hat{\mathbf{p}})$ in the denominator of Eq. (1.3.8), and can

be seen explicitly by calculating (1.3.14) for pulsars pointing in the same direction ($\hat{\mathbf{p}}_a = \hat{\mathbf{p}}_b$) and in opposite directions ($\hat{\mathbf{p}}_a = -\hat{\mathbf{p}}_b$).

- The minimum of the HD curve is not at $90°$. This is again due to the fact that the antenna pattern functions are not symmetric under the change $\hat{\mathbf{p}} \to -\hat{\mathbf{p}}$. A decomposition of the HD curve into Legendre polynomials $P_l(\cos\theta_{ab})$ shows that although the dominant contribution to the HD correlation function is the quadrupolar component for $l = 2$, there are small contributions from higher l-modes. A complete derivation showing the presence of these higher order contributions can be found in Gair et al. (2014). This combines to slightly offset the minimum of Eq. (1.3.15) to less than $90°$.

Our ability to constrain the presence of a GWB in our data stems from how well we can make out the above features and the distinct shape of the HD curve. It is therefore very beneficial to time many pulsars and time those that have a wide distribution in angular separations, allowing us to both observe the HD curve in its entirety as well as appropriately constrain our findings across all regions of Fig. 1.5. For an array of N pulsars, the total number of possible pairs is $N_{\text{pairs}} = N(N-1)/2$. For the NANOGrav 12.5-year dataset this constituted 990 pairs (45 pulsars), and for the NANOGrav 15-year dataset this grew to $2,211$ pairs (67 pulsars).

Finally we have a full prescription for defining the statistical properties of the timing repsonse, via Eq. (1.3.12). In Sec. 1.2.3 we characterized the GWB through its fractional energy density as a function of its characteristic strain. Now when considering the response we are again interested in h_c, this time as it relates to the cross-power spectral density of the induced timing residuals. This relation is given by

$$S_{ab}(f) = \Gamma_{ab}\frac{h_c^2(f)}{12\pi^2 f^3}. \tag{1.3.16}$$

As we will see in the following chapters, this is the description of the GWB that enters into our statistical inference models. If we assume the GWB to arise from a population of SMBHBs, then we can treat h_c as a power law according to Eq. (1.2.23). This is the predominant modeling choice for the signal in detection pipelines, and will be discussed further in the next chapter.

In 2023 PTA collaborations around the globe reported the first conclusive evidence of a stochastic GWB signal, consistent with the expected HD correlation signature. Assuming a $f^{-2/3}$ characteristic strain spectrum, the NANOGrav collaboration found $3-4\sigma$ confidence of the presence of correlations in their dataset, with a strain amplitude of $2.4^{+0.7}_{-0.6} \times 10^{-15}$, calculated for a reference frequency of $1\ \mathrm{yr}^{-1}$ (Agazie et al., 2023a). The EPTA collaboration similarly found 3σ support for a GWB with a strain amplitude of $2.5^{+0.7}_{-0.7} \times 10^{-15}$ (EPTA Collaboration et al., 2023), the PPTA measured an amplitude of $2.04^{+0.25}_{-0.22} \times 10^{-15}$ at 2σ confidence (Reardon et al., 2023a), and the CPTA found 4.6σ evidence for a correlated signal with log amplitude of $-14.4^{+1.0}_{-2.8}$ and a spectral index in the range $\alpha \in [-1.8, 1.5]$ (Xu et al., 2023). Detecting and characterizing the GWB is no longer a future goal but instead a reality of current PTA experiments, and with more time and observations we can continue to study the background signal, look for anisotropies, and eventually gain the sensitivity to probe individual sources.

1.4 DISSERTATION OVERVIEW

This dissertation covers the development and testing of novel data analysis methods aimed at studying primarily low-frequency GW sources with PTAs, as well as perusing preliminarily its applicability to LISA data and the millihertz GW band. In particular, these methods are designed to better deal with the complex parameter spaces and large data volumes that otherwise hinder traditional analysis pipelines, rendering them unten-

able long-term.

The remainder of this dissertation is structured as follows. First in Chapter 2 we give an overview of Bayesian statistics and Markov Chain Monte Carlo algorithms. We then discuss the Hamiltonian Monte Carlo algorithm that is used exclusively throughout the analysis pipelines developed for this dissertation. We conclude the chapter by introducing the framework for Bayesian analysis of PTA data and its potential pitfalls in computational performance. Chapter 3 is the first of two chapters where we apply a Hamiltonian sampling framework to PTA data analysis and robustly test its accuracy and efficiency against current pipelines. We thoroughly analyze a suite of simulated datasets and provide scaling arguments on real NANOGrav data supporting the efficacy of our methods to future data releases. In Chapter 4 we extend our analysis to a complete Bayesian pipeline for detection and joint analysis of both stochastic GWB signals and individual SMBHB sources in PTA data. We analyze multiple classes of source injections, including ones where we expect covariance between the GWB and SMBHB signals. We also demonstrate the improved speed and efficiency of our pipeline over current methods both in full detection runs as well as in setting upper limits across the full sky. In Chapter 5 we leave the PTA frequency band behind in favor of the millihertz GWs probed by LISA. We focus on one particular source, extreme mass ratio inspirals, that as of yet has no working robust method of parameter estimation to be included as part of a global fit for millihertz sources. This is due to their long-duration, complex waveforms that involve a large, intercorrelated parameter space, which leads to existing inference pipelines struggling to faithfully recover the complete signal. We present evidence supporting this by highlighting the difficulties of current LISA codes at accurately analyzing simulated source injections, and review preliminary work towards using Hamiltonian sampling to attack the problem of extreme mass ratio inspiral data analysis. Finally in Chapter 6 we summarize the key

contributions of this dissertation, discuss their implications, and outline potential future work.

CHAPTER 2

Bayesian Inference and Markov Chain Monte Carlo Methods

Physicists are quite fond of the concept of the quintessential "spherical cow"[1]. It is a metaphor poking fun at the host of assumptions physicists make when developing theoretical models. Simplifications like neglecting air resistance and friction are most commonly referenced, or assuming idealized noise in a detector. This greatly aids the development of testable models, but when it comes time to actually perform an experiment those assumptions are no longer necessarily true. One way or another everything is starting from some telescope, or oscilloscope, or any other non-idealized instrument. Simply put, the data we get are the data we get. It will contain information we care about, and lots of information about which we really don't care at all. Connecting back to GW experiments, the detector output data are composed of a combination of noise and, hopefully, a signal

$$d(t) = h(t) + n(t). \tag{2.0.1}$$

In general we are not able to model everything exactly, so distinguishing one from the other needs to be done in a probabilistic manner, treating all of the underlying processes statistically. We need some way of stating the level of confidence we have in any results we claim, for example how sure we are that GWs are present in our data. There are two leading approaches we can use to address our problem of statistical inference.

On one hand is the frequentist philosophy. The data are considered random, one of

[1]The origin of the spherical cow phrase is said to have come from, perhaps unsurprisingly, Wisconsin (Harte, 1985).

many hypothetical sets of observations, and the signal parameters are unknown yet fixed. The uncertainty comes solely from sampling the data. Probability in this context represents how frequently we should expect to measure our data given some signal parameters. There are numerous examples of frequentist statistics developed in the PTA literature (Chamberlin et al., 2015; Vigeland et al., 2018; Sardesai et al., 2023; Gersbach et al., 2025), and they remain a core part of GW detection. On the other hand is the Bayesian philosophy. The data are fixed and it is the signal parameters that are unknown and random. In this context the probability can be thought of as our degree of belief in a signal given the data we have taken, and as we take more data that degree of belief will change. What we infer are probability distributions of our model parameters. Bayesian parameter estimation is at the heart of modern PTA analysis pipelines (van Haasteren et al., 2009; van Haasteren & Levin, 2010; Ellis et al., 2013; Ellis, 2013; Lentati et al., 2013). The work presented in this dissertation is entirely Bayesian in nature, and we will not discuss frequentist statistics further.

2.1 BAYES' THEOREM

Bayesian inference is defined by the aptly name Bayes' Theorem (for an extended discussion on Bayesian inference for the sciences, see for example Jaynes & Baierlein (2004) and Gregory (2010)), which calculates the posterior probability distribution, denoted $P(\theta|d,\mathcal{H})$ for a set of signal parameters θ given the observed data d under a model assumption $\mathcal{H}$

$$P(\theta|d,\mathcal{H}) = \frac{P(\theta|\mathcal{H})P(d|\theta,\mathcal{H})}{P(d|\mathcal{H})}. \tag{2.1.1}$$

The term $P(\theta|\mathcal{H})$ is the prior probability distribution and represents any prior knowledge we have on the model. The likelihood $P(d|\theta,\mathcal{H}) \equiv L(\theta)$ is a function computing the probability distribution of the data given some signal parameters. Note that inference pipelines

normally work in terms of the log of the likelihood, $\mathcal{L}(\theta) \equiv \ln L(\theta)$. The denominator term $P(d|\mathcal{H})$ is called the evidence and describes the probability distribution of the data for all possible values of the signal parameters, and is akin to the likelihood of the model being tested $P(d|\mathcal{H}) = \int P(d|\theta, \mathcal{H})P(\theta|\mathcal{H})d\theta$. When considering only a single model at a time for parameter estimation, this term acts only as a normalization factor and is ignored. When considering multiple different models of the data, say $\mathcal{H}_1$ and $\mathcal{H}_2$, and selecting among them, computing the evidence is required. The methods discussed in this dissertation revolve around parameter estimation, rather than model selection, and from here on we will drop the $\mathcal{H}$ variable and corresponding evidence term from any pertinent equations.

Statistical inference is typically a multi-dimensional affair, as our signal parameters can be defined as a set $\theta = \{\theta_k\}$. Perhaps we are interested in the probability distributions of the chirp mass and frequency of a SMBHB, or the amplitude and slope of the PSD of a stochastic background signal. Bayes' theorem outputs the full multi-dimensional posterior distribution on θ, accounting for noise and other nuisance parameters that we otherwise do not care about. Breaking the full posterior down to one-dimensional distributions over the parameters of interest is done through a process called marginalization. For a single parameter $\theta_k \in \theta$ this is accomplished formally by integrating over all remaining variables in the set $\theta_l \in \theta$, $l \neq k$, written as:

$$P(\theta_k|d) = \int P(\theta|d) \prod_{l \neq k} d\theta_l \ . \tag{2.1.2}$$

For simple models of low dimension, it is possible that Eq. (2.1.2) will be analytically solvable and we can obtain exact expressions for the one-dimensional distributions of our parameters. No additional numerical methods would be necessary. For nearly all practical GW analysis examples, however, this will not be the case.

2.2 MARKOV CHAIN MONTE CARLO METHODS

We typically will not be able to carry out Eq. (2.1.2) exactly. As we will discuss later in the chapter, the models we are interested in are very high-dimensional, and their posterior distributions will not have a simple functional form. In these cases where we cannot obtain an analytic expression for the individual distributions we can turn to numerical methods and generate a set of random samples. Integrals can then be approximated as a sum over a large number of samples N as

$$\int d\theta f(\theta) P(\theta|d) \approx \frac{1}{N} \sum_{\theta_s \sim P(\theta|d)} f(\theta_s), \tag{2.2.1}$$

where $f(\theta)$ is an arbitrary multivariate function and $\theta_s \sim P(\theta|d)$ represents a sample θ_s drawn from a distribution $P(\theta|d)$. This technique is known as Monte Carlo integration. Markov chains are stochastic processes describing sequences of events where for each state its probability can only depend on the probability of the previous state. These processes can be used to generate random samples from a probability distribution. Markov chain Monte Carlo (MCMC) methods (Hastings, 1970) are a class of algorithms that combine both of the above approaches to approximate target probability distributions. For a large enough number of draws N we should reach the true posterior. MCMC algorithms must satisfy the principle of detailed balance, meaning the Markov chain must be reversible. MCMC chains are also ergodic, indicating that they can eventually reach the target distribution regardless of the starting point.

The most commonly used example of a MCMC method is the Metropolis-Hastings algorithm (MH, or MH MCMC; Metropolis et al. 1953). The outline for MH MCMC is shown in Algorithm 1. We start with an initial point in our parameter space θ^0 and the value of the posterior at this point $P(\theta^0|d)$. Beginning at sample number i a new sample y

is generated according to a proposal distribution $q(y|\theta^{i-1})$. Next we evaluate the posterior at the new sample and calculate the Hastings ratio,

$$H = \frac{P(y|d)q(\theta^{i-1}|y)}{P(\theta^{i-1}|d)q(y|\theta^{i-1})}. \tag{2.2.2}$$

The acceptance fraction is then given by $\alpha = \min\{1, H\}$. This proceeds iteratively until a desired number of samples N is reached.

Algorithm 1 Metropolis-Hastings

Given θ^0
for $i = 1, 2, ..., N$ **do**
  $y \sim q(\theta^{i-1}|y)$
  $\alpha = \min\left\{1, \frac{P(y|d)q(\theta^{i-1}|y)}{P(\theta^{i-1}|d)q(y|\theta^{i-1})}\right\}$
  $r \sim U(0, 1)$
  if $r < \alpha$ **then**
    $\theta^i \leftarrow y$
  else
    $\theta^i \leftarrow \theta^{i-1}$
  end if
end for

The output of an MCMC pipeline is a Markov chain for the parameter set. Initially the chain will spend some period of time searching for high probability regions representative of the target distribution. This is called the burn-in phase, and the standard operating procedure is to discard some percentage of samples from the beginning of the chain to remove this early exploration. The remaining samples can be plotted as histograms, applied to compute quantiles and upper limits, and used to make statistical statements about the model parameters.

An important notion that comes up in MCMC sampling is how to know when we have fully explored the target distribution. The principle of Monte Carlo integration only demands that N be large, but the actual number is left up to the user's discretion. As a

an initial diagnostic we can look directly at the Markov chain, a visual inspection called a trace plot, to check by eye if the MCMC sampler has converged to the target distribution or if it is still exploring the parameter space. There are also many diagnostic tests that can properly quantify chain convergence. The Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992), which is the primary test used for the MCMC analyses throughout this dissertation, splits the chain into multiple subsets and compares the parameter variances within each subset with the variances between the different split chains. The statistic converges to 1 in the limit $N \to \infty$, therefore we can quantify convergence by setting a threshold on this statistic for input Markov chains. Usually this is set as $\hat{R} < 1.01$.

One other property of MCMC analyses worth discussing is the chain autocorrelation length, sometimes also referred to as the autocorrelation time, denoted by τ. It measures the degree of similarity between successive points in the chain. The autocorrelation length assesses how far one must jump through the chain to find samples that are independent, and in this way is a direct measure of the Monte Carlo error on any integrals. It is then useful to instead consider the effective number of samples in an MCMC chain $N_{\text{eff}} = N/\tau$ in determining whether a analysis has been run for sufficient time.

At the center of many PTA Bayesian inference pipelines there is a MCMC random sampler at work. The vast majority of them employ the MH algorithm with added options for curated jump proposals to help with chain convergence. A major pitfall in using this algorithm for PTA inference is that it becomes increasingly inefficient as both the dimensionality and complexity of the problem grow. What occurs is most of the probability volume will be concentrated in narrow regions of high probability, making it difficult for a basic random-walk technique to sample smoothly through the space. Small proposal steps can lead to high autocorrelation, and large proposal steps can lead to a high rejection rate of samples. As we will see later in Sec. 2.4 the dimension of a typical GW search

with PTAs is on the order of $\mathcal{O}(100)$ parameters that can in cases be strongly correlated. While any GW search we want to perform can in theory be done by brute force, running the MCMC sampler indefinitely until it has satisfied some predetermined convergence metric, in reality it becomes a question of practicality. Time and computing resources are not infinite, at least on the scale of our experiments, and we need ways to more efficiently run our analyses even as our models grow more complex and we steadily accumulate data.

2.3 THE HAMILTONIAN MONTE CARLO ALGORITHM

A number of different MCMC samplers have been developed over years as methods to fully explore complicated parameter spaces and break the curse of dimensionality that plagues random-walk Metropolis alternatives. If the conditional probabilities, meaning the probabilities of each parameter assuming the others are fixed, are known then we can use Gibbs sampling (Geman & Geman, 1984) to directly sample the full posterior distribution. In cases where these conditional distributions are not analytically known we can instead use more sophisticated techniques to draw independent samples that have lower autocorrelation and are otherwise more efficient than a random-walk. One such example is the Hamiltonian Monte Carlo (HMC) algorithm (Duane et al., 1987; Neal, 2011), a method rooted in statistical mechanics that treats the evolution of trajectories in parameter space like that of a many-particle system. It is more adept at treating high-dimensional models and has significantly reduced autocorrelation compared to an MH MCMC sampler. In this section we provide a detailed derivation to Hamiltonian sampling.

Briefly diverging for what we hope is an interesting history lesson, the origins of the HMC algorithm dates back to the 1950s when two competing methods were being developed to simulate the states of molecules. There was the method of MH MCMC outlined

in Metropolis et al. (1953) treating their motions as random variables, and there was a separate method in which the motion was treated deterministically according to Newton's laws of motion (Alder & Wainwright, 1959). For over three decades they existed separately, then Duane et al. (1987) joined both methods into a single algorithm, using deterministic particle motion to inform a Markov chain generator. It was then applied to problems in lattice field theory of quantum chromodynamics. Considering the new algorithm was a hybrid of the two earlier techniques, it was originally called Hybrid Monte Carlo, not Hamiltonian Monte Carlo as we just defined. It was later in Neal (2011) that the method colloquially changed its name.

The HMC algorithm borrows from statistical mechanics the notion of the canonical ensemble, describing the states of an internal system in thermal equilibrium with a heat bath at a constant temperature T. Consider a distribution of n particles with positions q and momenta p, similar to a microstate of the system. We are interested in finding the probability of the system being in any particular microstate, a result given by the Boltzmann distribution

$$P(p,q) = \frac{1}{Z}\exp^{-E(p,q)/k_B T}, \tag{2.3.1}$$

where Z is the classical partition function, $E(p,q)$ the total energy of the system, and k_B the Boltzmann constant. The total energy is represented by a separable[2] Hamiltonian $H(p,q) = U(q) + K(p)$ for a potential energy function $U(q)$ and a kinetic energy function $K(p)$. Connecting back to the problem of statistical inference, we are treating the positions q of the particles like the points in parameter space in our model. What this has become is an example of data augmentation, where we are introducing an auxiliary set of variables in the momenta p to assist in calculating the probability distributions of said parameters.

[2]We choose the Hamiltonian such that it is seperable.

This means we have freedom to choose the form of our joint density, and we pick $P(p,q) = P(q)P(p|q)$. If we apply this against some data d, we can write

$$P(q|d)P(p|q,d) = \frac{P(q)P(d|q)}{P(d)}P(p|q,d) = \frac{1}{Z}e^{-(U(q)+K(p))/k_BT}, \tag{2.3.2}$$

where we have used Eq. (2.1.1) to relate the output of Bayes' Theorem to the Boltzmann distribution. Taking the temperature T to be such that $k_BT = 1$, we find the following relationships for the partition function, potential energy, and kinetic energy terms

$$Z = P(d), \qquad U(q) = -\ln\left[P(q)P(d|q)\right], \qquad K(p) = -\ln\left[P(p|q,d)\right]. \tag{2.3.3}$$

We explicitly choose the momenta p to be independent of parameters q and data d, making $P(p|q,d) = P(p)$. Note that this analog also implies that for the purposes of parameter estimation we can ignore the partition function term, represented by the model evidence. We are able to choose our kinetic energy function describing the auxiliary parameters. The standard choice in HMC implementations is that it is quadratic in each of its dimensions $K(p) = \sum_i p_i^2/2m_i$, translating to a zero-mean multivariate Gaussian distribution with the m_i representing independent variances. In practice the variances are taken to be unity. Our Hamiltonian is then

$$H(p,q) = -\ln\left[P(q)\right] - \mathcal{L}(q) - \sum_{i=1}^{n}\frac{p_i^2}{2m_i}, \tag{2.3.4}$$

where we have written the second term using the shorthand notation for the log of the likelihood function for our model.

We want to evolve this system to explore different states. This is accomplished by solving Hamilton's equation of motion

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial q}. \tag{2.3.5}$$

These equations need to be solved numerically. Since we require time-reversibility to satisfy the detailed balance condition of MCMC methods, a second-order symplectic integrator, often called a leapfrog method, is usually chosen. We also need to define an integration step size ϵ and number of steps L to take.

Algorithm 2 Hamiltonian Monte Carlo

Given q^0, ϵ, N, L:
Sample $p^0 \sim \mathcal{N}(0, I)$
for $i = 1, 2, ..., N$ **do**
  $q^i \leftarrow q^{i-1}, p^i \sim \mathcal{N}(0, I)$
  for $j = 1, 2, ..., L$ **do**
    $p^i, q^i \leftarrow \text{Leapfrog}(p^i, q^i, \epsilon)$
  end for
  $\alpha = \min\left\{1, \frac{\exp\{-H(p^i, q^i)\}}{\exp\{-H(p^{i-1}, q^{i-1})\}}\right\}$
  $r \sim U(0, 1)$
  if $r < \alpha$ **then**
    $p^i, q^i \leftarrow p^i, q^i$
  else
    $p^i, q^i \leftarrow p^{i-1}, q^{i-1}$
  end if
end for

function LEAPFROG(p, q, ϵ)
  $p' \leftarrow p + (\epsilon/2)\nabla_q \mathcal{L}(q)$
  $q' \leftarrow q + \epsilon p'$
  $p' \leftarrow p' + (\epsilon/2)\nabla_q \mathcal{L}(q')$
  return p', q'
end function

At last we have the necessary tools to describe the HMC algorithm, which is summarized in Algorithm 2. Given an initial point in the parameter space[3] q^0 and beginning at

[3] which is different than the joint position-momenta space on which our Hamiltonian is defined, but this will be rectified later

sample number i we can generate a new point in the following manner: Start by sampling the momenta variables, then evolve the system using Eq. (2.3.5) for L integration steps of size ϵ. The two integration hyperparameters are chosen by the user, and in typical implementations usually fall within $L \sim 10 - 100$ and $\epsilon \sim 0.001 - 0.1$. After stopping, evaluate the Hamiltonian at both the previous point and the new end point to calculate the acceptance fraction for HMC. Finally, since we are left with a joint distribution of both parameters and momenta $P(p, q)$, we can marginalize over the momenta to get our desired posterior distribution.

To summarize, the HMC algorithm generates a Markov chain by simulating Hamiltonian dynamics in a phase space corresponding to the target distribution plus some fictitious momenta. Proposal draws are informed by first-order gradient information of the model likelihood, leading to distant states being proposed that both are less correlated from previous steps and have a high rate of acceptance. This leads to a significant decrease in the autocorrelation of the Markov chains compared to random-walk Metropolis and an overall improvement in efficiency, particularly for high-dimensional and complex models. This improvement was quantified in Creutz (1988), showing that the generation of independent samples scales with the model dimension d as $\mathcal{O}(d^2)$ for MH MCMC and $\mathcal{O}(d^{5/4})$ for HMC. For details regarding the two scaling relations, see the Appendix after the conclusion of this dissertation. In Figure 2.1 we plot a side-by-side comparison of the two MCMC methods in sampling the Rosenbrock distribution (Rosenbrock, 1960), a popular test function in optimization problems that has also been adapted as a benchmark for MCMC algorithms (Goodman & Weare, 2010). Each sampler is run for 500 iterations, with identical step sizes of $\epsilon = 0.1$ and a number of integration steps $L = 16$ for the HMC example. We see that the random-walk sampler quickly gets stuck in a local extrema of the distribution and cannot break out to explore the full space, whereas the HMC sampler

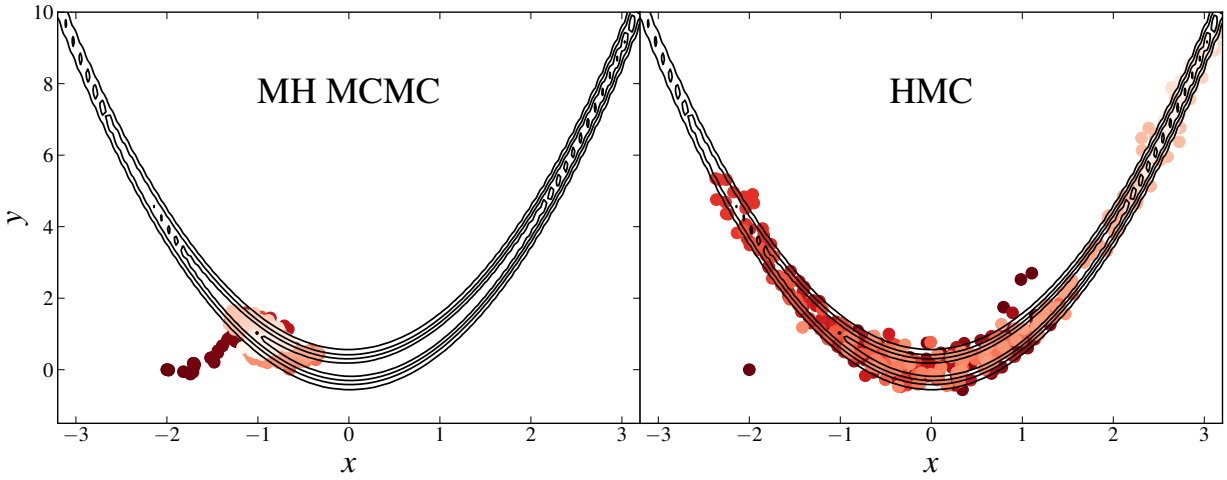

Figure 2.1: Two Markov chains run on the Rosenbrock density, given by $P(x, y|a, b) \propto \exp\{-[(a-x)^2 + b(y-x^2)^2]/20\}$ with $a = 1$ and $b = 100$. The left panel shows the performance of a MH MCMC sampler at exploring at the posterior, and the right panel displays similarly for a HMC sampler. Both routines were run for $N = 500$ samples and given similar starting points at $(x, y) = (-2, 0)$. The proposal distribution for the MH MCMC sampler was defined using a Gaussian proposal scheme centered on the current state $q(y|\theta_i) = \mathcal{N}(\theta_i, 0.25)$.

in the same number of samples can nearly traverse the entire target distribution.

2.3.1 No-u-turn Sampling

We have introduced Hamiltonian sampling as a more sophisticated alternative to MH MCMC methods, with direct application to the complex GW models we wish to analyze with PTAs. Implementation, however, comes with a few potential setbacks worth identifying and rectifying. First and foremost, recall that HMC proposals require calculating the gradient of the model likelihood. Gradients are computationally expensive to compute, and any differentiation scheme needs to be very precise to maintain high sample acceptance rates. For this reason numerical finite difference schemes are usually left behind in favor of automatic differentiation codes, which can work for all model likelihood functions, even those that are not analytically differentiable. Additionally, in defining the

HMC algorithm we have implicitly added two new hyperparameters, those being the two integration variables (L, ϵ). If they are not tuned properly the sampler risks defaulting to a host of different erroneous or inefficient behavior. For example if L is set too large the algorithm moves past acceptable sample proposals and wastes computation, and if it is set too small then it begins to default back to random-walk behavior.

Optimal values for the integration hyperparameters can be found through brute force, launching multiple preliminary tuning runs and interpreting the output to set the production level values for the final pipeline. This is very costly in terms of computation time, requires many instances of user intervention, and collectively counteracts the gains in efficiency we can otherwise achieve using a well-tuned HMC sampler. The No-u-turn sampler (NUTS), developed in Hoffman & Gelman (2011), avoids any hand tuning of the algorithm by dynamically determining both L and ϵ. The integration step size is adapted using a dual averaging algorithm described in Nesterov (2009) up to a number of samples $N_{\text{adapt}} < N$, akin to the burn-in phase discussed earlier, and then ϵ remains fixed for the remainder of the run.

Tuning the number of integration steps L is the crux of the NUTS algorithm. The procedure is as follows: at each iteration we construct a binary tree of sample proposals defining our trajectory through phase space defined by (p, q). The tree is defined to start at height $j = 0$ with the most current point in the chain. For increasing j we choose randomly for our simulation to move forwards or backwards in time, then simulate 2^{j-1} leapfrog steps in that direction adding to our path. To each new subtree we assign the states (p^+, q^+) and (p^-, q^-) to the rightmost and leftmost leaves, respectively. The simulation continues until the following condition is met for any subtree

$$\left(q^+ - q^-\right) \cdot p^- < 0 \quad \text{or} \quad \left(q^+ - q^-\right) \cdot p^+ < 0. \tag{2.3.6}$$

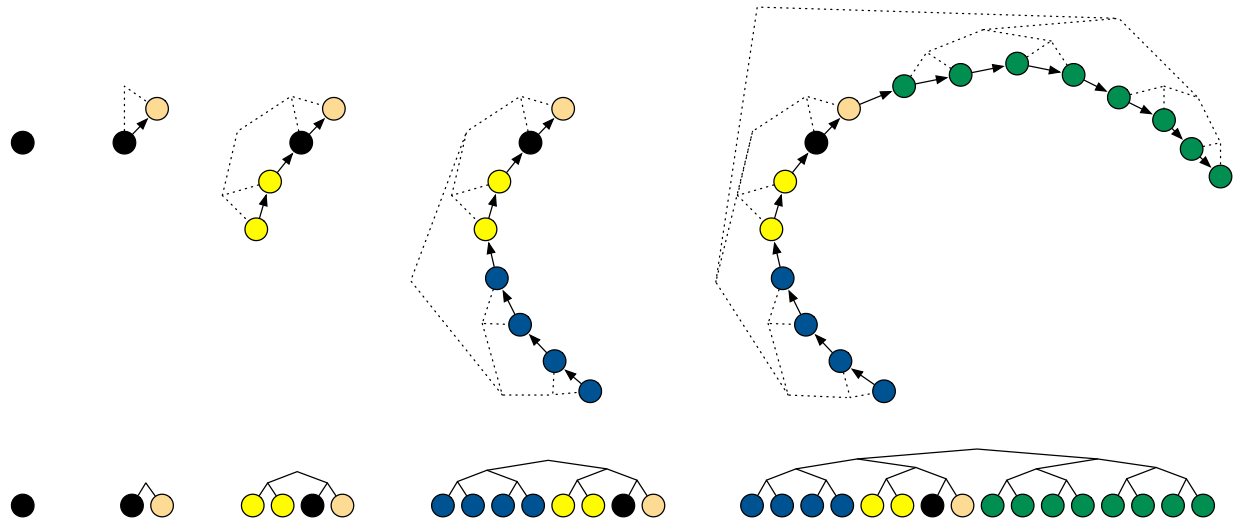

Figure 2.2: Diagram showing the construction of a binary tree for the NUTS algorithm. Four doublings of the tree are shown, with the top figures showing the two-dimensional trajectory and the bottom figures displaying the binary tree evolution. Figure from Hoffman & Gelman (2011).

This condition is equivalent to stating we have reached the point where the distance between proposal points has begun to decrease, i.e., our trajectory has begun to make a u-turn. Figure 2.2 illustrates the building of the binary tree in the NUTS algorithm. Once this stopping condition is met, the final sample is chosen randomly from the traced leapfrog path.

With these tools in hand, we now have the capability to run Hamiltonian sampling pipelines while dynamically tuning the integration hyperparameters to fit our model choice, completely eliminating the need for any additional user input or preliminary runs. This means that a NUTS routine allows us to fully leverage the increased efficiency of the HMC algorithm in analyzing our GW models. Indeed, NUTS has seen widespread adaptation over recent years to become the default implementation in many different analysis software and open-source projects utilizing HMC sampling.

2.4 PULSAR TIMING DATA ANALYSIS

Regardless of which Monte Carlo sampling method we ultimately wish to employ, we need a way to connect it to our models of interest. This brings us full circle back to the beginning of the chapter, asking ourselves how to define our signal and our noise, and how to relate this to Bayes' theorem. In this section we provide an overview of how we model PTA data, what noise components we consider, and how we construct the PTA likelihood function.

2.4.1 The Data Model

Recall from Eq. (1.3.2) that the pulsar timing residuals, the primary data products for GW searches, are written as the difference between the observed TOAs and the predicted arrival times from a deterministic timing model. Here we will focus on how we can model a pulsar's TOAs. Generally speaking the TOAs will contain a collection of both deterministic and stochastic processes, meaning we can at a top level write them for a single pulsar as

$$t_{\rm obs} = t_{\rm det} + t_{\rm sto}, \tag{2.4.1}$$

where the $t_{\rm det}$ and $t_{\rm sto}$ represent the induced delays from deterministic and stochastic sources, respectively.

First we consider the deterministic component. The primary factor is the pulsar timing model discussed in Sec. 1.3.2, and given explicitly in Eq. (1.3.1). If this best-fit timing ephemeris was exactly correct, the only factors left in the residuals should be GWs and noise. However, it is important to consider the possibility that the fit has absorbed some of the other processes. We assume that the difference between the timing model param-

eters from a full GW search and those from the best fit is very small. This means we can linearize the timing model and only consider first-order offsets to the corresponding parameters. The deterministic TOA component can then be written as

$$t_{\rm det} = t_{\rm tm} + \mathbf{M}\boldsymbol{\epsilon}, \tag{2.4.2}$$

where the vector $\boldsymbol{\epsilon}$ represents the linear offsets of each timing model parameter from the best fit, and M is known as the design matrix, constructed from the partial derivatives of the TOAs with respect to the different parameters. The matrix is of size $(N_{\rm TOA} \times m)$, with $N_{\rm TOA}$ denoting the number of TOAs and m the number of parameters used in the timing ephemeris, which is $\mathcal{O}(100)$ for a typical pulsar.

If we wish to include other deterministic processes we can add them to the right hand side of Eq. (2.4.2). For example we can add a CW to our signal model by including a term $s(t, \theta)$, using the expression given by Eq. (1.3.10) and parameterized by θ. This can in theory be done for any number of additional sources.

2.4.2 The Noise Model

Next we look at the stochastic component $t_{\rm sto}$, which encompasses the rest of the noise processes we wish to model in our data. It can be split into two main classifications: white noise and red noise. White noise is a random process that has equal power across all frequencies. Red noise processes have more power at lower frequencies than higher ones. Note that a stochastic GW signal, such as the GWB, is included here as it would manifest in the data as an excess of red noise. We will review the main contributors to the stochastic noise model.

There are three main sources of white noise to consider when constructing PTA analyses. First is a parameter called EFAC which acts as a correction factor to the template-

fitting uncertainties on the TOAs stemming from radiometer noise. It is treated as a multiplicative factor directly onto the TOA uncertainties. Other instrumental effects may lead to additional white noise in the data not compensated for by the EFAC model and can be included as an extra term added in quadrature, with a corresponding parameter called EQUAD. Pulsars are observed by multiple different combinations of telescope receivers and backends. For each unique combination per pulsar, we can assign both an EFAC and EQUAD parameter, contributing to an overall delay term we will denote $\mathbf{n}$. The covariance matrix for these two white noise components is

$$\mathbf{N} \equiv \langle n_{i,\alpha} n_{j,\beta} \rangle = W_\alpha^2 \left(\sigma_i^2 + Q_\alpha^2\right) \delta_{ij} \delta_{\alpha\beta}. \tag{2.4.3}$$

The σ_i are the TOA uncertainties, the W and Q represent the EFAC and EQUAD parameters, respectively, and the indices (α, β) denote different combinations of telescope receivers and backends. Note that the matrix $\mathbf{N}$ is diagonal.

There is one additional white noise term that is added labeled ECORR standing for extra correlated white noise. This is noise that is correlated within an observing epoch, but uncorrelated across different epochs. This could originate from, for example, an effect known as pulse phase jitter where across separate observing epochs there can be some residual shape changes in the pulse profile from the finite number of pulses being fit to the template. There are multiple different treatments for ECORR in analysis pipelines. Here we choose to model it using a basis of the number of observing epochs $N_{\rm epoch}$ which adds a component to the TOA vector of $\mathbf{Uj}$. The vector $\mathbf{j}$ is of length $N_{\rm epoch}$ and the matrix U, sometimes called an exploder matrix, is of size $(N_{\rm TOA} \times N_{\rm epoch})$ and contains values of 1 where a TOA aligns with its observing epoch, and zeros elsewhere. We model the ECORR terms as zero-mean Gaussian process, placing a Gaussian prior on the vector $\mathbf{j}$ given some hyperparameters such as the ECORR values. The full delay term is also split

by receiver and backend combinations.

Lastly we consider red noise contributions to the observed TOAs. They can be arranged given one of two designations: red noise that is intrinsic to each pulsar, and red noise that is common across all pulsars. Intrinsic red noise processes could arise from, for example, instabilities in the pulsar's period, spindown, or phase (Shannon & Cordes, 2010), whereas a common red noise process could be a source such as the GWB. We will distinguish the two in a moment, but their treatment in the overall PTA model is roughly the same. We characterize the processes using a Fourier basis with sampling frequency $1/T$, where T is the time span of the data. We extend this basis to some desired number of frequencies N_f. Then the added delay term representing the red-noise process is given by $\mathbf{Fa}$, with F the Fourier design matrix of size $(N_{\mathrm{TOA}} \times 2N_f)$ containing alternating columns of sines and cosines for each frequency, and $\mathbf{a}$ representing the vector of Fourier coefficients. Similar to ECORR, we model the Fourier coefficients as Gaussian processes with some hyperparameters defining the overall signal. The total red-noise covariance matrix $\boldsymbol{\varphi}$ can be written as (Arzoumanian et al., 2016)

$$[\boldsymbol{\varphi}]_{(ai),(bj)} = \rho_{\mathrm{irn},ai}\delta_{ab}\delta_{ij} + \Gamma_{ab}\rho_{\mathrm{gwb},i}\delta_{ij}, \tag{2.4.4}$$

where the (a, b) index different pulsars and (i, j) different frequency bins. The term Γ_{ab} is overlap reduction function, the HD curve for a GWB, and the $\rho_{\mathrm{irn},ai}$ and $\rho_{\mathrm{gwb},i}$ denote the spectrum of the intrinsic red-noise and common-process red-noise signals, respectively. There are many ways we can choose to model the spectrum of both red-noise processes. The standard choice, and the one used in the analyses throughout this dissertation, is to model both processes using power laws, similar to when we derived the PTA response to a stochastic GWB in Sec. 1.3.4. This gives

$$\rho_{\rm irn}(f) = A^2_{\rm irn}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma_{\rm irn}}, \qquad \rho_{\rm gwb}(f) = A^2_{\rm gwb}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma_{\rm gwb}}. \tag{2.4.5}$$

with $f_{\rm yr}$ a reference frequency equal to an inverse year, and the amplitude parameters quoted to that reference frequency. Bringing all of the pieces together, we can write the final form of our timing residuals with all its individual components

$$\delta t = \mathbf{M}\boldsymbol{\epsilon} + \mathbf{U}\mathbf{j} + \mathbf{F}\mathbf{a} + \mathbf{n} + \mathbf{s}. \tag{2.4.6}$$

In Figure 2.3 we plot examples of how the different signal or noise processes can affect measured TOA residuals in a single pulsar. The three plots show, in order from top to bottom, the inclusion of only white noise, the inclusion of both white noise and red noise, and an instance of white noise with the addition of a single CW source. In the bottom two plots we outline the contributions due to the red noise and CW signal, respectively, in red, demonstrating the relative differences in strength among the different noise and signal sources. The white noise is uncorrelated, the red noise is correlated in time, and the CW induces an expected sinusoidal behavior. Of course, in reality all processes be jumbled together in one dataset.

2.4.3 The PTA Likelihood

The final tool we need to start building and running PTA inference pipelines is a likelihood function for the residuals δt given a sample draw from the set of parameters in our model. In this section we will review the key concepts concerning the likelihood's derivation, and for a more complete description see for example Arzoumanian et al. (2016) or Lentati et al. (2013). We start by constructing an estimate of the delay term $\mathbf{n}$ for the Gaussian white noise. This is done through the noise-mitigated timing residuals

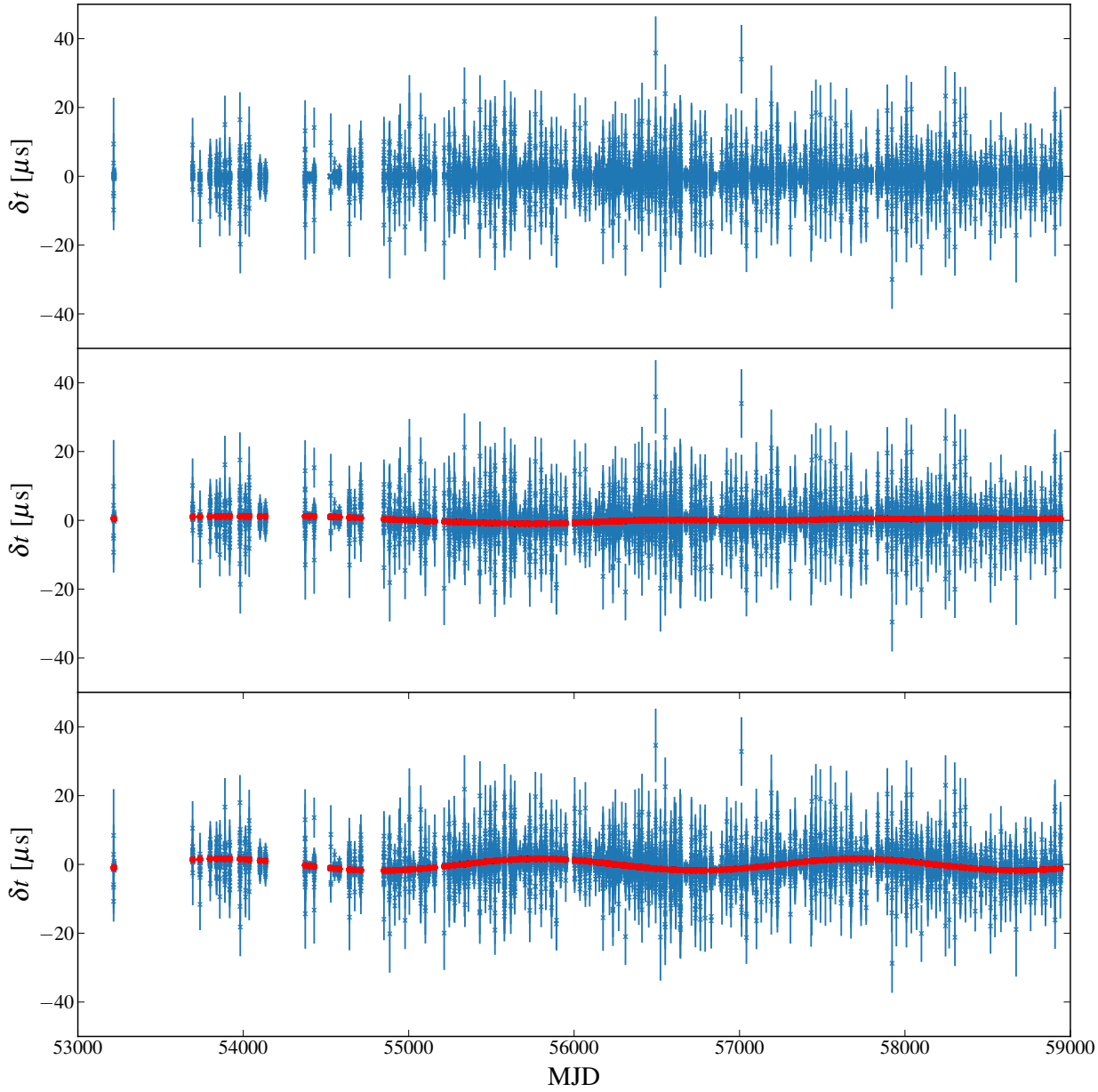

Figure 2.3: Simulated pulsar timing residuals for PSR J1744-1134 under the presence of different classes of noise sources. From top to bottom: the case of a white noise only injection, the case of white noise with an intrinsic red noise injection, and the case of white noise as well as a deterministic CW signal. In the bottom two panels the red dots correspond to the contributions from the red noise injection and CW signal, respectively.

$r = \delta t - \mathbf{M}\boldsymbol{\epsilon} - \mathbf{U}\mathbf{j} - \mathbf{F}\mathbf{a} - \mathbf{s}$, with a corresponding multi-variate Gaussian likelihood

$$P(\delta t|\boldsymbol{\epsilon}, \mathbf{j}, \mathbf{a}, \theta) = \frac{\exp\left(-\frac{1}{2}r^T\mathbf{N}^{-1}r\right)}{\sqrt{\det\left(2\pi\mathbf{N}\right)}}. \tag{2.4.7}$$

where θ is the set of hyperparameters describing the white noise and any additional deterministic signals contained in $\mathbf{s}$. The remaining terms all are modeled as Gaussian processes. For convenience we group the three sets of basis coefficients $\mathbf{b} = [\boldsymbol{\epsilon}\ \ \mathbf{j}\ \ \mathbf{a}]$ and basis matrices $\mathbf{T} = [\mathbf{M}\ \ \mathbf{U}\ \ \mathbf{F}]$. The prior over all of the coefficients is simply

$$P(\mathbf{b}|\theta) = \frac{\exp\left(-\frac{1}{2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}\right)}{\sqrt{\det\left(2\pi\mathbf{B}\right)}}. \tag{2.4.8}$$

The matrix $\mathbf{B} = \mathrm{diag}\{\boldsymbol{\infty}\ \ \boldsymbol{\mathcal{J}}\ \ \boldsymbol{\varphi}\}$ represents the prior covariance matrix for this grouping of Gaussian processes, with the term $\boldsymbol{\infty}$ meaning we placed unconstrained priors on the timing model offset parameters. We are more interested in the hyperparameters of these Gaussian processes than we are of the individual coefficients. It can be shown, as in Lentati et al. (2013) or van Haasteren & Levin (2013), that one can multiply Eq. (2.4.7) by Eq. (2.4.8) and analytically marginalize over $\mathbf{b}$ to write the likelihood function only for θ

$$P(\delta t|\theta) = \frac{\exp\left(-\frac{1}{2}(\delta t - s)^T\mathbf{C}^{-1}(\delta t - s)\right)}{\sqrt{\det\left(2\pi\mathbf{C}\right)}}, \tag{2.4.9}$$

with $\mathbf{C} = \mathbf{N} + \mathbf{T}\mathbf{B}\mathbf{T}^T$. By performing this marginalization, the dimensionality of the posterior is greatly reduced. The total covariance matrix $\mathbf{C}$ is a size $(N_{\mathrm{TOA}} \times N_{\mathrm{TOA}})$ matrix, which be of order $\mathcal{O}(10^3)$ for individual pulsars and $\mathcal{O}(10^6)$ for an entire PTA. We can use the Woodbury matrix identity (Max, 1950) to reduce the size of costly inversions

$$\left(\mathbf{N} + \mathbf{T}\mathbf{B}\mathbf{T}^T\right)^{-1} = \mathbf{N}^{-1} - \mathbf{N}^{-1}\mathbf{T}\boldsymbol{\Sigma}^{-1}\mathbf{T}^T\mathbf{N}^{-1}, \tag{2.4.10}$$

where $\boldsymbol{\Sigma} = \mathbf{B}^{-1} + \mathbf{T}^T\mathbf{N}^{-1}\mathbf{T}$. Recall that the white noise covariance matrix $\mathbf{N}$ is diagonal, so

the dense matrix inversion happens with $\boldsymbol{\Sigma}$, which is of a much smaller size $(N_b \times N_b)$ with N_b the number of overall basis components. This is ultimately the form of the likelihood used across the majority of PTA analyses.

Before concluding the chapter it is worthwhile to look at a complete example of a PTA model and consider its dimensionality and intercorrelated signals, since that forms the primary argument towards developing and using HMC sampling methods. We will consider a joint model containing both a stochastic GWB as well as a CW from a single source. The relevant parameters, their descriptions, and their typical priors are given in Table 2.1. The notation $U[p_{\min}, p_{\max}]$ denotes a uniform prior between $p_{\min}$ and $p_{\max}$. Each pulsar has its own set of power-law amplitude and spectral index parameters $(A_{\rm irn}, \gamma_{\rm irn})$. There are separate common amplitude and spectral index parameters $(A_{\rm gwb}, \gamma_{\rm gwb})$ for the GWB signal. There are also eight additional parameters for the CW source. The white noise hyperparameters are typically held fixed to maximum likelihood estimates obtained from single-pulsar noise-only inference runs. In total for a PTA containing 67 pulsars (the size of the NANOGrav 15-year dataset (Agazie et al., 2023b)), this amounts to $134+2+8 = 144$ free parameters in our model.

The model dimension is very large, and there are many potential intercorrelations between the parameters (e.g., the common-process signal and pulsar-intrinsic red noises, or the common-process signal and a low-frequency CW). Furthermore the noise covariance matrix $\mathbf{C}$ is still quite costly to invert even after the reduction in dimension by using the Woodbury matrix identity. Much work has been done to speed up the overall likelihood calculation, to factorize across pulsars where possible, and to develop rapid refitting techniques of other spectral characterizations (Taylor et al., 2022; Lamb et al., 2023). Regardless, this type of model is precisely the kind for which HMC sampling methods are designed to help improve, and the next two chapters will demonstrate this on multiple

Parameter	Description	Prior
$A_{\rm irn}$	Pulsar-intrinsic red noise amplitude	$U[-18, 11]$
$\gamma_{\rm irn}$	Pulsar-intrinsic red noise spectral index	$U[0, 7]$
$A_{\rm gwb}$	Common-process red noise amplitude	$U[-18, 11]$
$\gamma_{\rm gwb}$	Common-process red noise spectral index	$U[0, 7]$
$\log_{10} h$	Log of CW strain amplitude	$U[-18, 11]$
$\log_{10} \mathcal{M}$	Log of CW source chirp mass	$U[7, 10]$
$\log_{10} f_{\rm GW}$	Log of CW frequency	$U[-9, -6.5]$
$\cos\theta$	Cosine of CW source sky location polar angle	$U[-1, 1]$
ϕ	CW source sky location azimuthal angle	$U[0, 2\pi]$
$\cos\imath$	Cosine of CW source orbital inclination	$U[-1, 1]$
Φ_0	Initial CW orbital phase	$U[0, 2\pi]$
ψ	CW polarization angle	$U[0, \pi]$

Table 2.1: Summary of parameters included in a joint GWB and CW analysis. Included are the standard notations for the parameters, their descriptions, and typical prior ranges.

different PTA models. One final note making the bridge from the PTA likelihood into a Hamiltonian sampling pipeline: we need to be able to compute gradients of Eq. (2.4.9). In practice there are some components of the gradient that can be calculated analytically, but this is not true in general for all possible parameter combinations. Purely analytic gradient code becomes too bloated, and hybrid analytic/numerical differentiation solutions are not computationally feasible. It is best to calculate the gradient fully using autodifferentiation packages, with `JAX` (Bradbury et al., 2018) being the primary open-source example.

CHAPTER 3

Efficient Gravitational Wave Searches with Pulsar Timing Arrays using Hamiltonian Monte Carlo

This chapter originates from:
Efficient gravitational wave searches with pulsar timing arrays using Hamiltonian Monte Carlo
G. E. Freedman, A. D. Johnson, R. van Haasteren, and S. J. Vigeland
Physical Review D, 107, 043013, (2023)

3.1 INTRODUCTION

Pulsar timing arrays (PTAs) (Sazhin, 1978; Detweiler, 1979; Foster & Backer, 1990) seek to detect low-frequency gravitational waves (GWs) by looking for spatial correlations induced in the times of arrival (TOAs) pulses from millisecond pulsars. PTAs are most sensitive in the nanohertz frequency regime ($\sim$1–100 nHz), where the dominant source of GWs is expected to be a stochastic gravitational wave background (GWB) originating from a cosmic population of supermassive black hole binaries (SMBHBs) (Sesana et al., 2004, 2005; Rosado et al., 2015; Burke-Spolaor et al., 2019). The North American Nanohertz Observatory for Gravitational Waves (NANOGrav) (Ransom et al., 2019) has been collecting pulsar TOA data since 2004. NANOGrav, along with the European Pulsar Timing Array (Kramer & Champion, 2013), Parkes Pulsar Timing Array (Manchester et al., 2013), and the Indian Pulsar Timing Array Project (Tarafdar et al., 2022) form the International Pulsar Timing Array (IPTA) (Perera et al., 2019).

Detection of low-frequency GWs provides a valuable tool for studying parts of the dynamical universe not accessible through electromagnetic observations. Constraining the GWB shape and strength can provide useful constraints on properties of the SMBHB population including the black hole–host galaxy scaling relations (Ravi et al., 2015; Sesana,

2013) and the astrophysical environments of SMBHBs emitting GWs (Quinlan, 1996; Sesana et al., 2005; Haiman et al., 2009; van Haasteren et al., 2009; Sampson et al., 2015b). The GWB could also contain contributions from more speculative sources such as primordial GWs from inflation (Grishchuk, 1976; Lasky et al., 2016) and networks of cosmic strings (Siemens et al., 2007; Blanco-Pillado et al., 2014).

GW signals can be extracted as a correlated signal from pulsar timing data only after subtracting the pulsar's timing model and accounting for underlying sources of noise in both the pulsar and observing instruments. These analyses are frequently done using Bayesian techniques (van Haasteren et al., 2009; van Haasteren & Levin, 2013; Lentati et al., 2013; Arzoumanian et al., 2016), which we outline in Sec. 3.2. In order to perform the Bayesian searches, NANOGrav makes use of the parallel-tempering Markov chain Monte Carlo (MCMC) code `PTMCMCSampler` (Ellis & van Haasteren, 2017), which includes a variety of jump proposal schemes such as differential evolution, prior draws, and adaptive Metropolis.

MCMC methods work adequately for a large portion of statistical models, but simple MCMC algorithms such as random-walk Metropolis (Metropolis et al., 1953) or Gibbs sampling (Geman & Geman, 1984) become slow as the size and complexity of the model grow and take considerably longer to converge. Both of the aforementioned methods use random-walk proposals to generate samples and explore the parameter space, which tend to be increasingly inefficient when the target distribution includes correlations among the parameters (Neal, 2011). Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Neal, 2011) removes the requirement to sample the model randomly, and replaces it with a simulation of Hamiltonian dynamics on the distribution itself. This scheme allows samples to be drawn at much further distances from one another, and explores the full parameter space in a more efficient way. For a target distribution of dimension d, the cost of drawing an

independent sample with HMC goes roughly as $\mathcal{O}(d^{5/4})$, compared to $\mathcal{O}(d^2)$ for random-walk Metropolis (Creutz, 1988). The no-u-turn sampler (NUTS) (Hoffman & Gelman, 2011) algorithm provides a basis for performing analysis with HMC without pretuning the sampling.

The HMC algorithm was initially developed for the problem of performing lattice field theory simulations of quantum chromodynamics (Duane et al., 1987). The earliest approach applying HMC to PTA science was in the development of a model-independent method for performing Bayesian analyses on pulsar timing data (Lentati et al., 2013). The technique worked extremely well when applied to the IPTA Mock Data Challenge.[1] When applied to real data, however, the sampling could not fully explore the hierarchical model and became stuck in "Neal's funnel" (Neal, 2003). Applying data-aware coordinate transformations using the Cholesky decomposition helped deal with hierarchical funneling, and consequently there was a successful application of HMC to the targeted problem of outlier excision from PTA datasets (Vallisneri & van Haasteren, 2017). The trade-off was that the additional transformations made sampling the hierarchical likelihood slower than the typical marginalized likelihood that was already used. As a result, HMC was not further explored in this context and has since remained largely underutilized towards the broad array of PTA science.

In this paper, we present a method for performing PTA GW searches using HMC as the underlying sampling algorithm. This represents the first attempt at applying HMC to the marginalized PTA likelihood, where we can avoid the funneling that plagues hierarchical models while still leveraging the benefits of HMC in exploring high-dimensional distributions. We test this method on the NANOGrav 11-year dataset (Arzoumanian et al., 2018a), as well as realistic simulated data with similar red and white noise to the

[1]The first IPTA Mock Data Challenge was developed by Fredrick Jenet, Kejia Lee, and Michael Keith and administered in 2012.

NANOGrav 11-year dataset. We demonstrate that performing a Bayesian GWB search with HMC results in a significant reduction in required sample generation to give equivalent results to current methods.

We also show that the additional gradient calculations necessary for HMC to operate scale roughly the same as the current likelihood evaluation with respect to the number of pulsars in a given dataset. Additionally we demonstrate that when comparing the time to generate independent samples, HMC outperforms traditional MCMC methods for PTA models of varying size in accordance with the expected scaling. This is a necessary consideration as the sizes of PTAs will continue to grow and with that the number of parameters needed to sample over.

This paper is organized as follows. In Sec. 3.2, we describe the methods, signal models, and software used. In Sec. 3.3 we present the results of a GWB search using HMC, and compare the accuracy and efficiency of this method for both real and simulated PTA data. We conclude in Sec. 3.4 and discuss how this method could be utilized for future PTA work.

3.2 METHODOLOGY AND SOFTWARE

In this section, we provide a brief outline of a typical PTA Bayesian GW search. We then give an overview of the HMC and NUTS algorithms, and discuss how to apply these methods to existing PTA work.

3.2.1 PTA Signal Model

We now discuss the PTA likelihood function. Following the outline provided in (Arzoumanian et al., 2016), we start by considering a single pulsar and its timing residual vector $\delta \mathbf{t}$ with length equal to the number of TOAs in our dataset, $N_{\rm TOA}$. This timing residual

data can be decomposed into individual components:

$$\delta \mathbf{t} = M\boldsymbol{\epsilon} + F\mathbf{a} + U\mathbf{j} + \mathbf{n}. \tag{3.2.1}$$

Each term describes a different inaccuracy or source of noise that contributes to the residual data. The term $M\boldsymbol{\epsilon}$ represents inaccuracies stemming from the subtraction of the pulsar's timing model, with M the timing model design matrix, and $\boldsymbol{\epsilon}$ the vector of timing model parameter offsets. The effects due to low-frequency ("red") noise are encoded in the term $F\mathbf{a}$. We choose to define this in a rank-reduced basis where F represents our matrix of basis functions, in this case alternating sine and cosine functions, and $\mathbf{a}$ represents a set of Fourier coefficients. The term $U\mathbf{j}$ describes noise that is completely uncorrelated in time but completely correlated across observations of a similar epoch. The matrix U maps between N_{TOA} residual data and N_{epoch} observation sessions, and $\mathbf{j}$ accounts for the correlated noise in each epoch. The final term, $\mathbf{n}$, includes any other high-frequency ("white") noise that cannot be accounted for in the previous terms, such as radiometer noise.

Previous Bayesian analysis schemes (van Haasteren et al., 2009; van Haasteren & Levin, 2010; van Haasteren et al., 2011; Ellis, 2013; Ellis et al., 2013) have described the white noise with EFAC (constant multiplier to TOA uncertainties) and EQUAD (white noise added in quadrature to EFAC) parameters and employed a power-law model to describe the red noise. The sum of these white noise covariances we describe via a matrix N. The parameters describing $\boldsymbol{\epsilon}$, $\mathbf{a}$, and $\mathbf{j}$ we group as follows:

$$T = \begin{bmatrix} M & F & U \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \mathbf{a} \\ \mathbf{j} \end{bmatrix}. \tag{3.2.2}$$

We place a Gaussian prior on these parameters with covariance:

$$B = \begin{bmatrix} \infty & 0 & 0 \\ 0 & \varphi & 0 \\ 0 & 0 & \mathcal{J} \end{bmatrix}, \tag{3.2.3}$$

where ∞ represents a diagonal matrix of infinities corresponding to unconstrained uniform priors on all timing model parameters. The parameters that describe $\mathcal{J}$ we refer to as ECORR and correspond to the epoch-correlated white noise signals per receiving back end. The matrix φ defines the parameters involving red noise signals, which includes low-frequency noise intrinsic to each pulsar, as well as the stochastic GWB. For this paper, we performed our analysis by modeling the GWB using a fiducial power-law spectrum of the characteristic GW strain h_c and cross-power spectral density S_{ab}:

$$h_c(f) = A_{\rm gw} \left(\frac{f}{f_{yr}}\right)^{\alpha}, \tag{3.2.4}$$

$$S_{ab}(f) = \Gamma_{ab} \frac{A_{\rm gw}^2}{12\pi^2} \left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} f_{\rm yr}^{-3}, \tag{3.2.5}$$

where $\gamma = 3 - 2\alpha$. For a background generated by the GW emission from the evolution of a population of inspiraling SMBHBs in circular orbits, we have $\alpha = -2/3$, which implies $\gamma = 13/3$ (Phinney, 2001). The function Γ_{ab} is called the overlap reduction function (ORF) and describes the average correlations between any two pulsars a and b as a function of their angular separation. For an isotropic, stochastic GWB, this ORF is given by the Hellings-Downs correlation: (Hellings & Downs, 1983)

$$\Gamma_{ab} = \frac{3}{2} x_{ab} \ln x_{ab} - \frac{x_{ab}}{4} + \frac{1}{2}, \tag{3.2.6}$$

where $x_{ab} = (1 - \cos \xi_{ab})/2$ for two pulsars with angular separation ξ_{ab}.

We analytically marginalize over the timing model parameters to reduce the overall dimensionality of our posterior (Lentati et al., 2013; van Haasteren & Vallisneri, 2014) and are left with the form of the likelihood that is used for the analysis in this paper:

$$p(\delta\mathbf{t}|\phi) = \frac{\exp\left(-\frac{1}{2}\delta\mathbf{t}^T C^{-1}\delta\mathbf{t}\right)}{\sqrt{\det 2\pi C}}, \tag{3.2.7}$$

where $C = N + TBT^T$. We define ϕ as the set of all varying parameters in our model. We compute the likelihood and perform Bayesian searches using the NANOGrav package `enterprise` (Ellis et al., 2020).

3.2.2 Hamiltonian Monte Carlo

We now provide a description of the HMC algorithm. In HMC (Duane et al., 1987; Neal, 2011), we start by introducing an auxiliary momentum variable p_i alongside each target parameter q_i. In most implementations, the momenta are chosen to be independent of the q_i and follow a zero-mean Gaussian distribution, with a covariance matrix M that is typically taken to be the identity. The log of the joint density of $\mathbf{p}$ and $\mathbf{q}$ defines our Hamiltonian:

$$H(\mathbf{p},\mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}) = -\mathcal{L}(\mathbf{q}) + \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}, \tag{3.2.8}$$

where $\mathcal{L}(\mathbf{q}) \equiv \log p(\delta\mathbf{t}|\phi)$ is the log of the likelihood function for the distribution of our target parameters $\mathbf{q}$. Analogous to Hamiltonian dynamics, we have a potential energy term $U(\mathbf{q})$ and a kinetic energy term $K(\mathbf{p})$. We then simulate the evolution of this system over time according to Hamilton's equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}. \tag{3.2.9}$$

This can be solved numerically using a symplectic integrator such as a "leapfrog"

method, which for an integration step size ε uses an update scheme:

$$\mathbf{p}^{t+\varepsilon/2} = \mathbf{p}^{t} + \left(\frac{\varepsilon}{2}\right) \nabla_{\mathbf{q}} \mathcal{L}(\mathbf{q}^{t}), \tag{3.2.10a}$$

$$\mathbf{q}^{t+\varepsilon} = \mathbf{q}^{t} + \varepsilon \mathbf{p}^{t+\varepsilon/2}, \tag{3.2.10b}$$

$$\mathbf{p}^{t+\varepsilon} = \mathbf{p}^{t+\varepsilon/2} + \left(\frac{\varepsilon}{2}\right) \nabla_{\mathbf{q}} \mathcal{L}(\mathbf{q}^{t+\varepsilon}), \tag{3.2.10c}$$

where superscripts denote the time at which the particular quantity is evaluated. The standard method for producing a chain of samples using HMC then proceeds as follows: We first resample our momenta distribution. Then for a set number of leapfrog steps L, we use Eq. (3.2.10) to evolve our system through time and propose some final position and momentum vectors $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{p}}$. This proposal is accepted or rejected according to a similar condition to the Metropolis algorithm (Metropolis et al., 1953) using the ratio of the Hamiltonian evaluated at the initial and final points.

Mapping the path of the leapfrog integrator leads to a useful sanity check of HMC: trajectory divergences. These divergences occur when the trajectory taken via Hamiltonian simulation departs from the true trajectory, and risk biasing estimates or reducing HMC to random-walk behavior (Betancourt, 2016). By tracking the trajectories and alerting the user of large divergences, HMC offers another diagnostic to detect unsuitably parametrized models that is not possible with Metropolis-Hastings (MH) MCMC methods.

There are limitations to HMC and the models under which it can be used properly. Due to its origins in Hamiltonian dynamics, HMC can only operate in continuous state spaces and contains no internal recourse to deal with discrete variables. In such cases, the discrete variables can be handled with separate algorithms such as Gibbs sampling (Geman & Geman, 1984). HMC also requires that the log density of the target distribution is

differentiable almost everywhere with respect to the model parameters, with the exception coming at points of probability 0 (Neal, 2011). Additionally, HMC struggles when there is strong multimodality in the target distribution due to the modes being separated by regions of very low probability (Sminchisescu & Welling, 2011). The PTA models used in this paper satisfy the above conditions, and HMC remains a valid choice of underlying sampling algorithm.

3.2.3 No-u-turn Sampler

The performance of the HMC algorithm is particularly sensitive to two user-defined parameters: the number of leapfrog steps L and integration step size ε, defined in the above section. If these parameters are not properly tuned, the algorithm may waste computation time or begin to exhibit unwanted random walk behavior and in some cases may not even be ergodic (Neal, 2011). In general, tuning these parameters appropriately would require multiple preliminary runs.

The no-u-turn samlper (NUTS; Hoffman & Gelman, 2011) offers an extension to the HMC algorithm that dynamically tunes the number of leapfrog steps L. NUTS uses a recursive doubling algorithm, similar to the one outlined in (Neal, 2003), to determine when the generated proposal trajectory begins to double back on itself, or make a "U turn". The algorithm builds a binary tree, simulating Hamiltonian dynamics forwards and backwards randomly in time for 2^j steps, with j the height of the full tree. If we define $\mathbf{q}^+$, $\mathbf{p}^+$ and $\mathbf{q}^-$, $\mathbf{p}^-$ as the position-momenta pairs of the left- and right-most nodes of the bottom subtree, then the stopping condition for NUTS can be written as:

$$\left(\mathbf{q}^+ - \mathbf{q}^-\right) \cdot \mathbf{p}^- < 0 \quad \text{or} \quad \left(\mathbf{q}^+ - \mathbf{q}^-\right) \cdot \mathbf{p}^+ < 0. \tag{3.2.11}$$

The above procedure adaptively tunes the parameter L for each iteration in the chain.

The step size parameter ε in NUTS is set using the method of stochastic optimization with varying adaptation (Andrieu & Thoms, 2008). In particular, Hoffman and Gelman utilize the primal-dual averaging algorithm proposed by (Nesterov, 2009). With L and ε automatically tuned, NUTS can be run without any human intervention.

3.2.4 Coordinate Transformations and Software

Previous approaches to pulsar timing analyses with HMC utilized a hierarchical PTA likelihood. Initially these methods did not include coordinate transformations on the data, and as a result became stuck with hierarchical funneling. This funneling originates from the fact that within hierarchical models, random variables are very highly correlated when the data are sparse (Betancourt & Girolami, 2013). One can reduce the correlations between the random variables, and hence the funneling, by adopting a noncentered reparametrization of the data (Papaspiliopoulos et al., 2007). In regards to the hierarchical PTA likelihood, such a reparametrization using the Cholesky decomposition allowed HMC sampling to proceed but at the cost of slowing down the likelihood.

In this paper we are focused entirely on the marginalized PTA likelihood and can therefore leave behind the coordinate transformations designed for hierarchical models. We do employ a set of transformations designed to improve the performance of the NUTS algorithm. First we perform an interval transform, moving all parameters with bounded priors from their interval $[a, b]$ to the whole real line. We then whiten the data using Cholesky whitening to move to a set of transformed variables whose covariance matrix is the identity. This is accomplished through the Hessian calculated around the maximum *a posteriori* parameter vector. Neither transformation considerably alters the likelihood computation speed.

When determining the speed and efficiency of the HMC and NUTS pipeline, one must

depend almost entirely on the ability to calculate gradients of the likelihood and do so as quickly as possible. Numerical derivatives are comparably easier to write but slow in practice and prone to errors from approximations. By-hand analytic derivatives are fast but difficult to write into concise code for all but the simplest of models. An excellent solution for arbitrary likelihood functions and their gradients is the package `JAX` (Bradbury et al., 2018), which leverages both automatic differentiation and just-in-time compilation to efficiently differentiate native Python code and turn an otherwise slow gradient function into incredibly fast machine executables. The use of this technique is rather new, with `JAX` only recently becoming a mature code base, and consequently this marks the first time `JAX` and automatic differentiation have been utilized for HMC sampling of PTA data.

Summarizing the software used for the analyses to follow in this paper, the signal models and likelihood used in our analyses come from NANOGrav's flagship PTA analysis suite `enterprise` (Ellis et al., 2020). We utilize the automatic differentiation capabilities in `JAX` (Bradbury et al., 2018) to calculate the likelihood derivatives required for HMC to operate. We perform two coordinate transformations on our data to better interface with the NUTS algorithm. Lastly, for the sampling we use a custom-built NUTS code that is freely and openly available in `piccard`.[2] The combination of these three codes leads to an end-to-end pipeline for performing PTA analyses with HMC sampling.

3.3 RESULTS

In this section, we study the HMC sampling method both in its ability to accurately perform Bayesian searches for a stochastic GWB using PTA data, as well the efficiency of such a method when compared against the existing techniques employed by NANOGrav.

[2] https://github.com/vhaasteren/piccard

The GWB model that is analyzed in this paper arises from a PTA consisting of data from 45 pulsars. The parameters encompassing the signal model closely mimic those outlined in Sec. 3.2.1. We fix white noise parameters to their maximum likelihood values as obtained from individual pulsar noise runs. We model pulsar-intrinsic red noise with a power-law power spectral density (PSD) containing two search parameters $\log_{10} A_{\rm red} \in U[-18, -11]$ and $\gamma_{\rm red} \in U[0, 7]$. We model the GWB as a power-law PSD process that is common amongst all the pulsars. The corresponding parameters are an amplitude with log-uniform prior $A_{\rm CP} \in U[-18, -12]$ and a spectral index $\gamma_{\rm CP}$ that we fix to $13/3$. We do not include spatial correlations in our GWB model. This results in a total of $2N_{\rm psr} + 1$ varying parameters in the model.

We also generate a set of simulated PTA datasets using `libstempo` (Vallisneri, 2020). We inject both per-pulsar white and red noise parameters at their maximum likelihood values. The injected values again originate from individual pulsar noise runs, where all parameters for a given pulsar are allowed to vary. Again we include a common process signal representing the GWB with both a fixed amplitude and spectral index at $\log_{10} A_{\rm CP} = -14.7$ and $\gamma_{\rm CP} = 13/3$, and do not include interpulsar spatial correlations. We repeat the above procedure for 100 realizations of the GWB which results in a collection of 100 realistic simulated PTA datasets. When analyzing the simulated data, we use a similar signal model to the one described above but this time allow the common-process spectral index to vary as $\gamma_{\rm CP} \in U[0, 7]$.

Runs conducted with the MH MCMC algorithm use the `PTMCMCSampler` (Ellis & van Haasteren, 2017) code. The sampler is set up in similar fashion to the NANOGrav 11-year GWB search (Arzoumanian et al., 2018b). We include adaptive Metropolis and differential evolution jump proposals. For all varying parameters present in the model, we also add prior draw jump proposals. We do not utilize parallel-tempering in this work.

3.3.1 NANOGrav 11-year Data Comparison

We perform a stochastic GWB search with both the HMC and MH MCMC algorithms on the NANOGrav 11-year dataset (Arzoumanian et al., 2018a). This dataset encompasses the timing data for 45 millisecond pulsars. Figure 3.1 shows the posterior distributions for the background amplitude $A_{\rm CP}$ calculated using both Monte Carlo methods. We calculate 95% upper limits on $A_{\rm CP}$ and estimate uncertainties with bootstrap methods (Efron, 1979). The HMC algorithm produces results that are consistent with the base MH MCMC search, with corresponding 95% upper limits $A_{\rm CP,\,HMC} < 1.72(4) \times 10^{-15}$ and $A_{\rm CP,\,MH\,MCMC} < 1.74(3) \times 10^{-15}$.

The MH MCMC sampling routine was run for a total number of samples $M_{\rm MH\,MCMC} = 1,000,000$, whereas the HMC routine was run for $M_{\rm HMC} = 8,000$. The wall time for the MH run was approximately 4 hours, compared to just under 4 hours for the HMC run. Both sets of chains are checked for convergence using the Gelman-Rubin R-hat convergence test (Gelman & Rubin, 1992). It is worth reinforcing that the benefit of generating fewer samples is partially outweighed by the increased computational cost of proposing a new HMC sample. We explore the scaling of sample generation time in Sec. 3.3.3.

We also perform a direct comparison to the upper limit calculated in the NANOGrav 11-year GWB search (Arzoumanian et al., 2018b). In order to do such a comparison, we alter our signal model slightly to match that of the 11-year analysis and adjust the common-process amplitude from a log uniform to a uniform prior $A_{\rm CP} \in [10^{-18}, 10^{-12}]$. Performing this analysis with the HMC pipeline, again with $M = 8,000$ samples, recovers a 95% upper limit of $A_{\rm CP,\,HMC} < 1.64(3) \times 10^{-15}$. This is in relative agreement with the result in (Arzoumanian et al., 2018b) of $A_{\rm CP} < 1.61(2) \times 10^{-15}$ for a similar model with identical Jet Propulsion Laboratory (JPL) ephemeride DE436.

We further compare the efficiency of HMC sampling by looking at the autocorrelation

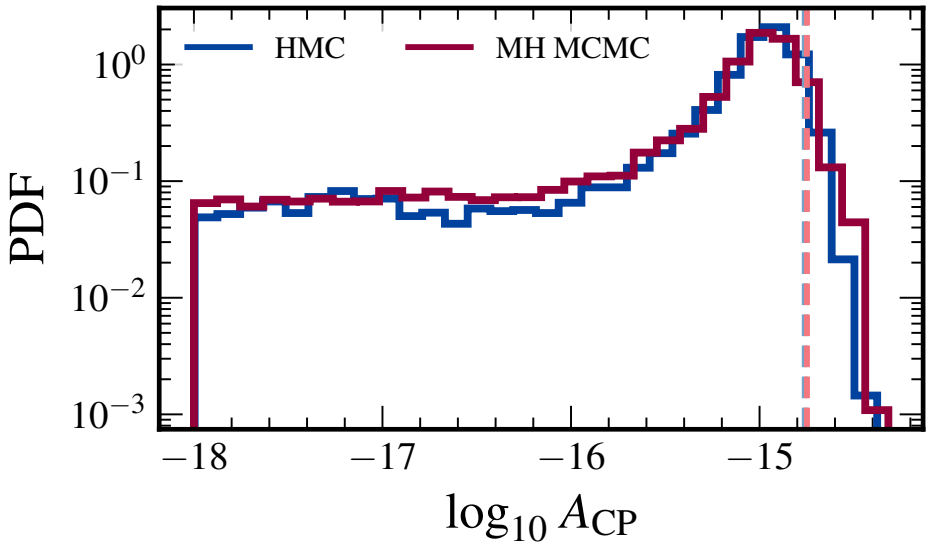

Figure 3.1: Posterior probability distributions for the amplitude $\log_{10} A_{\rm CP}$ of a common-process signal run using either MH MCMC or HMC as the primary sampling method, computed using the NANOGrav 11-year dataset. The common-process amplitude parameter is set with a log-uniform prior, the common-process spectral index is fixed at 13/3, and no spatial correlations are included. Vertical lines represent 95% upper limits calculated for posteriors generated using HMC [blue; $A_{\rm CP,\,HMC} < 1.72(4) \times 10^{-15}$] and MH MCMC [red; $A_{\rm CP,\,HMC} < 1.74(3) \times 10^{-15}$], though the two lines will be difficult to individually resolve due to the similarity in upper limits. We conclude that the two procedures produce consistent posteriors when applied to identical models.

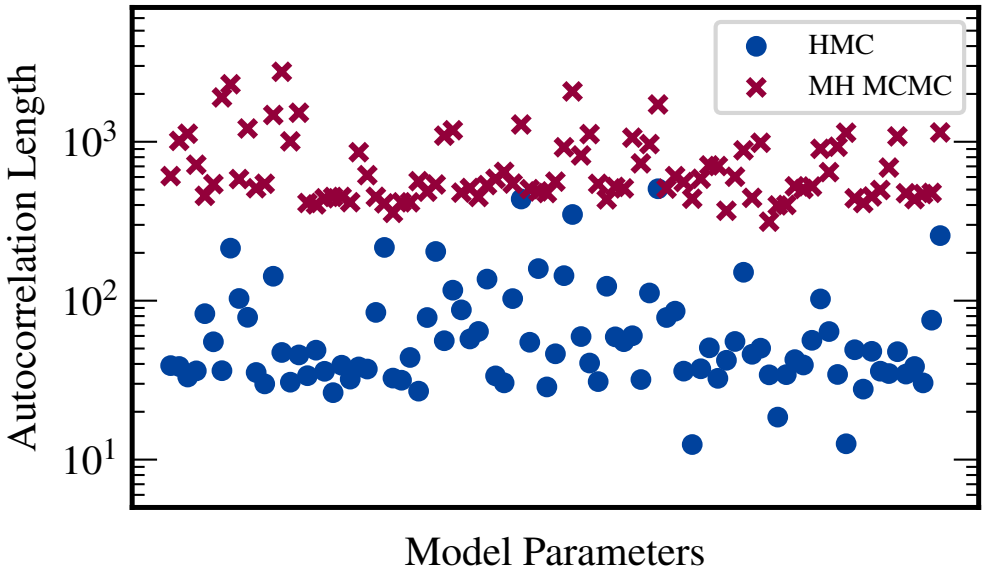

Figure 3.2: Autocorrelation lengths for 91 parameters ($2N_{\rm psr}$ individual pulsar red-noise parameters and a common process signal parametrized with an amplitude $A_{\rm CP}$ and spectral index $\gamma_{\rm CP} = 13/3$) present in a standard GWB model. The autocorrelation lengths are calculated from two sets of chains generated from sampling this model: one sampled with HMC (blue) and one with MH MCMC (red). Each mark represents the approximate number of steps one must jump through that particular parameter's chain to reach an independent sample.

lengths of the two sets of chains, measuring how far one must jump through the chain to find the next statistically significant sample. The autocorrelation lengths are calculated per parameter in the model. This was calculated for each set of chains generated with the two Monte Carlo sampling methods, and the results are shown in Fig. 3.2. We find that the HMC chains have autocorrelation lengths between 1 and 2 orders of magnitude smaller than those of identical parameters in the MH MCMC chains. This behavior is expected, as the HMC algorithm is designed to take larger, more-informed steps to avoid random walk-like behavior and produce a higher ratio of independent samples.

3.3.2 Simulated Data and Parameter Recovery

We also aim to test that the HMC algorithm behaves similarly to the standard MH MCMC technique when considering statistical coverage of a standard PTA model. To determine the capability of the sampling methods to accurately recover injected parameters, we consider 100 simulated PTA datasets and seek to verify if in $p\%$ of the realizations the injected parameter values fall within the $p\%$ credible region of the posteriors. We run standard Bayesian searches on all realizations using both sampling methods.

The results of the parameter recovery test described above are summarized in Fig. 3.3, with a particular focus on the two parameters describing the GWB. The HMC sampler recovers the injected GWB parameter values with the same consistency as the traditional analysis. Neither method recovers the injected parameters exactly, and therefore no line in Fig. 3.3 falls directly on the vertical line at $x = 0$. This is due to an inherent model mismatch present when simulating data with `libstempo` and recovering the posteriors separately with `enterprise`. The simulated GWB is generated with more frequencies than is searched over during the analysis, leading to a natural bias in recovery.[3]

3.3.3 Scaling of Gradient Computation Speed

The time per HMC sample generation is dominated by the time to calculate the gradient of the log likelihood necessary for leapfrog integration. The evaluation time for the base likelihood calculation present in `enterprise` is calculated by averaging the evaluation time for 50 calls of the log likelihood function. We first use a PTA with only a single pulsar, and repeat the above step adding one additional pulsar at a time up to $N_{\rm psr} = 45$. This produces an idea of how the base likelihood evaluation time, and by extension the MCMC sample generation time, scales with the number of pulsars present in a PTA (Fig. 3.4:

[3]For further details, see documentation for GWB simulation in the `toasim` module of `libstempo`.

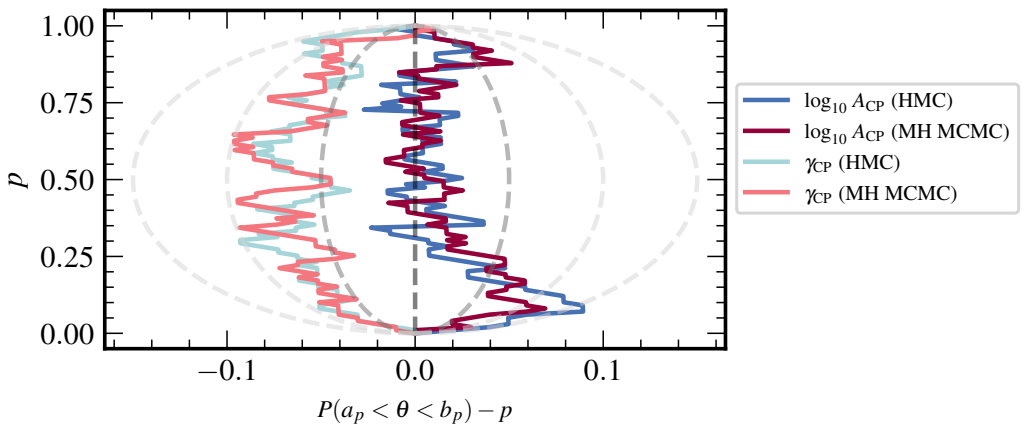

Figure 3.3: $p - p$ comparison of GWB parameter recovery for both the HMC and MH MCMC sampling methods operating on simulated PTA data. The x axis shows the difference between the fraction of realizations with which the injected values fall within the p% credible region of the posteriors and the p% credible region on the y axis. The vertical dark gray line at $x = 0$ represents a perfect recovery of the injected parameter values. The light gray lines represent 1σ, 2σ, and 3σ deviations.

dashed red line).

In order to accurately scale the computation time necessary to draw a sample with NUTS, we must account for the dynamic tuning of the HMC hyperparameter L and note that we likely require multiple evaluations of the log likelihood and gradient to generate a sample. First we consider the evaluation time of the log likelihood and gradient function compiled with `JAX`, and scale per pulsar following the same procedure defined above (Fig. 3.4: solid blue line). We then take the 45 separate PTA objects and run standard GWB analyses, with models defined in Sec. 3.3, through the HMC pipeline for $M = 10,000$ samples. The height j of the NUTS binary tree defines a total of $L = 2^j + 1$ gradient evaluations per new sample. By averaging this over the full run, we can approximate an $L_{\rm eff}$ and more accurately scale the time per HMC sample generation (Fig. 3.4: cyan triangles).

Finally, we look at the time to generate independent samples in our chain and how

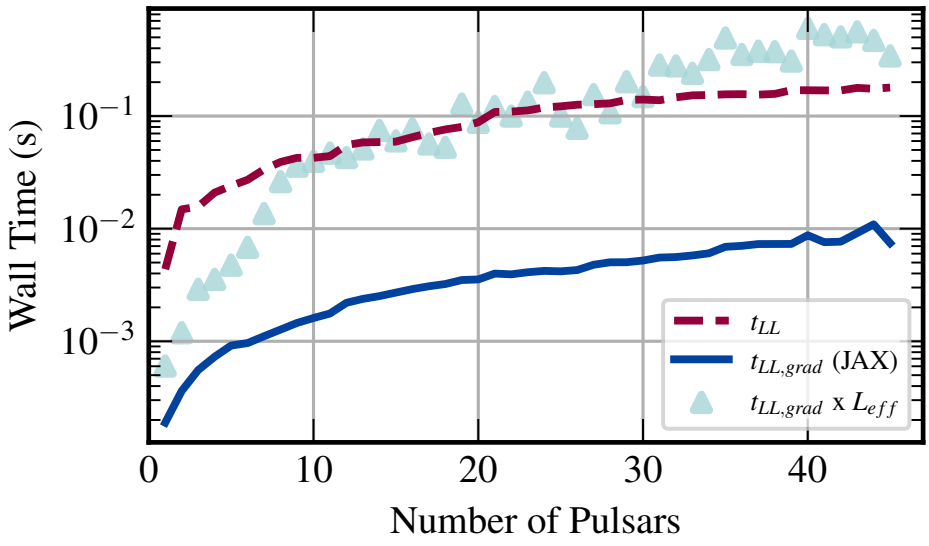

Figure 3.4: Wall time for calculating implementations of both the log of the PTA likelihood as well as its gradient, scaled by the number of pulsars present in a given model. The red dashed line represents the log-likelihood evaluation as present in the standard PTA analysis suite `enterprise`. The solid blue line shows the evaluation of the log likelihood and gradient function after being precompiled with `JAX`. The cyan triangles denote the evaluation times present in the blue line multiplied by a value L_{eff} representing the effective number of gradient evaluations required to generate a new HMC sample.

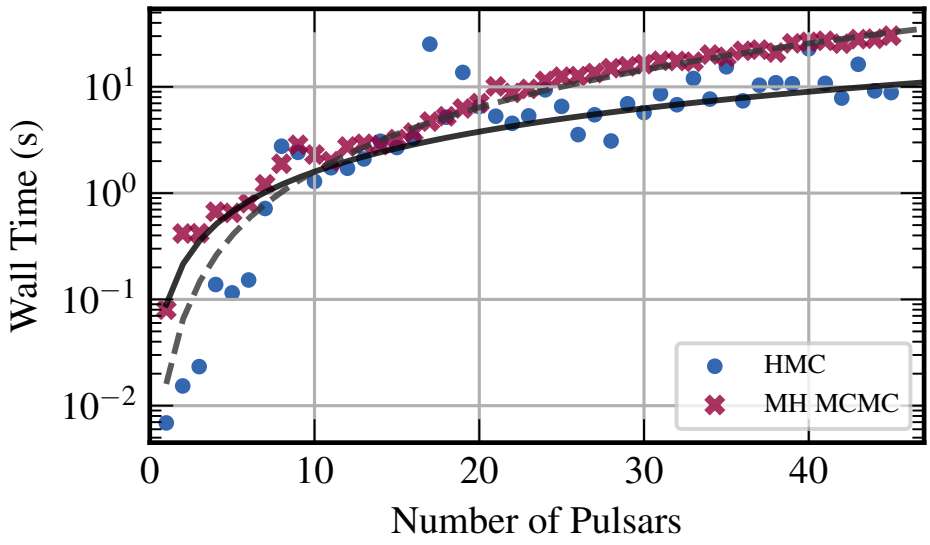

Figure 3.5: Wall time to produce an independent sample in Markov chains generated using HMC and MH MCMC methods, scaled by the number of pulsars N_{psr} present in the model. The total number of parameters in a given model is $d = 2N_{\text{psr}} + 1$. The solid black represents the expected scaling for HMC of $\mathcal{O}(d^{5/4})$. The dashed gray line denotes the expected scaling for MH MCMC of $\mathcal{O}(d^2)$.

it scales with increasing PTA size. This is ultimately the most important metric for testing the efficiency of HMC as independent samples and thinned Markov chains are what inevitably drive the statistical inferences made on the data. Independent samples in this context are defined here as samples that are separated by one autocorrelation length.

We take the 45 PTA objects of increasing N_{psr} and generate Markov chains with HMC and MH MCMC of size $M = 8,000$ and $M = 1,000,000$, respectively. Taking the median autocorrelation length of each chain and the base sampling speed calculated previously, we create a scaling of the wall time for both sampling methods in making independent samples for PTA models of increasing size. The results are summarized in Fig. 3.5. It shows that in the long run HMC will outperform MH MCMC techniques in making statistically relevant samples, despite the likely increase in upfront computational cost.

3.4 SUMMARY AND OUTLOOK

In this paper, we have implemented an efficient method for sampling the high dimensional distributions present in PTA GW Bayesian searches using the HMC algorithm. This method leverages a hybrid technique comprised of parts from both traditional stochastic Monte Carlo schemes as well as deterministic sampling methods derived from Hamiltonian dynamics. We show that utilizing HMC results in a reduction of approximately 2 orders of magnitude in the number of samples drawn to produce equivalent results to the existing Bayesian searches performed on PTA datasets.

The efficiency of this technique is largely defined by the speed at which derivatives of the log likelihood can be computed for the purpose of simulating Hamiltonian dynamics. We have shown that the current implementation of this calculation scales similarly to the present log-likelihood calculation with respect to the number of pulsars in a dataset, and improves upon traditional MCMC methods when comparing the production of independent samples of the distribution. This improvement in performance scaling is paramount because PTAs will continuously grow and add more pulsars to their data collection. The 11-year dataset featured in this paper contains 45 pulsars. Future NANOGrav datasets will have ≥ 60 pulsars and future IPTA datasets may contain close to ~ 100 pulsars. Increasing the data volume will further strain our computational capabilities to perform large parameter GW searches. HMC provides a way of resolving these limitations in a way that is more favorable to future PTA analyses.

It is worth emphasizing that the $\mathcal{O}(d^{5/4})$ scaling of HMC is not just with respect to the number of pulsars, but with respect to the total number of parameters. We analyzed a model with $2N_{\rm psr} + 1$ parameters (with the GWB spectral index held fixed), but this represents only one of many different approaches for GW searches in PTAs. For example, one can parametrize the GWB with a free spectrum model, increasing its number of pa-

rameters from 2 to 30. Likewise, one can parametrize the individual pulsar red noise in a similar fashion, increasing the parameter count from 2 to 30 *per pulsar*. The favorable scaling of HMC opens the door for more flexible models that are currently prohibitive with current MH MCMC runs.

Currently we have only applied the HMC algorithm to the problem of sampling a stochastic GWB model. PTAs are also sensitive to certain deterministic GW signals, and work towards tailoring this method to such searches is under development. This technique is particularly promising for searches for GWs from individual SMBHBs because of the large number of parameters necessary to describe the GW signal ($2N_{\rm psr} + 8$ for a circular binary, more if the source is eccentric). In general, this technique can be adapted to the full suite of PTA searches, provided the underlying models adhere to the limitations outlined in Sec. 3.2.2. The ultimate goal is a general purpose pipeline for performing any such PTA analysis that leverages the benefits of the HMC algorithm towards exploring complicated, high-dimensional models.

CHAPTER 4

Joint Searches for Continuous Gravitational Waves and a Gravitational Wave Background with Hamiltonian Sampling

This chapter originates from:
An efficient pipeline for joint gravitational wave searches from individual binaries and a gravitational wave background with Hamiltonian sampling
G. E. Freedman and S. J. Vigeland
Physical Review D, 110, 063038, (2024)

4.1 INTRODUCTION

The first evidence of a low-frequency stochastic gravitational wave (GW) signal reported (Agazie et al., 2023a; EPTA Collaboration et al., 2023; Reardon et al., 2023a; Xu et al., 2023) by the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) (Ransom et al., 2019), European Pulsar Timing Array (Kramer & Champion, 2013), Indian Pulsar Timing Array (Tarafdar et al., 2022), Parkes Pulsar Timing Array (Manchester et al., 2013), and Chinese Pulsar Timing Array (Xu et al., 2023) has opened a new chapter in the field of GW astrophysics. Pulsar timing array (PTA) (Sazhin, 1978; Detweiler, 1979; Foster & Backer, 1990) collaborations search for nHz frequency GWs by analyzing the the times-of-arrival (TOAs) of radio pulses emitted by millisecond pulsars. By regularly observing such pulsars over a decades-long timespan PTAs can reach the sensitivity necessary to probe the nHz band. The recently identified stochastic GW signal displayed, to varying levels of significance, the expected Hellings-Downs (HD) (Hellings & Downs, 1983) spatial correlation signature between pulsars that is indicative of the signal being a gravitational wave background (GWB).

One possible source describing the nHz GWB is the collective signal from the population of supermassive black hole binaries (SMBHBs) present in the observable universe (Sesana et al., 2005). All massive galaxies hold a supermassive black hole, typically of mass $10^6 - 10^{10} M_\odot$, at their centers (Kormendy & Ho, 2013). Galactic merger events consequently lead to the formation of SMBHB systems. When the component black holes reach inspiral phase, the emission of GWs becomes the dominant force behind the system's evolution. To date there have been no confirmed observations of SMBHBs at subparsec separations. With the discovery of a GWB signal, a next major step for PTA science is to search for particularly loud individual binaries that may be detected amongst the stochastic ensemble within the next decade (Rosado et al., 2015; Mingarelli et al., 2017; Kelley et al., 2018; Bécsy et al., 2022b). Measurements of GWs from individual sources, colloquially referred to as continuous waves (CWs) due to their minimal frequency evolution, would provide useful constraints on the astrophysical environments of SMBHBs (Quinlan, 1996; Haiman et al., 2009) and could be coupled with electromagnetic observations to study galactic evolution and further multimessenger astrophysical research (eg. (Charisi et al., 2022)).

Single source searches prove inherently more computationally taxing than a comparable GWB analysis. Modeling GWs from an individual SMBHB adds to the already large PTA parameter space, and such parameters bring covariances amongst themselves as well as with other red-noise processes present in the data. The computational cost additionally compounds with increased data volume, which includes longer observation span and new pulsars added to the array. This complication is particularly apparent for efforts at combining datasets from multiple PTAs, yielding highly sensitive yet computationally overwhelming data products. Multiple techniques at exploring the CW parameter space have been developed (Corbin & Cornish, 2010; Lee et al., 2011; Ellis, 2013; Taylor et al.,

2014; Bécsy et al., 2022a), and recent improvements have led to a 100-fold speed up of the full analysis through the use of a tailored likelihood calculation (Bécsy et al., 2022a).

In this paper we detail an additional procedure for achieving efficient CW searches through a Monte Carlo routine established through sample proposals based in the gradient of the model likelihood. This utilizes a Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Neal, 2011) sampler to replace the random-walk nature of traditional MCMC methods with a simulation of Hamiltonian dynamics on the target probability distribution. The algorithm concentrates on drawing subsequent samples at much further distances in the multidimensional parameter space, trading pure speedups of the likelihood calculation for higher sample acceptance rates and an efficient exploration of the distribution. Within the realm of PTA science, HMC was first utilized in the development of a model-independent approach to Bayesian inference with PTA data (Lentati et al., 2013), and soon after to the task of outlier removal in single-pulsar data. In a previous paper (Freedman et al., 2023), we demonstrated the effectiveness of using HMC to perform Bayesian GWB searches with the full marginalized PTA likelihood. Here we extend the methods and previous results to allow for joint inference of individual binary sources and common-process signals.

This paper is outlined as follows. In Sec. 4.2 we review the signal model of a single binary and describe the current Bayesian formalism for searching for such sources in the context of PTA data. We describe the Hamiltonian Monte Carlo sampling procedure and introduce a new pipeline, predicated on this algorithm, for performing the analyses in a more efficient manner. In Sec. 4.3 we validate this pipeline against a suite of simulated PTA datasets. We assess the efficiency of this new analysis prescription on the NANOGrav 12.5-year dataset in Sec. 4.4. Lastly, in Sec. 4.5 we summarize and discuss opportunities for future development of this work.

4.2 METHODOLOGY AND SOFTWARE

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

4.2.1 PTA Likelihood

First we discuss pulsar timing data and the structure of the PTA likelihood. Pulsar observational data exists in the form of pulse times-of-arrival (TOAs). After subtracting from each pulsar's TOAs a timing model comprised of parameters such as proper motion, parallax, spin period, spin period derivative, and other orbital parameters, we are left with timing residuals $\delta\mathbf{t}$ that we can characterize as a linear combination of noise sources and GW signals:

$$\delta\mathbf{t} = M\boldsymbol{\epsilon} + \mathbf{n}_{\text{RN}} + \mathbf{n}_{\text{CRN}} + \mathbf{n}_{\text{WN}} + \mathbf{s}\,. \tag{4.2.1}$$

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

$$h_c(f) = A_{\rm gw}\left(\frac{f}{f_{yr}}\right)^{\alpha}, \tag{4.2.2}$$

$$S_{ab}(f) = \Gamma_{ab}\frac{A^2_{\rm gw}}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} f_{\rm yr}^{-3}, \tag{4.2.3}$$

where the spectral index $\gamma = 3 - 2\alpha$. In the case where the common-process signal represents a background generated by the GW emission from a population of inspiraling SMBHBs in circular orbits, we have $\alpha = -2/3$ and $\gamma = 13/3$ (Phinney, 2001). The function Γ_{ab}, called the overlap reduction function (ORF), defines the average correlations between a set of two pulsars a and b based on their relative angular separation. For common uncorrelated red-noise (CURN) processes the ORF is equal to 1. For an isotropic, stochastic GWB it is given by the Hellings-Downs correlation function (Hellings & Downs, 1983):

$$\Gamma_{ab} = \frac{3}{2}x_{ab}\ln x_{ab} - \frac{x_{ab}}{4} + \frac{1}{2}, \tag{4.2.4}$$

with $x_{ab} = (1-\cos\xi_{ab})/2$ for an angular separation ξ_{ab} between two pulsars. Following the common-process noise signals is a term $\mathbf{n}_{\rm WN}$ encoding all high-frequency "white" noise sources present in the data, including constant multiplicative correction factors to TOA uncertainties (EFAC), additional noise added in quadrature (EQUAD), and observational epoch-correlated noise (ECORR) that is uncorrelated across separate epochs. Lastly the vector s represents the component of the timing residuals caused by additional deterministic signals. Here we treat s as the signal induced by an individual binary, and is outlined in more detail in Sec. 4.2.2.

Finally, we construct the form of the PTA likelihood for use in our Bayesian inference pipelines. First we dramatically reduce the dimensionality of our posterior by marginalizing over the timing model parameters (Lentati et al., 2013; van Haasteren & Vallisneri,

2014). Then by constructing the total PTA covariance matrix $C = N + TBT^T$, with N the white noise covariance matrix, T the design matrix for the timing model, red noise, and ECORR signals, and B the prior covariance matrix for those three sets of parameters, we can state the multivariate Gaussian likelihood function used for the analyses in this paper:

$$L\left(\delta\mathbf{t}|\mathbf{q}\right) = \frac{\exp\left(-\frac{1}{2}\left(\delta\mathbf{t}-\mathbf{s}\right)^T C^{-1}\left(\delta\mathbf{t}-\mathbf{s}\right)\right)}{\sqrt{\det 2\pi C}}\,, \tag{4.2.5}$$

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

4.2.2 CW Signal

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

$$s(t,\hat{\Omega}) = F^{+}(\hat{\Omega})\Delta s_{+}(t,\hat{\Omega}) + F^{\times}(\hat{\Omega})\Delta s_{\times}(t,\hat{\Omega})\,, \tag{4.2.6}$$

with the scripts $\{+\times\}$ denoting the plus and cross polarization modes, respectively, the two tensor polarizations allowed by general relativity, and $\hat{\Omega}$ is a unit vector pointing from the GW source to the solar system barycenter. The functions F^{+} and $F^{\times}$ represent the antenna pattern functions that describe the response of a given pulsar to the emitting source, and are composed of the binary sky location polar and azimuthal angles θ and ϕ, respectively, and GW polarization angle ψ (for a complete description, see (Arzoumanian et al., 2023). The terms $\Delta s_{+,\times}(t)$ account for the fact that the Earth and pulsar see the induced GW signal at different times in the binary evolution, and therefore define the difference between the "pulsar-term" and "earth-term":

$$\Delta s_{+,\times}(t,\hat{\Omega}) = s_{+,\times}(t_p) - s_{+,\times}(t)\,, \tag{4.2.7}$$

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

$$t_p = t - L\left(1 + \hat{\Omega}\cdot\hat{u}\right)\,, \tag{4.2.8}$$

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

$$s_E(t,\hat{\Omega}) = F^{+}(\hat{\Omega})s_{+}(t) + F^{\times}(\hat{\Omega})s_{\times}(t)\,. \tag{4.2.9}$$

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

$$s_{+}(t) = -\frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}\sin 2\Phi(t)\left(1+\cos^2\imath\right)\,, \tag{4.2.10}$$

$$s_{\times}(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}2\cos 2\Phi(t)\cos\imath\,. \tag{4.2.11}$$

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

$$\omega(t) = \omega_0 \left[1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}(t - t_0)\right]^{-3/8}, \tag{4.2.12}$$

$$\Phi(t) = \Phi_0 + \frac{1}{32}\mathcal{M}^{-5/3}\left[\omega_0^{-5/3} - \omega(t)^{-5/3}\right]. \tag{4.2.13}$$

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

$$h_0 = \frac{2\mathcal{M}^{5/3}\left(\pi f_{\rm GW}\right)^{2/3}}{d_L}, \tag{4.2.14}$$

with the GW frequency $f_{\rm GW}$ related to the initial angular frequency ω_0 by $\omega_0 = \pi f_{\rm GW}$. We note that Eq. 4.2.14 shows a degeneracy between h_0, $\mathcal{M}$, $f_{\rm GW}$, and d_L, allowing us to choose three of those four quantities when constructing our complete parameter vector. In practice we typically exclude the luminosity distance in favor of h_0, $\mathcal{M}$, and $f_{\rm GW}$. The full earth-term CW source is therefore completely parameterized by $(\theta, \phi, \imath, \psi, \Phi_0, h_0, \mathcal{M}_c, f_{\rm GW})$.

In order to further elucidate the complications in sampling joint CW and common-process models, we compute Eq. 4.2.5 for an individual binary and particular noise realization. We plot the earth-term only likelihood surface in Fig. 4.1 as a function of the CW sky location parameters. The surface displays highly nontrivial structure and demonstrates some of the difficulties in efficiently sampling the full parameter space. There are numerous local extrema where a random-walk MCMC sampler is liable to get trapped and be unable to fully explore the full posterior. This highlights the need for more sophisticated sampling routines and corresponding pipelines.

4.2.3 Hamiltonian Monte Carlo Sampling

The HMC algorithm (Duane et al., 1987; Neal, 2011), an extension of the traditional Metropolis-Hastings technique (Metropolis et al., 1953), tackles the problem of sampling high-dimensional

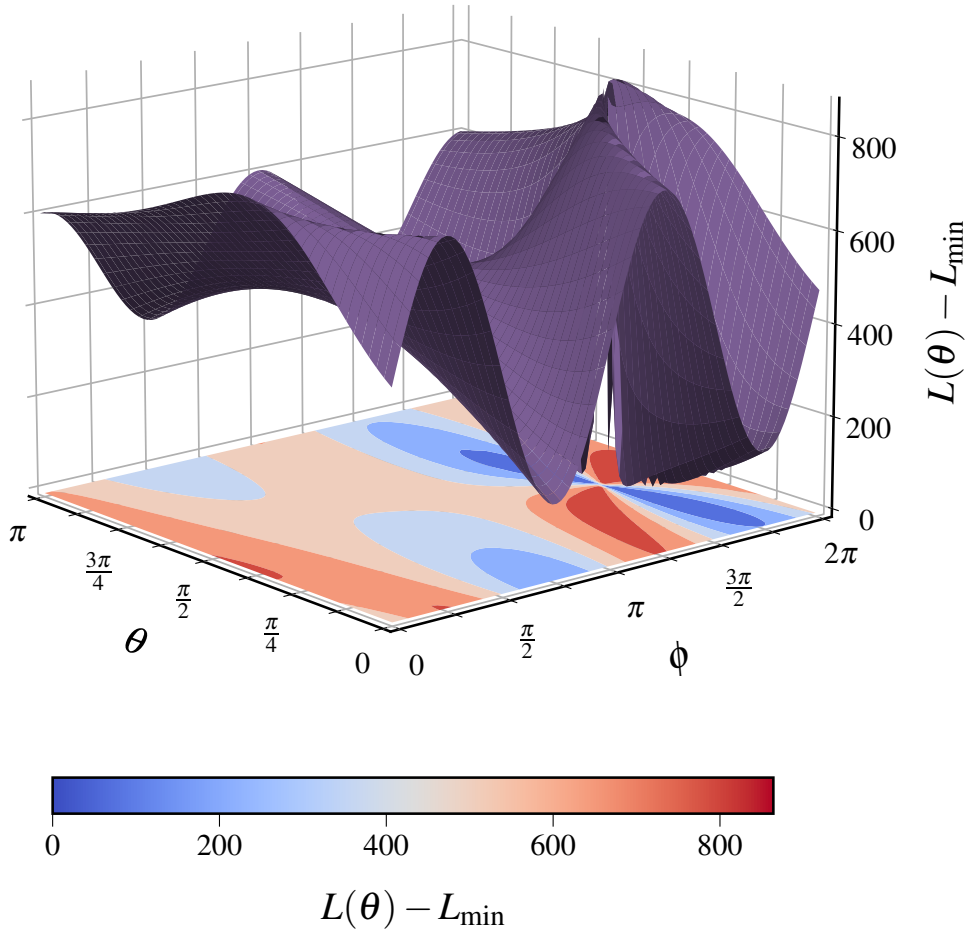

Figure 4.1: Likelihood surface for an earth-term only CW signal as a function of sky position. The x and y axes represent θ and ϕ for the source, respectively. The z axis shows the PTA log-likelihood function evaluated at a particular (θ, ϕ), then subtracting off the minimum log-likelihood value for the grid. On the plane $z = 0$ we plot a 2D colormap contour of the surface. We see that the contours of the likelihood surface have many sharp peaks and valleys, indicating difficult regions of parameter space to sample over.

and covariant state spaces by using Hamiltonian dynamics to generate proposal states that are distant relative to each other in the parameter space. Compared to a standard Gaussian proposal scheme, this reduces the overall correlation in the Markov chain while maintaining a high sample acceptance rate. It proceeds by first introducing an auxiliary momentum vector $\mathbf{p}$ alongside the model parameters $\mathbf{q}$. The Hamiltonian to be simulated is the log of the joint density of $\mathbf{p}$ and $\mathbf{q}$, which can be separated into a potential energy term $U(\mathbf{q})$ and kinetic energy term $K(\mathbf{p})$:

$$H(\mathbf{p}, \mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}) = -\mathcal{L}(\mathbf{q}) + \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}\,, \tag{4.2.15}$$

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

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}}, \qquad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}}\,. \tag{4.2.16}$$

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

The relative performance of the HMC algorithm is then determined by two factors: the computational cost of calculating the gradient of the log likelihood function for the target distribution, and the proper tuning of two user-defined parameters: the number of steps L and integration step size ϵ. The first factor is driven entirely by the complexity of the model in question, and whether or not the log likelihood gradient can be computed exactly or requires numerical differentiation. The second factor can be resolved by automatically tuning the two extra parameters through the use of a No-U-Turn Sampler

(Hoffman & Gelman, 2011). This sampler uses a recursive doubling algorithm to build a binary tree of sample proposals, simulating Hamiltonian dynamics either forwards or backwards in time at random for 2^j iterations with j denoting the height of the tree. The process continues until the position-momenta pairs $\mathbf{q}^+$, $\mathbf{p}^+$ and $\mathbf{q}^-$, $\mathbf{p}^-$ of the left- and rightmost leafs of the tree satisfy the condition:

$$\left(\mathbf{q}^+ - \mathbf{q}^-\right) \cdot \mathbf{p}^- < 0 \quad \text{or} \quad \left(\mathbf{q}^+ - \mathbf{q}^-\right) \cdot \mathbf{p}^+ < 0\,. \tag{4.2.17}$$

In effect this monitors the trajectory of proposals and stops when it begins to double back on itself, or make a "U turn". At this point a sample is chosen at random from the tree and accepted or rejected according to the Metropolis algorithm (Metropolis et al., 1953).

4.2.4 Software

Our new code, freely and publicly available on GitHub under the package `etudes` [1], includes an analysis suite capable of performing HMC sampling with PTAs. Although this paper focuses on joint searches for a CW signal and common red-noise process, the modularity of the code allows for the addition of a wide array of other GW sources of interest, such as multiple binaries (Babak & Sesana, 2012; Bécsy & Cornish, 2020), GW memory (Agazie et al., 2024a), eccentric binaries (Taylor et al., 2016; Susobhanan et al., 2020; Susobhanan, 2023; Agazie et al., 2024b), or advanced pulsar noise modeling (Reardon et al., 2023b; Larsen et al., 2024). It can also be natively run on GPUs, drastically dropping the runtime of CW analyses to timescales of hours for simulated data and days for production data. The code is under further active development to accommodate other searches of interest.

While there are renewed efforts towards utilizing hierarchical modeling for PTAs (van

[1] https://github.com/gabefreedman/etudes

Haasteren, 2024), This work solely uses the marginalized PTA likelihood, meaning that we need not apply coordinate transformations, such as a decentered reparameterization, designed to deal with Markov chain mixing rates and other sampling issues commonly associated with hierarchical funneling. Instead all analyses here, and the default setup for our pipeline, use a single change of coordinates known as an interval transform. This maps all model parameters from their default prior ranges $q \in [a, b]$ to the real line $q' \in (-\infty, \infty)$ via:

$$q' = \log\left(\frac{q-a}{b-q}\right), \tag{4.2.18}$$

$$q = \frac{(b-a)\exp(q')}{1+\exp(q')}, \tag{4.2.19}$$

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

Sampling with HMC necessarily requires taking derivatives of the model likelihood. We accomplish this by writing the PTA likelihood and its components computations entirely with `JAX` (Bradbury et al., 2018), allowing us to use automatic differentiation to calculate gradients. To do so we decouple the the entirety of the PTA computation from NANOGrav's analysis suite `enterprise` (Ellis et al., 2020), though we do make use of the code's data structures for holding per-pulsar TOAs, residuals, and other timing model information. We utilize the implementation of the NUTS algorithm present in the `blackjax` (Cabezas et al., 2024) package. All simulated PTA datasets are created using `libstempo` (Vallisneri, 2020).

4.3 SIMULATED DATA STUDY

First in order to gauge the accuracy of our pipeline and demonstrate its consistency in parameter estimation we created and analyzed a collection of simulated datasets. All datasets comprise identical TOAs, uncertainties, and timing model solutions to the NANOGrav 12.5-year dataset (Alam et al., 2021). This constitutes 45 pulsars in total, all with an observational baseline of at least 3 years.

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

The NANOGrav 12.5-year dataset contained a CURN process with a Bayes factor in excess of 10,000 relative to a model with only intrinsic pulsar noise (Arzoumanian et al., 2020). Therefore, for the most accurate prescription of a realistic PTA dataset, we also include a similar process in all of our simulations. The most recent dataset reported evidence for this process containing HD correlations, though we do not consider that in this study. We inject a CURN signal characterized by an amplitude $A_{\rm CURN} = 2 \times 10^{-15}$ and spectral index $\gamma_{\rm CURN} = 4.33$, in line with the expected power and shape of the spectrum.

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

the case of the high-frequency dataset, we include a binary emitting GWs at $f_{\rm GW} = 60$ nHz with an SNR of 9.3. In both cases the SNR is calculated as:

$$\mathrm{SNR} = \sqrt{(s|s)} = \sqrt{s^T C^{-1} s}\,, \tag{4.3.1}$$

where s is the template waveform and C is the same noise covariance matrix present in Eq. 4.2.5. This can also be considered the expected SNR that is independent from any particular noise realization. The dataset without any CW injection allows us to verify the ability of our methods to place upper limits on source properties in the absence of a detection. For the purposes of validating our pipeline, we create 100 simulated datasets with both the 6 nHz and 60 nHz injection properties. The high- and low-frequency source properties remain fixed across their respective simulations. This allows us to test our methods across numerous noise realizations.

Next we outline the basic procedure for setting up our models before performing Bayesian inference through our HMC pipeline. Rather than simultaneously search over the hundreds of white-noise parameters, we fix them to their maximum likelihood values used in creating the datasets, a commonplace procedure in production-level PTA analyses. We model the pulsar intrinsic red-noise with a power-law power spectral density (PSD) defined by an amplitude $\log_{10} A_{\rm red} \in U[-18, -11]$ and spectral index $\gamma_{\rm red} \in U[0, 7]$. Additionally we search over the two parameters characterizing the CURN process, using priors of $A_{\rm CURN} \in U[-18, -12]$ and $\gamma_{\rm CURN} \in U[0, 7]$. When modeling the CW signal all parameters are given uniform priors. In the case of upper limit analyses, the prior on $\log_{10} h_0$ is shifted from uniform in log space $\log_{10} h_0 \in U[-18, -12]$ to uniform in linear space $\log_{10} h_0 \in U[10^{-18}, 10^{-12}]$. The coordinate space outlined by these priors is later transformed via the procedure outlined in Sec. 4.2.4 prior to beginning the inference.

Lastly we benchmark the speed and efficiency of both the `etudes` pipeline and com-

parable run with `enterprise` through a pilot inference run on one of the simulated 6 nHz injection datasets. The full joint CURN and CW search here constitutes 100 free parameters ($2N_{\rm psr}$ intrinsic red-noise parameters for 45 pulsars, 2 parameters for the CURN, and 8 describing the CW signal model). For the traditional MCMC pipeline with `enterprise` the average likelihood evaluation time is 200 ms on a 12-core Intel(R) Xeon(R) E5-2680 v3 processor. Using the same CPU, the average likelihood and gradient evaluation times with `etudes` is 170 ms and 3.7 s, respectively, and on an NVIDIA Tesla A100 GPU they are 17 ms and 1.6 s, respectively. All of the above results are then scaled by the average autocorrelation lengths of the corresponding MC chains to calculate the timescales of statistically independent sample generation. This gives an estimate of 450 s to get an independent sample with `enterprise` compared to 52 s for runs on a CPU and 25 s on a GPU for the HMC pipeline. Overall, Hamiltonian sampling provides an increase of roughly an order of magnitude in computational efficiency.

4.3.1 Low-frequency (6 nHz) Signal

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

source's inclination, polarization, and initial phase. Together, all of the above allows to fully classify our injected signal.

Taking the resulting chains from our analyses, we plot both the one- and two-dimensional posterior distributions for all eight binary parameters in Fig. 4.2. All parameters have their true injected values lying within their respective posteriors. Both the GW frequency and amplitude distributions are tightly constrained. The posterior for the chirp mass remains entirely unconstrained as we expect for earth-term only searches and sources with slow frequency evolution. The sky location of the source is very well localized to its true value. The initial phase and polarization angles display a set of multimodal posteriors, which we can efficiently sample but are unable to break the multimodality.

4.3.2 High-frequency (60 nHz) Signal

The long-term prospects of CW detection play crucial role in multi-messenger analyses and astrophysical interpretation of SMBHB populations and sources, and it is important that we have the ability to do accurate parameter estimation on possible binary candidates. With this in mind, we analyzed a signal with a GW frequency of $f_{\rm GW} = 60$ nHz, chosen to closely mimic that of the potential SMBHB candidate 3C 66B (Jenet et al., 2004; Iguchi et al., 2010; Agazie et al., 2024b). The remaining parameters describing the source properties and sky location are $(\theta, \phi, \imath, \psi, \Phi_0, \mathcal{M}_c, d_L) = (2\pi/3, 3\pi/2, 3\pi/4, \pi/3, 3\pi/2, 10^9 M_\odot, 91.1\text{Mpc})$. At frequencies this high the CURN is very weak and therefore we do not have to worry with covariances between the common-process and binary signals.

In Fig. 4.3 we plot the posterior distributions for the eight binary parameters for this model, similar to Fig. 4.2. Again we find that we are able to efficiently sample the entire CW parameter space alongside both a CURN process as well as all intrinsic pulsar noise. We see the same structure in the nearly all of our posteriors: the source GW frequency, GW

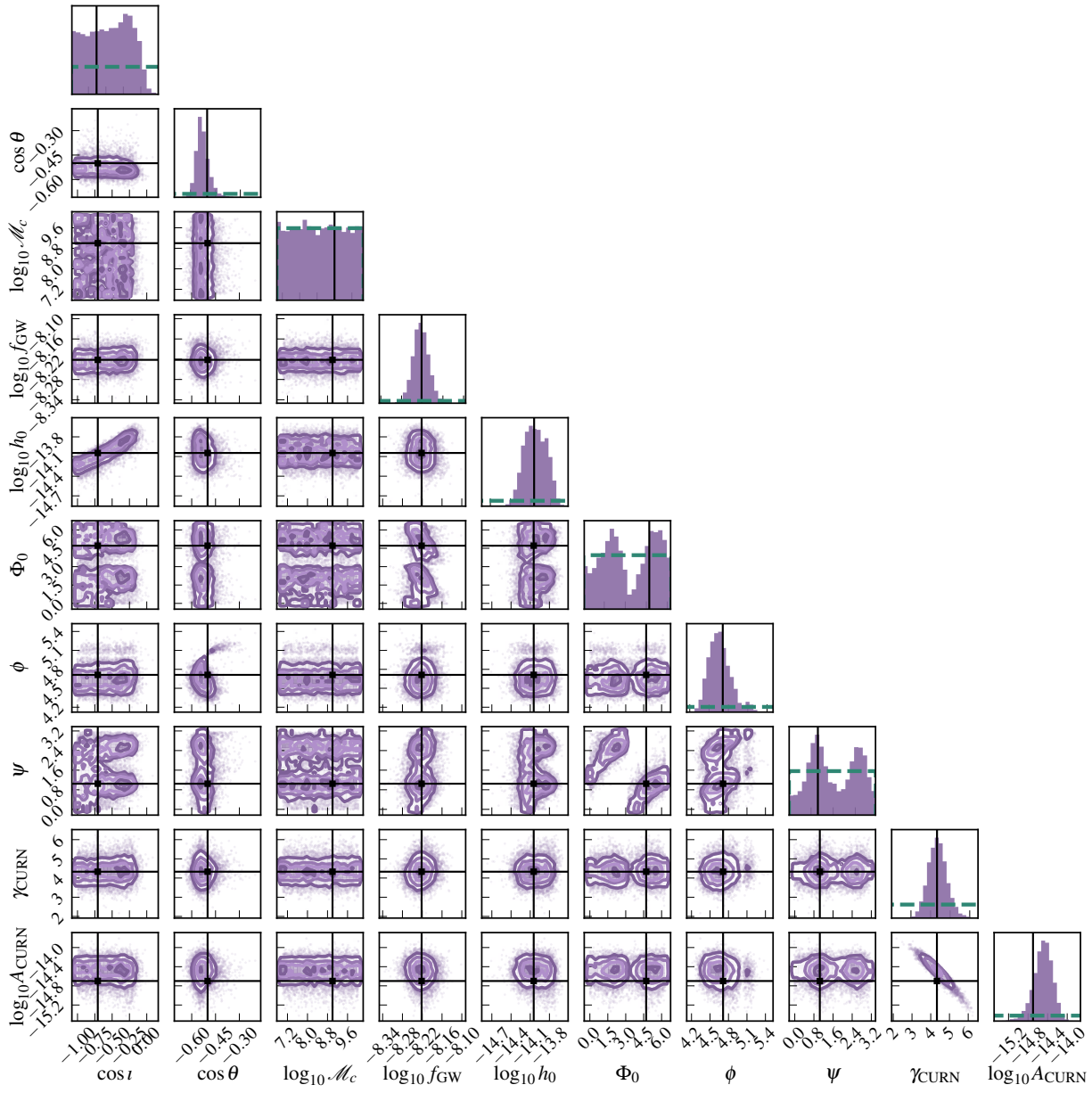

Figure 4.2: 1D and 2D posterior distributions for the eight parameters describing a SMBHB signal emitting GWs at $f_{\mathrm{GW}} = 6$ nHz at an SNR of 10.8. The true values of the injected parameters are shown as solid black lines, and the priors are plotted on the 1D histograms as horizontal, green dashed lines. All true values fall within their posteriors, with the sky location, GW frequency, and GW strain parameters being tightly constrained. This demonstrates the capability of the HMC pipeline in accurate parameter estimation for full CW searches.

amplitude, and sky location are very tightly constrained, and the multimodal structure in the polarization angle and initial phase persist. Most importantly, all injected values once again fall squarely within their 1D posteriors. One notable difference is the emerging constraint on the binary chirp mass. High-mass binaries emitting at this frequency should show significant evolution over the 12.5-year observing window of our simulated datasets. Consequently we find across the 100 realizations of the data that we can place an upper limit on the binary chirp mass. In 30 of the realizations, the chirp mass posterior was less constrained than what the frequency evolution would predict.

4.3.3 Parameter Estimation Consistency

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

The results of this analysis, called a $p - p$ plot, are summarized in Fig. 4.4. We plot lines for CW sky location parameters, $\log_{10} h_0$, $\log_{10} f_{\rm GW}$, and the CURN amplitude and spectral index. The dotted gray lines represent 1σ, 2σ, and 3σ confidence intervals. All parameters fall largely within the 3σ boundary indicating an unbiased recovery of the injected values. The chirp mass, being entirely unconstrained across all realizations due to the minimal evolution of the particular signal, was left out off this figure. The cosine of the binary inclination was also largely unconstrained across all realizations and was likewise excluded.

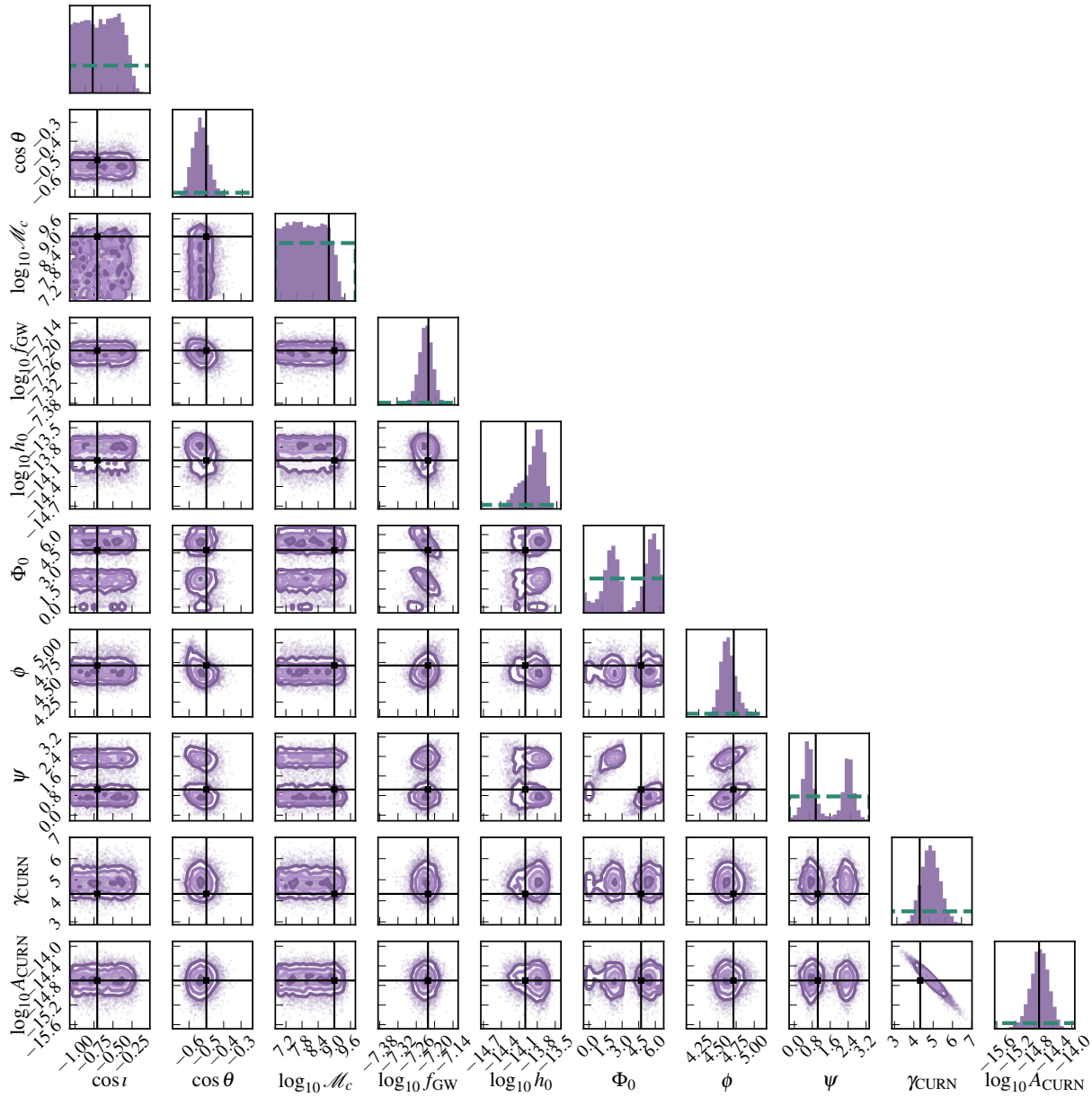

Figure 4.3: 1D and 2D posterior distributions for the eight parameters describing a SMBHB signal emitting GWs at $f_{\rm GW} = 60$ nHz at an SNR of 9.3. The true values of the injected parameters are shown as solid black lines, and the priors are plotted on the 1D histograms as horizontal, green dashed lines. Similar to the low-frequency injection analysis, all true values fall within their posteriors, with parameters such as the sky location, GW frequency, and GW strain parameters being tightly constrained. The binary chirp mass posterior now features an upper limit excluding sources that would have undergone significant frequency evolution over the data timespan.

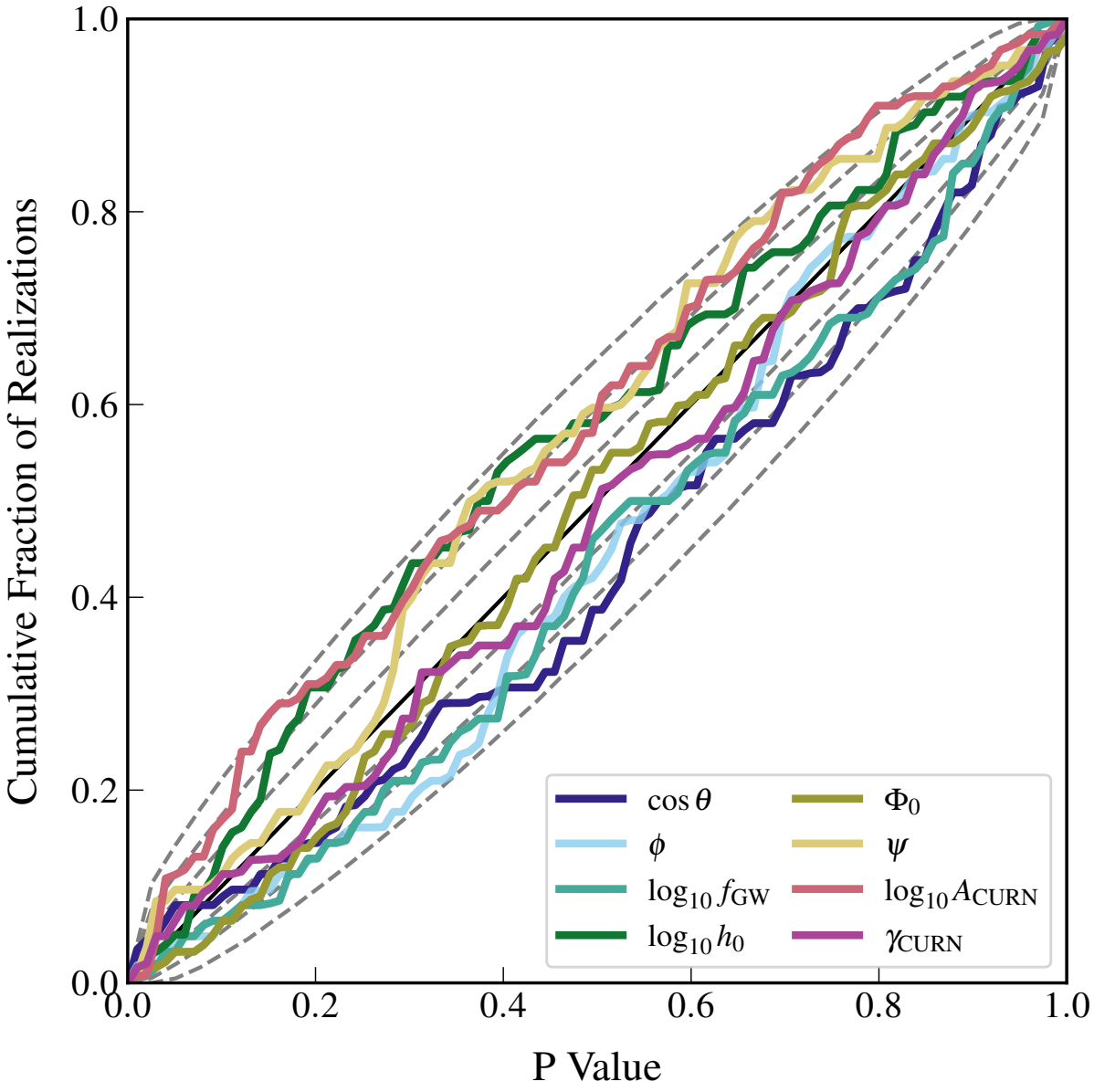

Figure 4.4: $p - p$ plot displaying recovery of injected parameters across 100 simulated PTA datasets. All datasets contain a CURN process and a 6nHz CW injection. Plotted are six lines corresponding to the CW sky location parameters, log strain, log frequency, and CURN amplitude and spectral index. The solid black line along the diagonal represents the line of perfect recovery. Dotted gray lines represent 1σ, 2σ, and 3σ confidence intervals. All plotted parameters lie within these boundaries indicating no significant bias in parameter recovery.

4.4 ANALYSIS OF REAL PTA DATA

Ultimately we want to validate our methods against real data and published results. We use the full NANOGrav 12.5-year dataset (Alam et al., 2021) to benchmark our analysis, and focus on the particular challenge of creating sensitivity sky maps. Given the anisotropic distribution across the sky of the pulsars in our array, it is important to quantify how our observing limits change in different areas. The sky maps typically describe, for a given GW frequency, the 95% upper limits on h_0 as a function of sky location. The typical strategy for generating the plots is to bin the sky into 768 separate pixels and run an MC analysis on each individual partition. This dense pixelation is due in part to our inability to get similar number of MCMC samples across the full parameter space in an all-sky search.

We analyze a CW model including a CURN process for sky locations bounded by $\theta \in [\pi/2, 3\pi/4]$, $\phi \in [3\pi/2, 2\pi]$. The bounds were chosen so as to include the most sensitive sky location from the NANOGrav 12.5-year CW analysis (Arzoumanian et al., 2023), at an RA of $19^{\mathrm{h}}07^{\mathrm{m}}30^{\mathrm{s}}$ and a Dec of $30'00''$. This range of parameter space corresponds to 72 distinct pixels, and therefore typically 72 independent analyses, in the resolution of the sky map from the NANOGrav 12.5-year CW paper. The CW frequency is held fixed at $f_{\mathrm{GW}} = 7.65 \times 10^9$ Hz, the most sensitive frequency in the NANOGrav 12.5-year dataset.

Our strategy is to run one single chain with HMC sampling and leverage the pipeline's efficiency to fully explore across the broader sky range, allowing us to compute a series of GW strain upper limits as a function of sky location and populate the sky map in post-processing. We run one single analysis for $M = 80,000$ samples, after which we break up our chains into sky location bins consistent with the full 768-pixel map. With all autocorrelation lengths of order $\mathcal{O}(1)$, after thinning this results in between $700 - 1,200$ independent samples per reduced sky pixel.

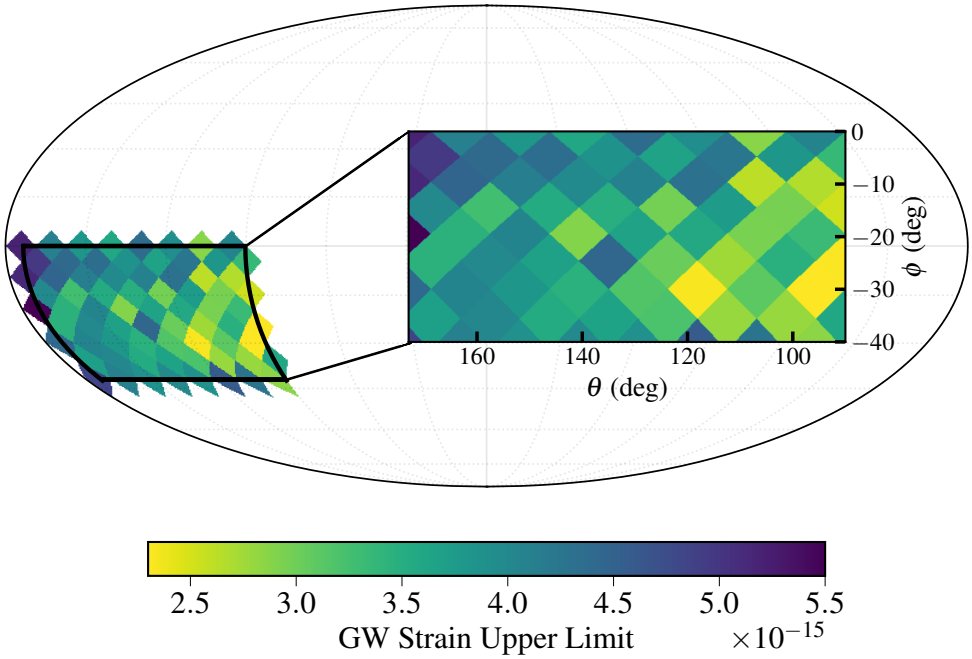

Figure 4.5: Map displaying CW strain 95% upper limits for a range of sky location parameters bounded by $\theta \in [\pi/2, 3\pi/4]$, $\phi \in [3\pi/2, 2\pi]$. The data are taken from a single chain run with an HMC pipeline and pixelated to match the resolution of the analogous map for the 12.5-year data set. The analysis is run for $f_{\rm GW} = 7.65 \times 10^9$ Hz, the most sensitive frequency searched. Pixel to pixel uncertainties range between $1.03 \times 10^{-16} < \sigma_{h_0} < 1.81 \times 10^{-15}$.

In Fig. 4.5, we plot the results of our reconstructed sky map. The strain upper limit at the most sensitive sky location is $h_0 < (2.15 \pm 0.30) \times 10^{-15}$. Its coordinates exactly match that of the most sensitive region from the full NANOGrav 12.5-year analysis, which reported a strain upper limit for that pixel of $h_0 < (2.66 \pm 0.15) \times 10^{-15}$ (Arzoumanian et al., 2023). We find the upper limit at the least sensitive sky location to be $h_0 < (5.45 \pm 0.36) \times 10^{-15}$. Unlike the corresponding analysis and map in the NANOGrav 12.5-year CW paper, we marginalize over the amplitude and spectral index of the CURN process instead of fixing the signal parameters to their maximum likelihood values. We also only search over the Earth term of our CW signal. Therefore we do not expect to find perfect agreement between the two results when comparing on a pixel-to-pixel basis.

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

4.5 DISCUSSION

In this paper we have presented an end-to-end pipeline for performing efficient Bayesian searches of the high dimensional and complicated parameter spaces for joint CW and common red-noise process signal analyses with PTA data. Our code employs HMC sampling to conduct accurate parameter estimation. We demonstrated the performance of

this sampling routine through numerous tests across both simulated and real PTA data. We showed that by using HMC sampling we can effectively do parameter estimation for both high- and low-frequency CW signals. The methods are robust towards conducting these analyses while simultaneously marginalizing over a common-process signal and can accurately recover both GW signals.

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

A significant long-term advantage of this pipeline is its modularity and ability to adapt to a wide range of signal modeling choices. The code is not designed solely for the task of CW searches and can develop and grow into a general purpose analysis suite similar to the current analysis suite `enterprise`. For example, it can modified to run on models considering only a GWB signal, for which previous efforts have already shown HMC sampling to be increasingly useful (Freedman et al., 2023). Further development can also add the possibility of more sophisticated pulsar noise models or additional deterministic sources of interest in the PTA band. The future of PTA GW analyses is in part defined by its potential computational pitfalls: an ever-increasing data span, noise modeling of growing complexity, and the goal of combined international datasets. These methods will prove a valuable tool alongside the range of computational developments in the PTA community towards addressing these issues before they arise, and keeping our analyses tractable to the future.

CHAPTER 5

Exploring the Problem of Extreme Mass Ratio Inspiral Data Analysis with the Laser Interferometer Space Antenna

5.1 INTRODUCTION

Data analysis algorithms are functionally universal. Clustering algorithms like K-means find uses spanning from classifying galaxy spectra (Sánchez Almeida et al., 2010) to filtering spam emails. Supervised learning algorithms like Support Vector Regression models can both estimate photometric redshifts (Wadadekar, 2005) and also analyze MRIs for abnormalities. Any new model or novel method will likely have far-reaching relevance across the sciences and industry. Therefore it stands to reason that developments in data analysis techniques in one area of GW physics will find practical applications across the full range of GW frequencies and their corresponding experiments.

With this in mind, we will now shift focus away from pulsar timing array experiments and the low-frequency GW spectrum and instead explore the middle-frequency ($\sim 10^{-4} - 10^{-1}$ Hz) range lying squarely between the sensitivities of PTAs and ground-based detectors. Exploring this portion of frequency space requires both the complete suppression of seismic noise sources that would otherwise drown out all signals, as well as longer interferometer arm lengths compared to ground-based GW experiments such as LIGO. Both constraints necessitate the use of of space-based laser interferometry experiments. The future Laser Interferometer Space Antenna (LISA) mission (Amaro-Seoane et al., 2017), formally accepted by the European Space Agency (ESA) in 2024 with a planned launch in or around 2035, will consist of a triangular interferometric constellation designed to study in detail the mHz GW regime. This region will be rich with potential

sources of great impact to both astrophysics and the study of gravity. We expect that from the moment the spacecraft turns on there will be thousands of potential GW sources all concurrently contributing to the data stream. Subsequent data analysis pipelines must be capable of decoupling and characterizing the global set of sources alongside detector noise. The development of such "global-fit" MCMC pipelines is of paramount importance, with current prototypes already demonstrating their effectiveness on simulated data (Littenberg & Cornish, 2023; Katz et al., 2024). Not all primary LISA sources are included in the latest iterations of the global fit. Extreme mass-ratio inspirals (EMRIs), sources with great scientific potential but considerably complex waveforms, are not fully integrated and may require more sophisticated MCMC routines to analyze.

This chapter is outlined as follows. In Sec. 5.2 we introduce the Laser Interferometer Space Antenna mission, its observables, and its potential GW sources. Then in Sec. 5.3 we hone in on one particular LISA source, EMRIs, and review standard kludges for their waveforms and responses. We provide insight into the difficulties doing parameter estimation of EMRIs in Sec. 5.4 and present preliminary arguments in support of using Hamiltonian sampling to assuage these issues. Finally in Sec. 5.5 we discuss what further steps are necessary to validate this method for production-level use by LISA.

5.2 THE LASER INTERFEROMETER SPACE ANTENNA

The LISA project is an ESA mission done in collaboration with the National Aeronautics and Space Administration. The design consists of three identical satellites arranged in an equilateral triangle configuration, each side of length 2.5×10^9 m. Each satellite will contain gold/platinum test masses. The satellites will have attached thrusters to maintain their configuration and achieve zero drag for the interior test masses. There will be two optical benches in each arm pointed at the other two spacecraft, creating six links for the

full interferometer. The entire constellation will trail the Earth by 20° as it orbits the Sun, and will be tilted at 60° from the ecliptic plane. This design is illustrated in Figure 5.1

Although the mission idea was first pitched in the 1990s, it wasn't until the 2010s that LISA began to be more formally funded as an ESA project. In 2013, the ESA selected themes for its large class mission slots, including for a theme titled "The Gravitational Universe." The LISA Pathfinder mission, a single satellite designed as a technological demonstration of the full LISA project, launched in 2015 and reached the L1 Lagrange point in 2016. Across its mission duration, the LISA Pathfinder successfully demonstrated its drag-free satellite concept and achieved noise precision near the required levels for the full proposed project. One year later LISA was proposed and accepted as a candidate large class mission for "The Gravitational Universe" theme. In January 2024 the mission was formally adopted by the ESA, acknowledging the mission concept as sufficiently advanced to proceed with instrumentation and hardware development. As of the writing of this dissertation, LISA is expected to launch in 2035.

5.2.1 The LISA Response Function

This section briefly reviews important definitions and calculations related to the response function for GWs in LISA data, derived explicitly in multiple papers (Cutler, 1998; Cornish & Rubbo, 2003).

To begin we start in a detector-based coordinate system with the LISA constellation at rest. Following the standard labeling conventions outlined in the LISA Data Challenge (LDC) Manual (Babak & Petiteau, 2020), we label the spacecrafts clockwise from 1 to 3. Variables are typically given the subscripts (s, l, r) for *sender*, *link*, and *receiver*. Links are numbered by the spacecraft vertex opposite the laser path, i.e., $l = 2$ indexes the laser link traveling from spacecraft 1 ($s = 1$) to spacecraft 3 ($r = 3$). Moving counterclockwise

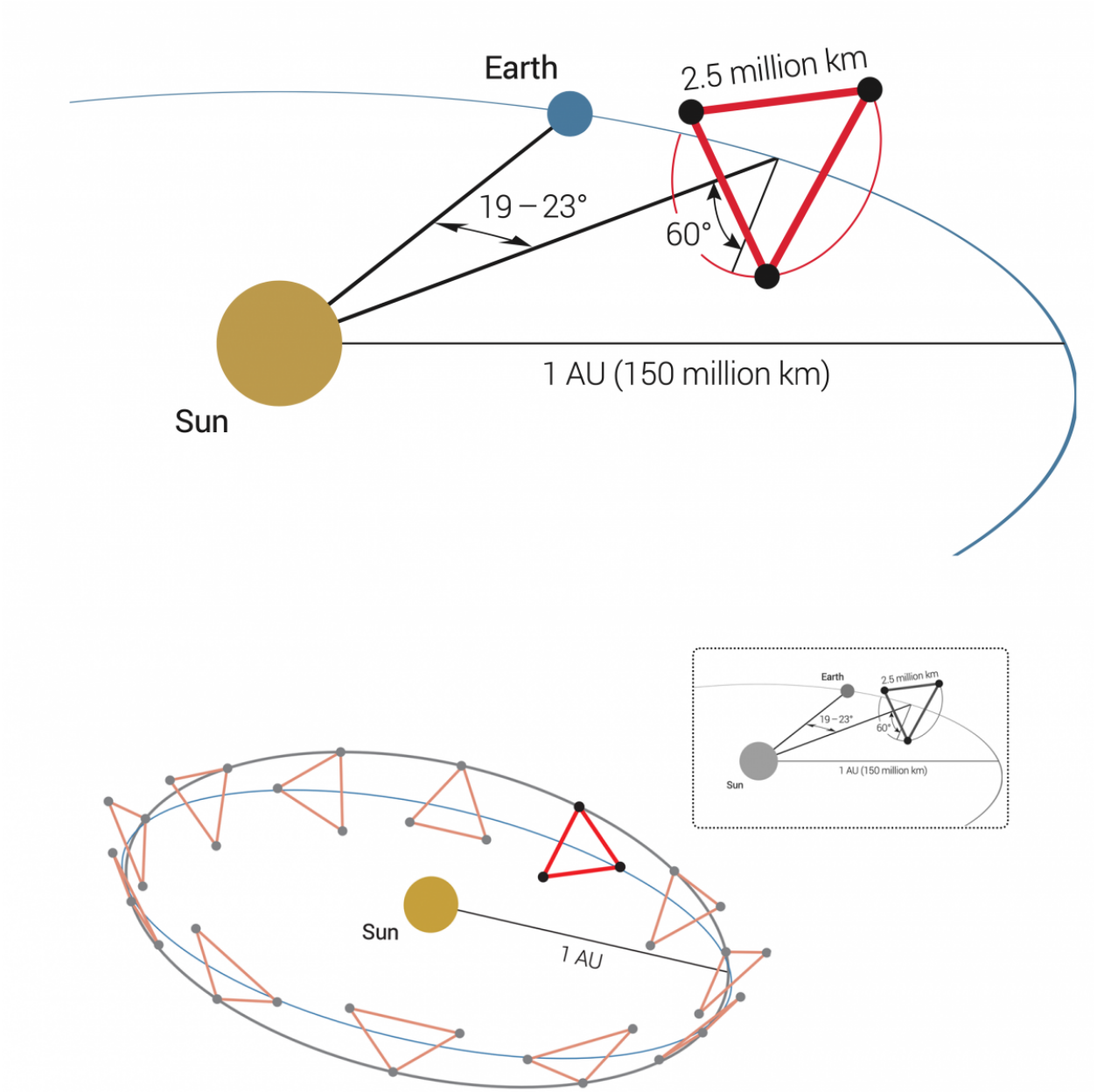

Figure 5.1: Two drawings of the LISA orbit, not to scale. The top panel shows the triangular constellation at one point in its heliocentric orbit. The three satellites are arranged in an equilateral triangle that lag the Earth's orbit by about $20°$. The constellation is inclined at $60°$ relative to the ecliptic plane. The bottom panel displays the annular rotation of the LISA orbit about the ecliptic. Figure from Amaro-Seoane et al. (2017).

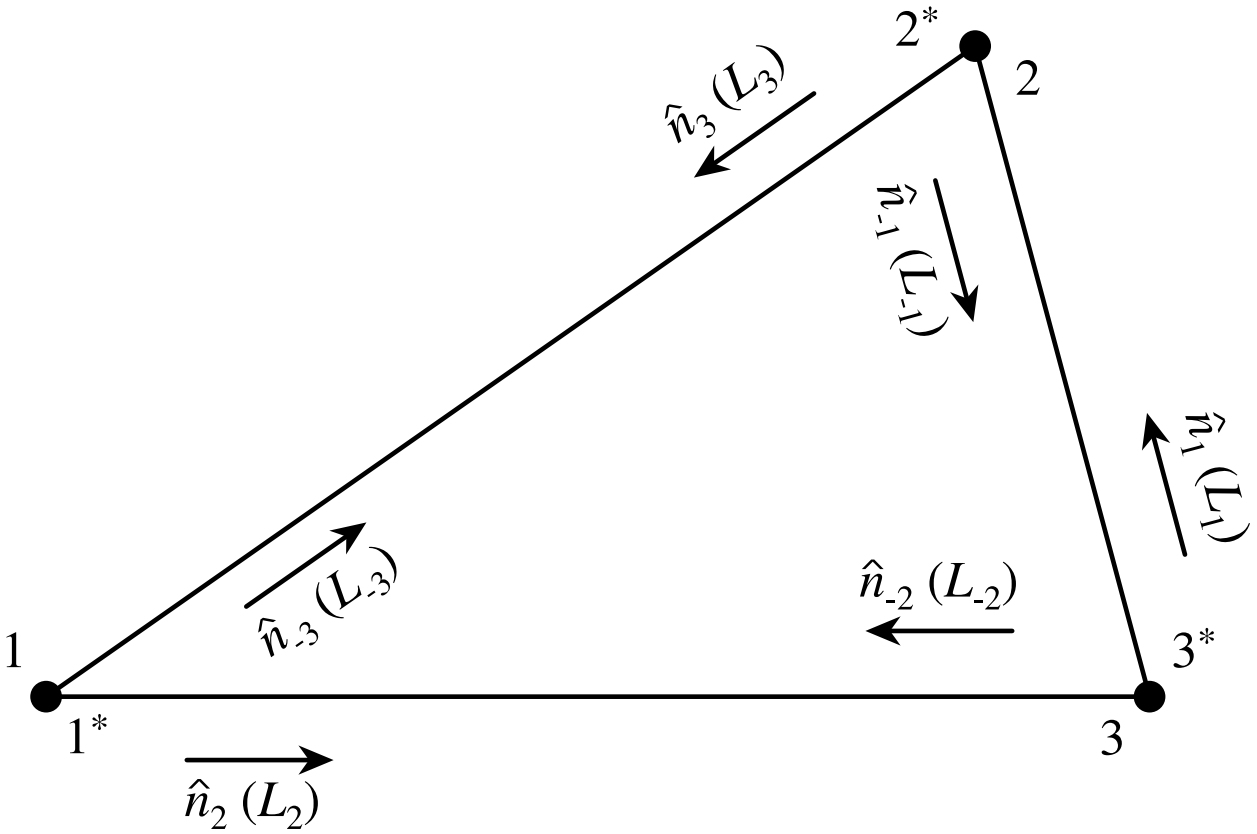

Figure 5.2: Schematic of the LISA constellation defining the indexing convention used in calculating the TDI response variables. All six possible laser path configurations are shown along with their notation for unit vectors $\hat{n}_i$ and light-path lengths L_i Figure from Vallisneri (2005)

primes the indices[1], where for a similar example the laser traveling from 3 to 1 is assigned $l = 2'$. Figure 5.2 provides a helpful schematic for remembering this labeling convention.

A GW source can be described spatially by a unit vector $\mathbf{k}$ pointing from the source to the origin, which in this system is the center of the equilateral triangle. The vector $\mathbf{k}$ is parameterized in terms of the source sky position variables (θ, ϕ),

[1]Occasionally in the LISA literature the indices are negated to mark the counterclockwise laser paths. An example would be the link $l = -3$ denoting the laser traveling from 1 to 2.

$$\mathbf{k} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}. \tag{5.2.1}$$

Additionally, let $\mathbf{R}_i$ denote the position of the "i"-th spacecraft, L_j the length of the "j"-th link, and $\mathbf{n}_j$ its corresponding unit vector. Taking the difference between the strain induced at laser emission and laser reception, the response of a single arm to a GW signal, given as a relative frequency shift of the laser, can be written as (Babak & Petiteau, 2020)

$$\frac{\delta\nu}{\nu} \equiv y_{slr}(t) = \frac{\mathbf{n}_l^i \mathbf{n}_l^j}{2} \frac{h_{ij}(t - \mathbf{k}\cdot\mathbf{R}_s - L_l) - h_{ij}(t - \mathbf{k}\cdot\mathbf{R}_r)}{1 - \mathbf{k}\cdot\mathbf{n}_l}. \tag{5.2.2}$$

Given that LISA has three arms, it can function at any time as a pair of two-arm detectors with two orthogonal signal responses and consequently measure both polarizations of a gravitational wave. Cutler (1998) derived the two orthogonal signal outputs for detectors I and II, given by

$$s_\alpha(t) = \frac{\sqrt{3}}{2}\left[F_\alpha^+(t)h^+(t) + F_\alpha^\times(t)h^\times(t)\right], \qquad (\alpha = I, II), \tag{5.2.3}$$

where the $F_\alpha^{+,\times}$ are the corresponding antenna pattern functions (Apostolatos et al., 1994):

$$\begin{aligned} F_I^+ &= \frac{1}{2}\left(1 + \cos^2\theta\right)\cos(2\phi)\cos(2\psi) - \cos\theta\sin(2\phi)\sin(2\psi), \\ F_I^\times &= \frac{1}{2}\left(1 + \cos^2\theta\right)\cos(2\phi)\sin(2\psi) - \cos\theta\sin(2\phi)\cos(2\psi), \end{aligned} \tag{5.2.4}$$

$$\begin{aligned} F_{II}^+ &= \frac{1}{2}\left(1 + \cos^2\theta\right)\sin(2\phi)\cos(2\psi) - \cos\theta\cos(2\phi)\sin(2\psi), \\ F_{II}^\times &= \frac{1}{2}\left(1 + \cos^2\theta\right)\sin(2\phi)\sin(2\psi) - \cos\theta\cos(2\phi)\cos(2\psi). \end{aligned} \tag{5.2.5}$$

The additional factor of $\sqrt{3}/2$ in Eq. (5.2.3) originates from the arms being separated by $60°$ instead of $90°$ such as in ground-based interferometers such as LIGO.

5.2.2 Time-Delay Interferometry

Due to the large distances between spacecraft, LISA observations are taken via pairs of independent lasers, contrary to ground-based observatories and their classical Michelson interferometers. The measured signals will contain not only any embedded GW signals but also many different noise sources. Of particular importance is the laser frequency noise. The lasers will emit at wavelength 1064 nm, which corresponds to a frequency of $\nu_0 \approx 282$ THz. Noise fluctuations in the lasers will be on the order of MHz, making the laser noise level more than 8 orders of magnitude larger than the expected gravitational wave signals in the data stream. In order to completely suppress this effect, LISA analysis pipelines employ a numerical method called Time Delay Interferometry (TDI). A detailed description and derivation of TDI can be found in Estabrook et al. (2000) and Tinto & Dhurandhar (2021). Here we review the general system of TDI and its base first-generation combinations.

The principle of TDI is given right in its moniker: use linear combinations of delayed interferometric measurements to numerically eliminate the laser noise. The time delays correspond to the light travel time over some multiples of spacecraft distances. To introduce some notation, again keeping to the indexing conventions outlined in the LDC manual, we define a delay operator

$$\mathcal{D}_i y(t) \equiv y(t - L_i), \tag{5.2.6}$$

where L_i denotes the light travel time along link i, ranging from 1 to 3. Chained delays are then simply

$$\mathcal{D}_{i_1,i_2,\ldots,i_n} y(t) = y\left(t - \sum_{m=1}^{n} L_m\right). \tag{5.2.7}$$

One can then construct Michelson-like variables from linear combinations of delayed signal streams that cancel out the laser frequency noise. One such example, the X-TDI variable, can be calculated as (Estabrook et al., 2000)

$$\begin{aligned} X = {} & y_{32'1} + \mathcal{D}_{2'} y_{123} + \mathcal{D}_{2,2'} y_{231} + \mathcal{D}_{322'} y_{13'2} \\ & - \left[y_{231} + \mathcal{D}_{3} y_{13'2} + \mathcal{D}_{3'3} y_{32'1} + \mathcal{D}_{2'3'3} y_{123}\right], \end{aligned} \tag{5.2.8}$$

where the primed indices represent the reverse of the standard cyclic link naming convention[2]. Applying a cyclic permutation of the spacecraft indices yields the Y- and Z-TDI channels. These three variables represent the core observables produced by the interferometer that can be then be utilized in GW search algorithms and analysis pipelines.

There are a few remaining details worth mentioning regarding the application of TDI. First note that the above expression for X makes no assumptions towards the link lengths being equal (similarly for Y and Z). It does, however, assume that the lengths do not vary as a function of time, which for realistic LISA orbits will not be true. In order to account for time dependence of the arm lengths, a set of second-generation TDI has been calculated where the delay operators no longer commute and order is important (Tinto et al., 2023). Lastly, it is worth pointing out that the X, Y, Z Michelson-like TDI variables have correlated noise properties. One can construct an uncorrelated set of TDI variables, typically denoted with the variables A, E, T by the following linear combinations (Prince et al., 2002)

[2] i.e., L_2 denotes the light travel time from the laser emitted at spacecraft 1 and received at spacecraft 3, and $L_{2'}$ the light travel time for the laser emitted at spacecraft 3 and received at spacecraft 1

$$
\begin{aligned}
A &= \frac{1}{\sqrt{2}}\left(Z - X\right), \\
E &= \frac{1}{\sqrt{6}}\left(X - 2Y + Z\right), \\
T &= \frac{1}{\sqrt{3}}\left(X + Y + Z\right).
\end{aligned}
\tag{5.2.9}
$$

The A, E, T variables are the most commonly used form of TDI observables in current LISA analysis pipelines, including the work discussed later in this chapter.

5.2.3 LISA Sources

The mHz frequency regime probed by LISA is replete with a variety of astrophysical sources. The most numerous of sources will be compact binaries in the Milky Way, comprised primarily of white-dwarfs, emitting nearly monochromatic and continuous GW signals. This galactic binary population is so large that it is expected to create a stochastic foreground or confusion signal, with some subset being individually resolvable through GW measurements (Korol et al., 2020; Robson & Cornish, 2017). A smaller collection of the white-dwarf binary population already have well-defined properties through electromagnetic observations and should therefore be detectable by LISA soon after it begins taking data, allowing for quick verification of the instrument's capabilities (Kupfer et al., 2018).

Another set of prime targets for LISA are black hole binary systems across a very wide range of the masses. There is potential for LISA to observe stellar mass ($M \sim 10^1 M_\odot$) and intermediate mass ($10^2 M_\odot < M < 10^4 M_\odot$) black hole binaries at early stages in their inspirals, before being later detectable by ground-based interferometers (Cutler et al., 2019). This presents an exciting opportunity to study GW astrophysics across multiple

frequency bands. Larger black hole binary systems ($10^4 M_\odot < M < 10^7 M_\odot$) should coalesce in the mHz frequency band, meaning LISA should detect the inspiral, merger, and ringdown phases of these interactions (Colpi et al., 2019).

One particularly important potential source are EMRIs, also sometimes called "capture sources" in the literature. These systems describe the capture of stellar-mass compact objects (CO; $\mu \sim 10^0 - 10^2 M_\odot$)[3], by massive black holes (MBH; $M \sim 10^4 - 10^7 M_\odot$) in a galactic center (Amaro-Seoane et al., 2007). The small mass ratio ($q = \mu/M \sim 10^{-5}$) means that EMRI orbits will evolve slowly up until the final plunge of the CO, having roughly $10^4 - 10^5$ waveform cycles over periods of years during the LISA experiment. The waveforms of these systems are in general highly eccentric and displaying extreme relativistic precession, making them incredibly complicated to faithfully model and analyze. The payoff for doing so is great, with EMRIs presenting an excellent opportunity to study MBH populations, probe stellar dynamics in galactic centers, and test strong-field gravity near the MBH (Gair et al., 2010; Berti et al., 2019). The remainder of this chapter will focus entirely on these particular LISA sources, review current strategies in waveform modeling, and outline new methods at performing parameter estimation of these exciting systems.

5.3 GWS FROM EXTREME MASS RATIO INSPIRAL SOURCES

Constructing accurate EMRI waveform models is an ongoing project within the GW community. There exist multiple different model strategies, each with its own positives and drawbacks regarding accuracy and speed. The three most common are the Analytical Kludge (AK; Barack & Cutler 2004), Augmented Analytical Kludge (AAK; Chua et al. 2017), and Numerical Kludge (NK; Gair & Glampedakis 2006). Here a kludge means any

[3]Following the convention in Barack & Cutler (2004) the variable μ denotes the mass of the CO and not the reduced mass for which it was used in Sec. 1.2.2.

approximate but computationally efficient model used for the purpose of data analysis. The AK waveforms are the fastest of the four but also induce the most deviation from realistic EMRI orbits. The NK waveforms are more accurate, even for highly eccentric systems, but considerably slower to calculate. The AAK waveforms use qualities of both the AK and NK models to produce accurate waveforms while maintaining computational speeds comparable to the AK model. The work highlighted in this chapter exclusively uses AK waveforms.

5.3.1 Analytical Kludge Waveform Model

The following is an condensed description of the AK waveform derivation in Barack & Cutler (2004), going through the steps to compute the GW strain polarizations h_+ and $h_\times$, and describing the full parameter space for the GW signal and response.

We begin with a MBH-CO system, letting M denote the mass of the MBH and μ the mass of the CO, located at some distance r from a detector. Recall from Sec. 1.1.2 the expression for the metric perturbation in the transverse, traceless gauge

$$h_{ij}^{TT} = \frac{2}{r}\left(P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}\right)\ddot{I}^{kl}, \tag{5.3.1}$$

where P_{ij} is once again the projection operator and I^{ij} the inertia tensor. In Sec. 1.2.2 we derived $h_+(t)$ and $h_\times(t)$ for a circular binary. Here we will consider a general Newtonian binary orbit parameterized by an eccentricity e and semi-major axis a. The orbital frequency of the binary is then $\nu = (2\pi)^{-1}M^{1/2}a^{-3/2}$. The inertia tensor can now be expressed as a sum $I^{ij} = \sum_n I_n^{ij}$ over the harmonics of the orbital frequency. It has three non-vanishing, independent components that can be described, following Peters and Mathews (Peters & Mathews, 1963), by

$$
\begin{aligned}
\ddot{I}^{11} &= a_n + c_n, \\
\ddot{I}^{12} &= b_n, \\
\ddot{I}^{22} &= c_n - a_n,
\end{aligned}
\tag{5.3.2}
$$

with the coefficients a_n, b_n, c_n defined as

$$
\begin{aligned}
a_n &= -n\mu \left(2\pi\nu M\right)^{2/3} \left[J_{n-2}(ne) - 2eJ_{n-1}(ne) + (2/n)J_n(ne) + 2eJ_{n+1}(ne)\right. \\
&\qquad \left. - J_{n+2}(ne)\right] \cos\left[n\Phi(t)\right], \\
b_n &= -n\mu \left(2\pi\nu M\right)^{2/3} \left(1-e^2\right)^{1/2} \left[J_{n-2}(ne) - 2J_n(ne) + J_{n+2}(ne)\right] \sin\left[n\Phi(t)\right], \\
c_n &= 2\mu \left(2\pi\nu M\right)^{2/3} J_n(ne) \cos\left[n\Phi(t)\right].
\end{aligned}
\tag{5.3.3}
$$

Here J_n denote Bessel functions of the first kind, and the parameter $\Phi(t)$ represents the mean anomaly. If we denote $\Phi_0 = \Phi(t_0)$ for some initial time t_0, then for a Newtonian binary we have the explicit form for $\Phi(t)$[4],

$$
\Phi(t) = 2\pi\nu \left(t - t_0\right) + \Phi_0. \tag{5.3.4}
$$

We next need to define the coordinate basis in which to analyze the detector-source system. Starting with a unit vector $\hat{n}$, which points from the detector directly to the source, we can construct the remaining two right-handed basis vectors

[4]Note that over the timescale of LISA observations $T \sim \mathcal{O}(1\mathrm{yr})$ the detectors' orbital motion will induce an additional non-negligible Doppler phase modulation that needs to be faithfully modeled

$$\hat{p} = \frac{\left(\hat{n} \times \hat{L}\right)}{\left|\hat{n} \times \hat{L}\right|}, \qquad \hat{q} = \hat{p} \times \hat{n}. \tag{5.3.5}$$

The unit vector $\hat{L}$ aligns with the orbital angular momentum of the CO, and in general will vary with time ($\hat{L} = \hat{L}(t)$). From this we can write the GW polarization basis tensors,

$$\begin{aligned} e^{+}_{ab}(t) &= \hat{p}_i \hat{p}_j - \hat{q}_i \hat{q}_j, \\ e^{\times}_{ab}(t) &= \hat{p}_i \hat{q}_j - \hat{q}_i \hat{p}_j, \end{aligned} \tag{5.3.6}$$

We can then fully describe the GW strain using the corresponding n-harmonic components of the polarization amplitude coefficients A^{+}_{n} and $A^{\times}_{n}$ as

$$h_{ij}(t) = \sum_{n,\alpha} A^{\alpha}_{n}(t) e^{\alpha}_{ij} \qquad \alpha \in (+, \times)\,. \tag{5.3.7}$$

Using Eq. (5.3.1) and Eq. (5.3.7) we can explicitly determine A^{+}_{n} and $A^{\times}_{n}$, where we introduce the parameter γ measuring the direction of pericenter,

$$\begin{aligned} A^{+}_{n} &= -\left(1 + \left(\hat{L} \cdot \hat{n}\right)^2\right)\left(a_n \cos(2\gamma) - b_n \sin(2\gamma)\right) + \left(1 + (\hat{L} \cdot \hat{n})^2\right) c_n, \\ A^{\times}_{n} &= 2\left(\hat{L} \cdot \hat{n}\right)\left(b_n \cos(2\gamma) + a_n \sin(2\gamma)\right). \end{aligned} \tag{5.3.8}$$

The remainder of the AK parameter space comes from applying the antenna pattern functions of Eqs. (5.2.4) and (5.2.5) in computing the detector response. Whereas these equations are expressed in terms of the source sky location and polarization angles in a

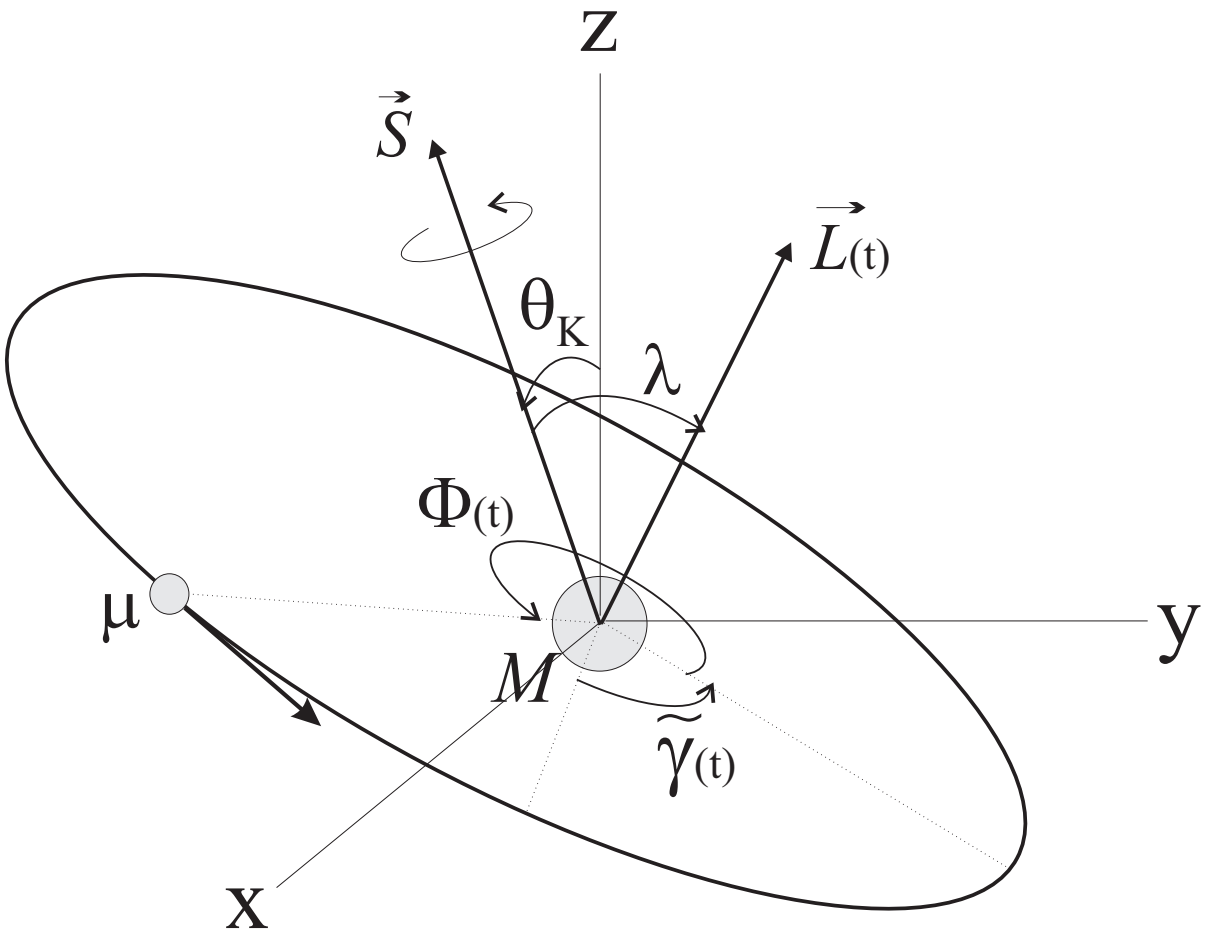

Figure 5.3: Orbital configuration and notation for an EMRI system represented in a Cartesian coordinate system. The MBH and CO masses are given by M and μ, respectively. The orbital angular momentum vector is given by $\vec{L}(t)$, and the spin vector of the MBH is denoted $\vec{S}$. The angle θ_K represents the polar angle of the MBH spin vector. The parameter λ is the angle between $\vec{L}(t)$ and $\vec{S}$. Lastly, the variables $\tilde{\gamma}(t)$ and $\Phi(t)$ define the direction of pericenter and the mean anomaly of the orbit, respectively. Figure from Barack & Cutler (2004).

Parameter	Description	Units
ν_0	Orbital frequency at some time t_0	Hz
M	Mass of MBH	$M_\odot$
μ	Mass of CO	$M_\odot$
S/M^2	Magnitude of the MBH spin	1
e_0	Orbital eccentricity, evaluated as $e(t_0)$	1
γ_0	Direction of pericenter, evaluated as $\gamma(t_0)$	Rad
Φ_0	Mean anomaly, evaluated as $\Phi(t_0)$	Rad
θ_S	Source sky location polar angle	Rad
ϕ_S	Source sky location azimuthal angle	Rad
θ_K	MBH spin polar angle	Rad
ϕ_K	MBH spin azimuthal angle	Rad
α_0	Azimuthal direction of orbital angular momentum, evaluated as $\alpha(t_0)$	Rad
λ	Angle between orbital angular momentum and MBH spin	Rad
D	Distance to the source	Gpc

Table 5.1: Summary of parameters comprising the AK waveform model, along with their descriptions and units.

reference frame moving with the detector, the AK waveform model instead uses a fixed, ecliptic-based frame. The variables θ and ϕ are now given in terms of the source locations parameters in ecliptic-based coordinates θ_S and ϕ_S, and the polarization angle ψ can be written in terms of θ_S, ϕ_S, the magnitude of the MBH spin angular momentum S/M^2, the direction of the MBH spin θ_K, ϕ_K, the angle λ between the spin vector $\hat{S}$ and orbital angular momentum vector $\hat{L}$, and an azimuthal angle α that tracks the precession of $\hat{L}$ around $\hat{S}$. Table 5.1 summarizes the full parameter space for the EMRI analytic kludge waveform and detector response model, and Figure 5.3 provides a visual diagram of the orbital configuration and notation.

Five of the fourteen parameters (ν_0, e_0, γ_0, Φ_0, α_0) are provided at some fiducial reference time t_0, alluding to the fact that they are in general time-dependent functions. While the waveforms are initially based on the Peters and Matthews lowest-order quadrupolar waveforms for general Keplerian orbits, they are then corrected for effects such as Lense-Thirring precession, pericenter precession, and radiation reaction allowing those five or-

bital parameters to evolve with time via Post-Newtonian formulae governing a system of coupled ordinary differential equations (Junker & Schaefer, 1992; Brumberg, 1991; Ryan, 1996; Barker & O'Connell, 1975). The general strategy is then to pick some set of initial parameters, evolve the orbit through some predetermined time or until plunge, then use the result to calculate the coefficients from Eq. (5.3.3), the polarization amplitudes in Eq. (5.3.8), the antenna patterns via Eqs. (5.2.4) and (5.2.5), and finally the signal response given in Eq. (5.2.3).

5.4 BAYESIAN METHODS FOR EMRI DATA ANALYSIS

In this section we review current efforts in performing Bayesian inference of EMRI signals and highlight specific hindrances in the analyses. Specifically we explore the prevalence of many secondary modes in the likelihood surfaces, in large part due to degeneracies of the signal space, and their effect on subsequent parameter estimation runs with MCMC methods. Finally we discuss ongoing work to build such a pipeline, and present preliminary results towards using Hamiltonian sampling for EMRI data analysis.

Performing inference of EMRIs presents a particularly difficult task compared to analyzing other LISA sources. As shown in the previous section the parameter space is quite large with fourteen free parameters, and can be extended further to seventeen if we wish to include the magnitude and orientation of the spin of the CO. Waveforms need to be generated at sub-second speeds in order to search across the full parameter space, and templates need to minimize the phase error across the long duration signals (Amaro-Seoane et al., 2011). New computational frameworks are being developed to aid in fast and accurate waveform generation (Katz et al., 2021), but their immediate applications to parameter estimation prove they are still unable to access the full posterior.

Additionally we must consider the possibility of correlations in the signal space, broadly

separated into the classes of confusion and degeneracy. Self-confusion in EMRI signals arises from correlations among multiple different signals in the LISA data volume. This is not expected to be a problem for analysis pipelines due largely to their highly uncertain event rates and the large volume of the signal space (Babak et al., 2017). Degeneracy, on the other hand, presents a major problem for the development of EMRI parameter inference methods (Chua & Cutler, 2022). This manifests from non-local correlations in the signal space for a given waveform model, and has already been noted by participants in the Mock LISA Data Challenges (Babak et al., 2010).

For our exploratory analyses we make use of the LDC data products. The first iteration of the new LDC, given the name Radler[5], contains an EMRI-specific challenge denoted LDC 1-2. The data contain a single GW signal injection from an EMRI under an idealized noise model. The injection is modeled with the AK waveforms. The waveform and response are given in the time domain, and the TDI channels are provided in terms of the X, Y, Z variables. The signal is produced assuming an observing cadence of 15 s. It is derived from a source with MBH and CO masses of $1.1 \times 10^6 M_\odot$ and $29.5 M_\odot$, respectively. It has a spin magnitude of 0.97, initial eccentricity of 0.23, and initial orbital frequency of 7.4×10^{-4} Hz.

In order to characterize the likelihood surface for this problem, we vary the likelihood separately a function of the two components masses to visualize the expected multimodal structure. The actual calculation uses the waveform generating code from Katz et al. (2021). One key difference between this code and the AK model is that it is parameterized in terms of the initial semi-latus rectum p_0 rather than the initial orbital frequency ν_0. We calculate p_0 following the numerical procedure outlined in Barack & Cutler (2004).

Figure 5.4 displays the two sets of negative log-likelihood values, with the source in-

[5] Link to Radler Data Challenge

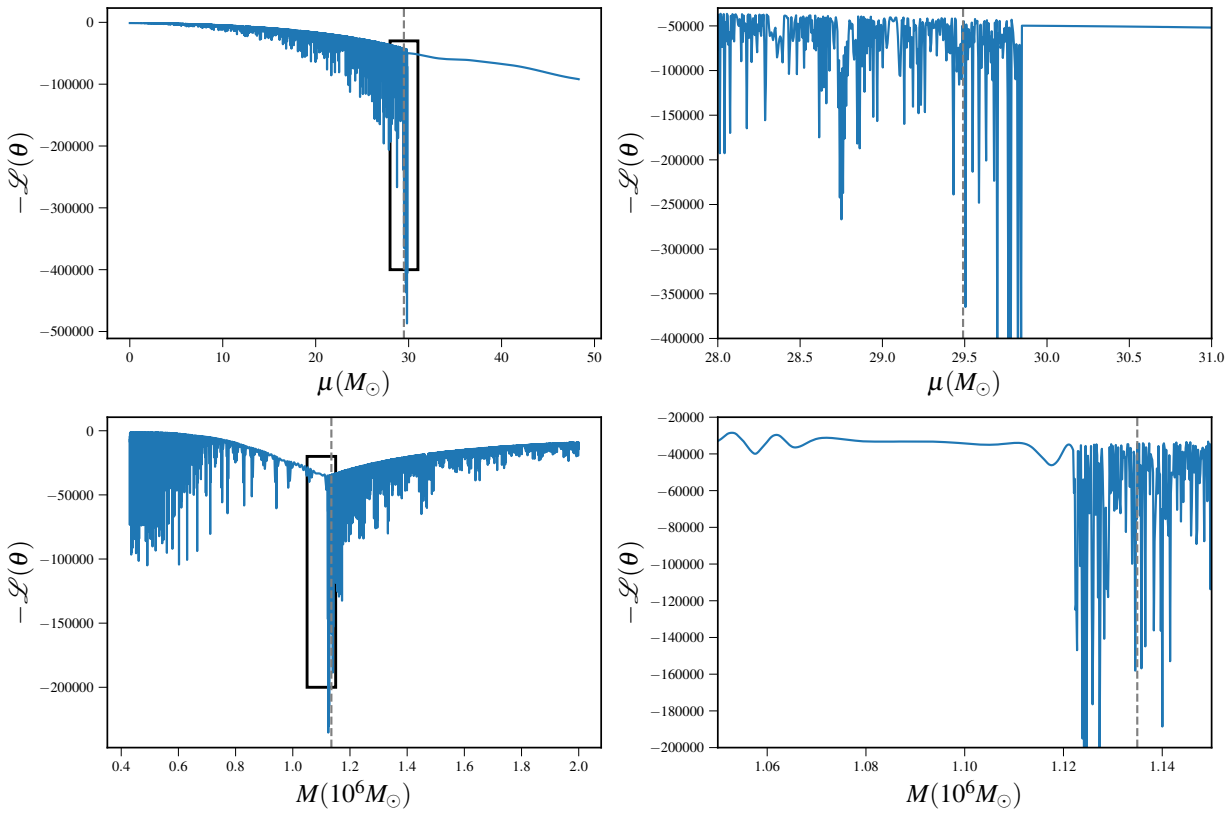

Figure 5.4: (Top) The negative log-likelihood evaluated as a function of the CO mass μ across the full width of its prior. The right panel shows a close-up of a portion of the log-likelihood emphasizing the jagged behavior. All parameters not shown are fixed to to those of the source injection in the Radler dataset. The dashed gray lines represent the injected value of the CO mass. (Bottom) Similar plots for the MBH mass M.

jections denoted by dashed gray lines. The two left panels show the full widths of the priors, while the two right panels contain zoomed-in views on narrow ranges of the parameters. In all four plots the jagged structure is evident, indicating numerous secondary extrema where an MCMC sampler may become trapped. We also note that the injected values do not exactly align with the points of maximal likelihood, implying a more accurate estimate of p_0 may be necessary.

The consequences of this behavior in the likelihood is apparent when running a full inference scheme. Standard parallel-tempering MCMC pipelines struggle to find the posterior maximum (Chua & Cutler, 2022), and likely necessitate either increasing the number of different temperatures for the chains to values much larger than that for other LISA analyses or greatly reducing the prior widths. There is currently no existing code to run inference on EMRI data with other sampling algorithms. The goal of this work is develop, test, and utilize a Hamiltonian sampler for these analyses.

We, as a first step, have built a package to calculate quick TDI X, Y, Z response variables for AK waveforms of EMRIs, based on the original code for this purpose found in the `EMRI_Kludge_Suite` package of the Black Hole Perturbation Toolkit. The new code is written entirely in Python using `JAX` (Bradbury et al., 2018), making it fully GPU-compatible. By using the AK waveforms, we also directly utilize the full capabilities of autodifferentiation. In its current state we can compute the TDI variables for a set of waveform parameters in $\mathcal{O}(1\,\mathrm{s})$ on a CPU, and $\mathcal{O}(0.1\,\mathrm{s})$ on a GPU. If we use the source injection values for the LDC, we can very nearly recover the output response from the Radler data set. This is shown explicitly in Figure 5.5, with the Radler TDI variables plotted in black and the result from this work in blue.

More must be done to transform our standalone TDI generator into a full fledged HMC inference pipeline. In particular we need to add implementations of the likelihood

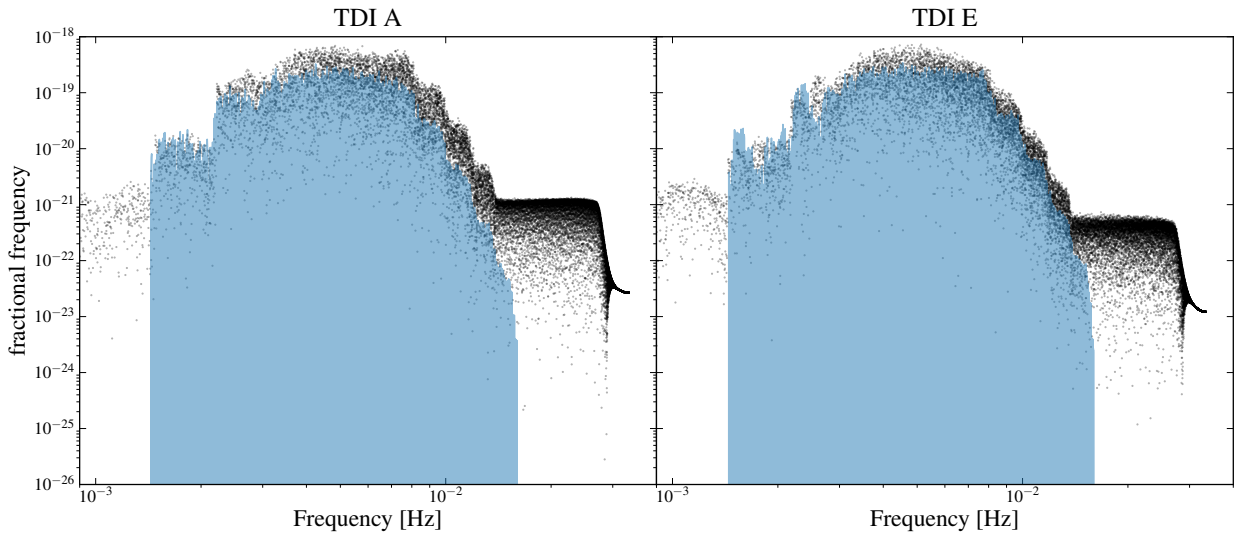

Figure 5.5: Plots of TDI values as a function of frequency for an EMRI matching the source parameters in the Radler LDC dataset. The blue denotes the TDI as computed using our newly developed package and the black dots are the true values from the simulated data. The left and right plots show the TDI A and E channels, respectively. We find good agreement between the simulated data and our new AK code for calculating TDI variables.

function and priors for all parameters. At that point this will allow us to take complete gradients of the likelihood with `JAX`. Then we can determine the effect to which using Hamiltonian sampling proposals aid in efficiently exploring the EMRI parameter space, We will see whether it can do so accurately compared to present attempts with parallel-tempering MCMC, and to what extent we can widen the priors on our parameters while still maintaining a similar level of accuracy. This work is ongoing.

5.5 DISCUSSION

Despite being one of the most scientifically rich sources in the LISA band, there is to date no robust implementation of parameter estimation for EMRIs either as a standalone pipeline or as an integration into the global fit. Deeper studies into sampling these models reveal further complications but little in the ways of practical solutions. The large

dimensionality and numerous degeneracies pose significant challenges to both efficiently exploring the signal space as well as being able to converge on a single solution. It is evident that we will require treating the inference in a manner different to the remainder of LISA analyses in order to accurately detect these sources.

Here we reviewed the LISA response function, the construction of TDI variables, and GW signals from EMRIs given through the AK waveform model. We briefly summarized previous work in studying the correlations in this parameter space and its effects on parameter estimation. Even using the idealized source injections in the LDC datasets we still encounter similar difficulties in. The problems will be exacerbated in realistic LISA data when multiple sources are simultaneously present. We have begun the development of a new pipeline for EMRI analysis that will use Hamiltonian sampling to aid in dealing with these complexities. Currently we have demonstrated a working implementation of the full response function that can accurately match the source injection in the Radler simulated dataset. The code can also be natively run on GPUs, which lowers the TDI computation time to speeds more conducive for production-level analyses.

Further development will provide the first insight as to what degree HMC sampling can improve upon existing detection algorithms in exploring the source posterior distributions. It will also establish whether the less accurate AK waveforms are still acceptable given an improvement in sampling efficiency, or if adjustments to include more up-to-date AAK waveforms is necessary. Additional extension to newer-generation TDI variables (Tinto et al., 2023) is also required to bring this package in line with current LISA data infrastructure requirements. Later tests against multi-source injections can demonstrate if EMRIs need to be treated separately from other GW sources in its own distinct pipeline, or if a true global fit for all of LISA data is truly attainable.

CHAPTER 6

Conclusion

This dissertation comprises a complete outline of the development and application of Hamiltonian sampling tools towards the detection and characterization of various nHz GW sources using pulsar timing. We have demonstrated its effectiveness and provided arguments supporting the method's full adoption into Bayesian inference pipelines. The usefulness and applicability of these methods extend across the GW frequency spectrum, and to that end we have begun applying these methods to searches within LISA data. We have particularly honed in on the problem of data analysis of EMRI sources.

Chapter 3 presents the primary evidence that HMC sampling is a more effective analysis method long term for PTAs (Freedman et al., 2023). We performed many different statistical tests comparing Hamiltonian sampling against the MCMC techniques currently in use, and found general consistency between the two for standard GWB inference problems. Furthermore, we found that both methods followed their expected scaling relations for sample computation time vs model dimension, a strong indicator that building a full Hamiltonian sampling pipeline will be more beneficial long term. This is clearly evident when considering that future PTA datasets will include ever-increasing data volume, both in terms of the number of pulsars as well as their overall observing timespan, and combined IPTA datasets will contain $\mathcal{O}(100)$ pulsars.

Much effort in the PTA community goes towards faster and more efficient computation of stochastic background signals, however oftentimes those methods apply only to that specific subset of models and not to the broader class of potential sources. Our work has no such restriction, and we show this in Chapter 4 by extending our code to jointly search for both stochastic and deterministic GW signals (Freedman & Vigeland, 2024). We

show that we can accurately recover both injected signals, even in cases where the two may have stronger covariance. Moreover, by using our method we can more easily search wider areas on the sky rather than gridding the analysis into hundreds of small batches. Accurately sampling the GWB alongside additional sources is a necessary addition to any analysis pipeline, as we expect the significance of the background to only grow in future data releases. Although in this work we only consider circular SMBHB signals, it is straightforward to extend to SMBHBs in eccentric orbits, or even other deterministic sources. It can also be applied to advanced noise modeling of individual pulsars.

Data analysis methods are in many cases agnostic of the data we input into them, so to that effect we also aim to apply HMC sampling techniques that we have proven to work for PTA searches to other GW experiments. In Chapter 5 we honed in specifically on EMRI signals, which to date represents an as of yet unresolved piece of the LISA data science puzzle. We outline our strategy and plans for building an HMC sampler explicitly for EMRIs, and aim to develop this code and make it available as an integral part of the greater collection of LISA data analysis tools.

6.1 FUTURE PROSPECTS

There are several further directions to explore based on the research outlined in this dissertation, broadly classified into projects related to PTAs and projects related to LISA. Several of these projects are ongoing. Regarding GW inference in the low-frequency regime, there is the motivation to expand the work presented in Chapters 3 and 4 to utilize Hamiltonian sampling for the full class of PTA searches. This in essence would comprise an updated and enhanced version of `enterprise`, the current workhorse analysis code. This would open up enhancements to other core searches of interest, such as generic GW bursts, multiple SMBHB signals, or any of the plethora of new and exotic physics sources

that can be constrained with PTAs. Using a similar framework to the code developed here, specifically using the `JAX` ecosystem, would also enable the full leveraging of GPU computing resources, providing the necessary speedups for our ever increasing data volume. This comprehensive development endeavor is a continuing effort across the wider PTA community.

Regarding future extensions to LISA projects, the immediate tasks are outlined in Sec. 5.5. We expect in the coming months to have the first tests of using Hamiltonian sampling methods towards efficiently performing inference of EMRI GW signals. This will need to be validated against a suite of simulated EMRI datasets, ideally beyond the sole one provided through the LDC. A future HMC pipeline for these sources could in theory be integrated alongside current global fit codes for one single robust analysis product in advance of mission launch. If we can perform accurate inference of the signals, it leads naturally to follow-up questions on the possibilities for multi-messenger astrophysics with EMRIs, and to what degree we can determine their sky location to perform any sort of targeted electromagnetic followup campaign.

6.2 FINAL REMARKS

The past ten years mark the dawn of the golden age of gravitational wave astronomy as an observational science. In just one decade's time we have gone from the very first detection of GWs from a merging compact binary system to a growing catalog of events in the high-frequency regime. PTA collaborations have gone from setting gradually increasing constraints on the stochastic GWB to providing the first conclusive evidence of its existence. The next ten years promise to be just as exciting, if not more. We expect to see many more detections from ground-based interferometers, a decisive detection of the low-frequency GWB and perhaps even GWs from single SMBHB sources, and the con-

clusion of the next ten years may mark the official launch of the LISA mission and the beginning phases of mHz GW astronomy. As our experiments expand in size, complexity, and observing capacity, it is imperative that we continue to be innovative with our data analysis methods in order to keep pace with this growth.

BIBLIOGRAPHY

Abbott B. P., et al., 2016, Phys. Rev. Lett., 116, 061102

Acernese F., et al., 2015, Classical and Quantum Gravity, 32, 024001

Agazie G., et al., 2023a, ApJ, 951, L8

Agazie G., et al., 2023b, ApJ, 951, L9

Agazie G., et al., 2024a, ApJ, 963, 61

Agazie G., et al., 2024b, ApJ, 963, 144

Akutsu T., et al., 2021, Progress of Theoretical and Experimental Physics, 2021, 05A101

Alam M. F., et al., 2021, ApJS, 252, 4

Alder B. J., Wainwright T. E., 1959, The Journal of Chemical Physics, 31, 459

Amaro-Seoane P., Gair J. R., Freitag M., Miller M. C., Mandel I., Cutler C. J., Babak S., 2007, Classical and Quantum Gravity, 24, R113

Amaro-Seoane P., Schutz B., Thornburg J., 2011, arXiv e-prints, p. arXiv:1102.3647

Amaro-Seoane P., et al., 2017, arXiv e-prints, p. arXiv:1702.00786

Andrieu C., Thoms J. A., 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

Apostolatos T. A., Cutler C., Sussman G. J., Thorne K. S., 1994, Phys. Rev. D, 49, 6274

Arzoumanian Z., et al., 2016, ApJ, 821, 13

Arzoumanian Z., et al., 2018a, ApJS, 235, 37

Arzoumanian Z., et al., 2018b, ApJ, 859, 47

Arzoumanian Z., et al., 2020, ApJ, 905, L34

Arzoumanian Z., et al., 2023, ApJ, 951, L28

Babak S., Petiteau A., 2020, https : / / lisa-ldc . lal . in2p3 . fr / static / data / pdf / LDC-manual-002.pdf

Babak S., Sesana A., 2012, Phys. Rev. D, 85, 044034

Babak S., et al., 2010, Classical and Quantum Gravity, 27, 084009

Babak S., et al., 2017, Phys. Rev. D, 95, 103012

Barack L., Cutler C., 2004, Phys. Rev. D, 69, 082005

Barker B. M., O'Connell R. F., 1975, Phys. Rev. D, 12, 329

Bécsy B., Cornish N. J., 2020, Classical and Quantum Gravity, 37, 135011

Bécsy B., Cornish N. J., Digman M. C., 2022a, Phys. Rev. D, 105, 122003

Bécsy B., Cornish N. J., Kelley L. Z., 2022b, ApJ, 941, 119

Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307

Berti E., et al., 2019, BAAS, 51, 32

Betancourt M., 2016, arXiv e-prints, p. arXiv:1604.00695

Betancourt M. J., Girolami M., 2013, arXiv e-prints, p. arXiv:1312.0906

Blanco-Pillado J. J., Olum K. D., Shlaer B., 2014, Phys. Rev. D, 89, 023512

Bradbury J., et al., 2018, JAX: composable transformations of Python+NumPy programs, http://github.com/google/jax

Brumberg V. A., 1991, Essential relativistic celestial mechanics.. CRC Press

Burke-Spolaor S., et al., 2019, A&A Rev., 27, 5

Cabezas A., Corenflos A., Lao J., Louf R., 2024, BlackJAX: Composable Bayesian inference in JAX (arXiv:2402.10797)

Chamberlin S. J., Creighton J. D. E., Siemens X., Demorest P., Ellis J., Price L. R., Romano J. D., 2015, Phys. Rev. D, 91, 044048

Charisi M., et al., 2022, MNRAS, 510, 5929

Chua A. J. K., Cutler C. J., 2022, Phys. Rev. D, 106, 124046

Chua A. J. K., Moore C. J., Gair J. R., 2017, Phys. Rev. D, 96, 044005

Colpi M., et al., 2019, arXiv e-prints, p. arXiv:1903.06867

Corbin V., Cornish N. J., 2010, arXiv e-prints, p. arXiv:1008.1782

Cornish N. J., Rubbo L. J., 2003, Phys. Rev. D, 67, 022001

Creighton J., Anderson W., 2011, Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis.

Creutz M., 1988, Phys. Rev. D, 38, 1228

Cutler C., 1998, Phys. Rev. D, 57, 7089

Cutler C., et al., 2019, BAAS, 51, 109

Detweiler S., 1979, ApJ, 234, 1100

Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956

Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Physics Letters B, 195, 216

EPTA Collaboration et al., 2023, A&A, 678, A50

Efron B., 1979, Annals of Statistics, 7, 1

Einstein A., 1916, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 688–696

Ellis J. A., 2013, Classical and Quantum Gravity, 30, 224004

Ellis J., van Haasteren R., 2017, jellis18/PTMCMCSampler: Official Release, doi:10.5281/zenodo.1037579, https://doi.org/10.5281/zenodo.1037579

Ellis J. A., Siemens X., van Haasteren R., 2013, ApJ, 769, 63

Ellis J. A., Vallisneri M., Taylor S. R., Baker P. T., 2020, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, Zenodo, doi:10.5281/zenodo.4059815, https://doi.org/10.5281/zenodo.4059815

Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152

Estabrook F. B., Tinto M., Armstrong J. W., 2000, Phys. Rev. D, 62, 042002

Foster R. S., Backer D. C., 1990, ApJ, 361, 300

Freedman G. E., Vigeland S. J., 2024, Phys. Rev. D, 110, 063038

Freedman G. E., Johnson A. D., van Haasteren R., Vigeland S. J., 2023, Phys. Rev. D, 107, 043013

Gair J. R., Glampedakis K., 2006, Phys. Rev. D, 73, 064037

Gair J. R., Tang C., Volonteri M., 2010, Phys. Rev. D, 81, 104014

Gair J., Romano J. D., Taylor S., Mingarelli C. M. F., 2014, Phys. Rev. D, 90, 082001

Gelman A., Rubin D. B., 1992, Statistical Science, 7, 457

Geman S., Geman D., 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721

Gersbach K. A., Taylor S. R., Meyers P. M., Romano J. D., 2025, Phys. Rev. D, 111, 023027

Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Gregory P., 2010, Bayesian Logical Data Analysis for the Physical Sciences

Grishchuk L. P., 1976, Soviet Journal of Experimental and Theoretical Physics Letters, 23, 293

Haiman Z., et al., 2009, ApJ, 700, 1952

Harte J., 1985, Consider a spherical cow: A course in environmental problem solving. American Assn. for Artificial Intelligence,Menlo Park, CA, https://www.osti.gov/biblio/6147157

Hastings W. K., 1970, Biometrika, 57, 97

Helfand D. J., Manchester R. N., Taylor J. H., 1975, ApJ, 198, 661

Hellings R. W., Downs G. S., 1983, ApJ, 265, L39

Hewish A., Bell S. J., Pilkington J. D. H., Scott P. F., Collins R. A., 1968, Nature, 217, 709

Hoffman M. D., Gelman A., 2011, arXiv e-prints, p. arXiv:1111.4246

Iguchi S., Okuda T., Sudou H., 2010, ApJ, 724, L166

Jaynes E. T., Baierlein R., 2004, Physics Today, 57, 76

Jenet F. A., Lommen A., Larson S. L., Wen L., 2004, ApJ, 606, 799

Jensen J. L. W. V., 1906, Acta Mathematica, 30, 175

Junker W., Schaefer G., 1992, MNRAS, 254, 146

Katz M. L., Chua A. J. K., Speri L., Warburton N., Hughes S. A., 2021, Phys. Rev. D, 104, 064047

Katz M. L., et al., 2024, arXiv e-prints, p. arXiv:2405.04690

Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2018, MNRAS, 477, 964

Kormendy J., Ho L. C., 2013, ARA&A, 51, 511

Korol V., et al., 2020, A&A, 638, A153

Kramer M., Champion D. J., 2013, Classical and Quantum Gravity, 30, 224009

Kupfer T., et al., 2018, MNRAS, 480, 302

LIGO Scientific Collaboration et al., 2015, Classical and Quantum Gravity, 32, 074001

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lamb W. G., Taylor S. R., van Haasteren R., 2023, Phys. Rev. D, 108, 103019

Larsen B., et al., 2024, arXiv e-prints, p. arXiv:2405.14941

Lasky P. D., et al., 2016, Physical Review X, 6, 011035

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., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, Phys. Rev. D, 87, 104021

Littenberg T. B., Cornish N. J., 2023, Phys. Rev. D, 107, 063004

Lorimer D. R., Kramer M., 2012, Handbook of Pulsar Astronomy

Maggiore M., 2018, Gravitational Waves: Volume 2: Astrophysics and Cosmology, doi:10.1093/oso/9780198570899.001.0001.

Manchester R. N., et al., 2013, PASA, 30, e017

Max A. W., 1950, in , Memorandum Rept. 42, Statistical Research Group. Princeton Univ., p. 4

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem. Phys., 21, 1087

Miles M. T., et al., 2023, MNRAS, 519, 3976

Milosavljević M., Merritt D., 2003, ApJ, 596, 860

Mingarelli C. M. F., Mingarelli A. B., 2018, Journal of Physics Communications, 2, 105002

Mingarelli C. M. F., Sidery T., Mandel I., Vecchio A., 2013, Phys. Rev. D, 88, 062005

Mingarelli C. M. F., et al., 2017, Nature Astronomy, 1, 886

Moore C. J., Cole R. H., Berry C. P. L., 2015, Classical and Quantum Gravity, 32, 015014

Neal R. M., 2003, The Annals of Statistics, 31(3), 705–767

Neal R., 2011, in , Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC, pp 113–162, doi:10.1201/b10905

Nesterov Y., 2009, Mathematical Programming, 120, 221

Ostriker J. P., Hausman M. A., 1977, ApJ, 217, L125

Papaspiliopoulos O., Roberts G., Sköld M., 2007, Statist Sci, 22

Perera B. B. P., et al., 2019, MNRAS, 490, 4666

Peters P. C., Mathews J., 1963, Physical Review, 131, 435

Phinney E. S., 2001, arXiv e-prints, pp astro–ph/0108028

Prince T. A., Tinto M., Larson S. L., Armstrong J. W., 2002, Phys. Rev. D, 66, 122002

Quinlan G. D., 1996, New A, 1, 35

Ransom S., et al., 2019, in Bulletin of the American Astronomical Society. p. 195 (arXiv:1908.05356), doi:10.48550/arXiv.1908.05356

Ravi V., Wyithe J. S. B., Shannon R. M., Hobbs G., 2015, MNRAS, 447, 2772

Reardon D. J., et al., 2023a, ApJ, 951, L6

Reardon D. J., et al., 2023b, ApJ, 951, L7

Richstone D., et al., 1998, Nature, 385, A14

Robson T., Cornish N., 2017, Classical and Quantum Gravity, 34, 244002

Rosado P. A., Sesana A., Gair J., 2015, MNRAS, 451, 2417

Rosenbrock H. H., 1960, The Computer Journal, 3, 175

Ryan F. D., 1996, Phys. Rev. D, 53, 3064

Sampson L., Cornish N. J., McWilliams S. T., 2015a, Phys. Rev. D, 91, 084055

Sampson L., Cornish N. J., McWilliams S. T., 2015b, Phys. Rev. D, 91, 084055

Sánchez Almeida J., Aguerri J. A. L., Muñoz-Tuñón C., de Vicente A., 2010, ApJ, 714, 487

Sardesai S. C., Vigeland S. J., Gersbach K. A., Taylor S. R., 2023, Phys. Rev. D, 108, 124081

Sazhin M. V., 1978, Soviet Ast., 22, 36

Sesana A., 2013, MNRAS, 433, L1

Sesana A., Haardt F., Madau P., Volonteri M., 2004, ApJ, 611, 623

Sesana A., et al., 2005, ApJ, 623, 23

Shannon R. M., Cordes J. M., 2010, ApJ, 725, 1607

Siemens X., Mandic V., Creighton J., 2007, Phys. Rev. Lett., 98, 111101

Sminchisescu C., Welling M., 2011, Pattern Recognition, 44, 2738

Susobhanan A., 2023, Classical and Quantum Gravity, 40, 155014

Susobhanan A., Gopakumar A., Hobbs G., Taylor S. R., 2020, Phys. Rev. D, 101, 043022

Tarafdar P., et al., 2022, PASA, 39, e053

Taylor S. R., 2021, arXiv e-prints, p. arXiv:2105.13270

Taylor J. H., Weisberg J. M., 1982, ApJ, 253, 908

Taylor S., Ellis J., Gair J., 2014, Phys. Rev. D, 90, 104028

Taylor S. R., Huerta E. A., Gair J. R., McWilliams S. T., 2016, ApJ, 817, 70

Taylor S. R., Simon J., Schult L., Pol N., Lamb W. G., 2022, Phys. Rev. D, 105, 084049

Tinto M., Dhurandhar S. V., 2021, Living Reviews in Relativity, 24, 1

Tinto M., Dhurandhar S., Malakar D., 2023, Phys. Rev. D, 107, 082001

Vallisneri M., 2005, Phys. Rev. D, 71, 022001

Vallisneri M., 2020, Astrophysics Source Code Library, p. ascl:2002.017

Vallisneri M., van Haasteren R., 2017, MNRAS, 466, 4954

Vigeland S. J., Islo K., Taylor S. R., Ellis J. A., 2018, Phys. Rev. D, 98, 044003

Wadadekar Y., 2005, PASP, 117, 79

Wald R. M., 1984, General Relativity

White S. D. M., 1980, MNRAS, 191, 1P

Xu H., et al., 2023, Research in Astronomy and Astrophysics, 23, 075024

Yu Q., 2002, MNRAS, 331, 935

van Haasteren R., 2024, arXiv e-prints, p. arXiv:2406.05081

van Haasteren R., Levin Y., 2010, MNRAS, 401, 2372

van Haasteren R., Levin Y., 2013, Mon. Not. Roy. Astron. Soc., 428, 1147

van Haasteren R., Vallisneri M., 2014, Phys. Rev. D, 90, 104012

van Haasteren R., Levin Y., McDonald P., Lu T., 2009, MNRAS, 395, 1005

van Haasteren R., et al., 2011, Monthly Notices of the Royal Astronomical Society, 414, 3117

APPENDIX

Deriving the Hamiltonian Monte Carlo Scaling Relation

A key characteristic of the HMC algorithm is that its time to produce a statistically independent sample in a Markov chain scales with the dimensionality d of the model as $\mathcal{O}(d^{5/4})$ compared to a MH MCMC sampler, which scales as $\mathcal{O}(d^2)$. Here we briefly go over how these two quantities are derived. The complete description is found in Creutz (1988), and an updated overview added in Neal (2011). Our derivation closely follows the latter.

First assume that we are sampling the probability density $P(x) = (1/Z)\exp(-E(x))$ where $E(x)$ is our energy function, such as the Hamiltonian for HMC algorithms, for our variables x (so in keeping to the conventions outlined in Ch. 2, $x \sim \theta$ for MH MCMC and $x \sim (p, q)$ for HMC sampling) and Z the partition function. Now we consider the class of Monte Carlo algorithms that use the Metropolis-Hasting acceptance condition with probability

$$\alpha = \min\left\{1, e^{-\Delta}\right\}, \tag{6.2.1}$$

where $\Delta = E(x^*) - E(x)$ is the energy difference between the proposed and current states. Creutz (1988) notes that the expectation value of the energy difference term is $\mathbb{E}[\exp(-\Delta)] = 1$. This equality implies acceptance of every sample if they are all drawn exactly from the distribution $P(x)$. The choice of MC sampler, however, induces some error associated with how it traverses the parameter space. From Jensen's inequality (Jensen, 1906) we have

$$\mathbb{E}\left[\Delta\right] \geq 0, \tag{6.2.2}$$

meaning that on average, proposed states will have higher energy than the current state. If the proposal steps are not chosen carefully, this can lead to the rejection of nearly all samples. If we denote $\mathbb{E}\left[\Delta_d\right]$ as the total energy difference for the full parameter space, the goal is to keep $\mathbb{E}\left[\Delta_d\right] \approx 1$.

We consider a generic Gaussian proposal distribution $\mathcal{N}(0, \sigma^2)$. In a random-walk Metropolis sampler, the energy displacement will grow proportional to σ^2 times the number of steps n. In order to obtain a nearly independent sample we require $n\sigma^2 \sim 1$ or $n \sim \sigma^{-2}$. Additionally we have $\mathbb{E}\left[\Delta_d\right] \propto d\sigma^2$, which in order to maintain a reasonable acceptance rate we must scale $\sigma \propto d^{-1/2}$. Putting the two together results in the number of steps needed to de-correlate as $n \sim d$, and since each iteration requires updating all d parameters this leads to an overall computation cost of $\mathcal{O}(d^2)$.

Following a similar prescription for the HMC algorithm, we start by assuming that we are using a leapfrog integrator to simulate Hamiltonian dynamics. We define our simulation step size to be ε and total number of steps as L. In order to obtain an independent sample we need $L\varepsilon \sim 1$, or $L \sim \varepsilon^{-1}$. There is a numerical error of $\mathcal{O}(\varepsilon^2)$ due to our

choice of integration method. Across all dimensions in our space this error accumulates to $\mathbb{E}[\Delta_d] \propto d\varepsilon^4$. Once again we require this to be of order 1 to maintain reasonable sample acceptance, which means that $\varepsilon \sim d^{-1/4}$. The number of leapfrog steps per independent trajectory then scales as $L \sim d^{1/4}$. Each step will require d computations which gives us our main result, that the total computational cost of generating an independent sample with the HMC algorithm scales as $\mathcal{O}(d^{5/4})$.