WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Graduate Theses, Dissertations, and Problem Reports

2022

Bayesian Methods for Multi-Messenger Analysis of Supermassive Black Hole Binaries: Pulsars and Quasars and Gravitational Waves, Oh My!

Caitlin A. Witt
West Virginia University, caw0057@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

 Part of the Cosmology, Relativity, and Gravity Commons, and the External Galaxies Commons

Recommended Citation

Witt, Caitlin A., "Bayesian Methods for Multi-Messenger Analysis of Supermassive Black Hole Binaries: Pulsars and Quasars and Gravitational Waves, Oh My!" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11259.
https://researchrepository.wvu.edu/etd/11259

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Bayesian Methods for Multi-Messenger Analysis of Supermassive Black Hole Binaries: Pulsars and Quasars and Gravitational Waves, Oh My!

Caitlin A. Witt

Dissertation Submitted to
The Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Physics

Sarah Burke-Spolaor, Ph.D., Chair
Maura McLaughlin, Ph.D.
Kathryn Williamson, Ph.D.
Jenny Greene, Ph.D.

Morgantown, West Virginia, USA
2022

Keywords: Supermassive Black Holes, Gravitational Waves, Pulsars, Active Galactic Nuclei
Copyright 2022 Caitlin A. Witt

Abstract

Bayesian Methods for Multi-Messenger Analysis of Supermassive Black Hole Binaries: Pulsars and Quasars and Gravitational Waves, Oh My!

Caitlin A. Witt

Supermassive black hole binaries (SMBHBs) can lurk, often unseen, in the centers of post-merger galaxies, and pulsar timing arrays (PTAs) are rapidly approaching the sensitivities required to detect nanohertz gravitational waves (GWs) from these giant pairs. Independently, numerous electromagnetic surveys are seeking evidence of these dynamic duos' effects on their host galaxies by searching for periodicities in time-domain observations. Combining these two methods to use multi-messenger techniques allows us to learn more about these binaries than using one messenger alone. In this thesis, we have created Bayesian methods to search for SMBHBs using electromagnetic observations of quasars and through GW emission in PTA data.

By using electromagnetic observations to identify an SMBHB candidate, we gain numerous pieces of information that also define the source's GW emission, including the location of and distance to the SMBHB's host galaxy and the orbital period of the SMBHB. In our study, we developed the first multi-messenger techniques used by the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), and applied them to a well-known supermassive black hole binary candidate, 3C 66B. We placed the lowest chirp mass limit to date on an SMBHB within 3C 66B of $\mathcal{M} < 1.65 \times 10^9 M_\odot$. Furthermore, we learned that multi-messenger techniques can lead to a factor of two reduction over all-sky GW searches at the same frequency, and an order of magnitude reduction for completely "blind" searches.

Next, we analyzed the capabilities of Bayesian methods to search for electromagnetic signatures of these binaries in simulated time-domain data sets from next-generation surveys. We developed a Bayesian model selection technique to identify periodicities induced into a quasar light curve by the orbital motion of an SMBHB from within intrinsic red noise. We discovered that future surveys, such as the Legacy Survey of Space and Time (LSST), will identify more robust SMBHB candidates than current surveys, such as the Catalina Real-time Transient Survey (CRTS).

Finally, we present the results of searches for bright continuous GWs (CWs) from individual SMBHBs in NANOGrav's 12.5-year data set. A red noise process, which could be the first signs of an emerging stochastic GW background (GWB), was previously detected for the first time in this data set. In our work, we searched for CWs alongside this common noise process for the first time in real PTA data, and developed necessary data-handling techniques which will be critical for the detection of a CW, which may come soon after the potentially imminent detection of the GWB.

Dedication

Dedication, if any, goes here.

Acknowledgements

There are many people that deserve thanks for all of the effort, time, and care they put in to help me reach this point. First I would like to thank my advisor Sarah Burke-Spolaor for all of the support she has given me over the last five years.

Much of this thesis is the work of not just me, but contains contribution from the entire NANOGrav collaboration. In particular, I'd like to acknowledge the work of the collaborators, mentors, and friends that went out of their way to help me find my place in the collaboration. Sarah Vigeland, Steve Taylor, Joe Simon, Maria Charisi, Michael Lam, Jeff Hazboun, and Megan Jones, thank you for repeatedly taking the time to lift me up along the way.

I would also never have reached this point, motivation intact, without the graduate student community within WVU Physics and Astronomy and NANOGrav. Thank you for the unwavering support, neverending sass, and constant encouragement as we worked to create a more equitable department and collaboration. Additionally, thank you to PAGSO for the trust you placed in me while I served as president. The opportunity to advocate on behalf of this group let me find a level of professional bravery that I never knew I possessed, and it has been a joy to watch you all flourish in the years since.

Among the friends and family that got me through the last few years, first and foremost is Brent Shapiro-Albert. You've supported me unconditionally over the past six years, and poured your time and energy into countless practice talks, applications, and paper drafts. Thank you for the miles of hikes you took me on to decide our next steps, the lazy nights at home to help me recharge, and the excited discussions that reminded me why I loved astronomy. Thank you for walking this path alongside me, I wouldn't want it any other way.

A few more of you also deserve your own shout-outs. Jeff Ouellette, thanks for teaching me how to finish a drink before grad school made it a necessity. Marissa Tranquilli and Alyssa Biscoglio, thank you for many TV nights, distractions, and rants. See you on New Year's Eve like always! Megan Jones, Amber Lenon, and Belinda Cheeseboro, thank you for the many ladies' nights, makeup Fridays, and always being around for a quick chat. Mitch and Elena Shapiro-Albert, thanks for making me an honorary sibling. I'll always be grateful that you made me feel so welcome, especially when we gang up on Brent. And thank you to my parents and grandma for helping me begin the journey.

Finally, the fact that I've finished my PhD with even a grain of sanity remaining is certainly due to the community I've found in Morgantown that's helped me take breaks away from research, especially the friends (human and equine) that I made at Valley Ridge Farm. Thank you Michele, Liz, and Logan for constantly reminding me that I already had the skills I needed, and thank you Mick for helping me find the confidence to show them off.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Frequently Used Terms:

Supermassive Black Hole . . . . . SMBH
Gravitational Wave . . . . . GW
Supermassive Black Hole Binary . . . . . SMBHB
Active Galactic Nucleus . . . . . AGN
Damped Random Walk . . . . . DRW
Pulsar Timing Array . . . . . PTA
Time-of-Arrival . . . . . TOA
Millisecond Pulsar . . . . . MSP
Modified Julian Day . . . . . MJD
Very Long Baseline Interferometry . . . . . VLBI
Continuous Wave . . . . . CW
Gravitational Wave Background . . . . . GWB
Common Red Noise . . . . . CRN
Markov-Chain Monte Carlo . . . . . MCMC
Bayes Information Criterion . . . . . BIC

Collaborations, Experiments, and Surveys:

North American Nanohertz Observatory for Gravitational waves . . . NANOGrav
European Pulsar Timing Array . . . . . EPTA
Parkes Pulsar Timing Array . . . . . PPTA
Indian Pulsar Timing Array . . . . . InPTA
Chinese Pulsar Timing Array . . . . . CPTA
International Pulsar Timing Array . . . . . IPTA
Laser Interferometer Gravitational-wave Observatory . . . . . LIGO
Laser Interferometer Space Antenna . . . . . LISA
Legacy Survey of Space and Time . . . . . LSST
Catalina Real-Time Transient Survey . . . . . CRTS
Palomar Transient Factory . . . . . PTF
Zwicky Transient Facility . . . . . ZTF

Frequently Used Symbols:

Solar Mass . . . . . $M_\odot$

Chapter 1

Introduction

Over a century has passed since the introduction of General Relativity (Einstein, 1916), and in the time since, black holes have risen to the forefront of astronomical investigation. From massive stellar remnants to gargantuan galactic cores, these compact mysteries have captured both light and the imagination. Despite the fact that they cannot be observed directly, black holes' significant gravitational impact on the material in their environments and on space-time itself have allowed astronomers to study these invisible giants. This thesis will focus on the largest members of the black hole family, supermassive black holes (SMBHs), as well as the numerous experiments that will unite to provide an unprecedented picture of SMBHs. By combining electromagnetic observations with gravitational wave (GW) analyses, this thesis seeks to forge the tools necessary to enable a multi-messenger detection of an SMBH binary (SMBHB), and explore the improvements that multi-messenger techniques can offer to our understanding of these dynamic duos.

1.1 Supermassive Black Holes

Black holes come in all shapes and sizes, with many orders of magnitude separating the masses of the smallest stellar-mass black holes from their largest cousins, SMBHs. Recent experiments have been able to dramatically expand our

knowledge of stellar-mass black holes by electromagnetic observations with GW detections to populate a "graveyard" of these stellar remnants (Abbott et al., 2021).

With masses exceeding 10^6 $M_\odot$, SMBHs are thought to reside in the centers of nearly every massive galaxy (Kormendy & Ho, 2013), with evidence ranging from gravitational analyses of stellar distributions within galactic cores (Ferrarese & Merritt, 2000; Ghez et al., 2008), to statistical analyses of active galactic nuclei (AGN), to direct imaging (Event Horizon Telescope Collaboration et al., 2019).

The formation of stellar-mass black holes can be clearly linked to the life-cycle of massive stars, but the processes needed to grow an SMBH are not as clear. Similar mechanisms, such as the direct collapse of an enormous gas cloud, could be possible (Bromm & Loeb, 2003). Alternatives formation methods include steady growth of a seed black hole through ongoing accretion (Kulier et al., 2015), but it is unclear if this is possible within the age of the universe. Finally, it is speculated that SMBHs may grow through mergers of seed black holes, which may occur in dense clusters of many black holes or through mergers of galaxies each containing black holes (Kulier et al., 2015). These formation theories each have implications for the resulting SMBHs, from accretion driven activity (discussed in subsection 1.1.1) to the formation and eventual merger of SMBHs binaries (SMBHBs), as discussed in subsection 1.1.2.

1.1.1 Active Galactic Nuclei

In this thesis, the main electromagnic markers we will use to study SMBHs

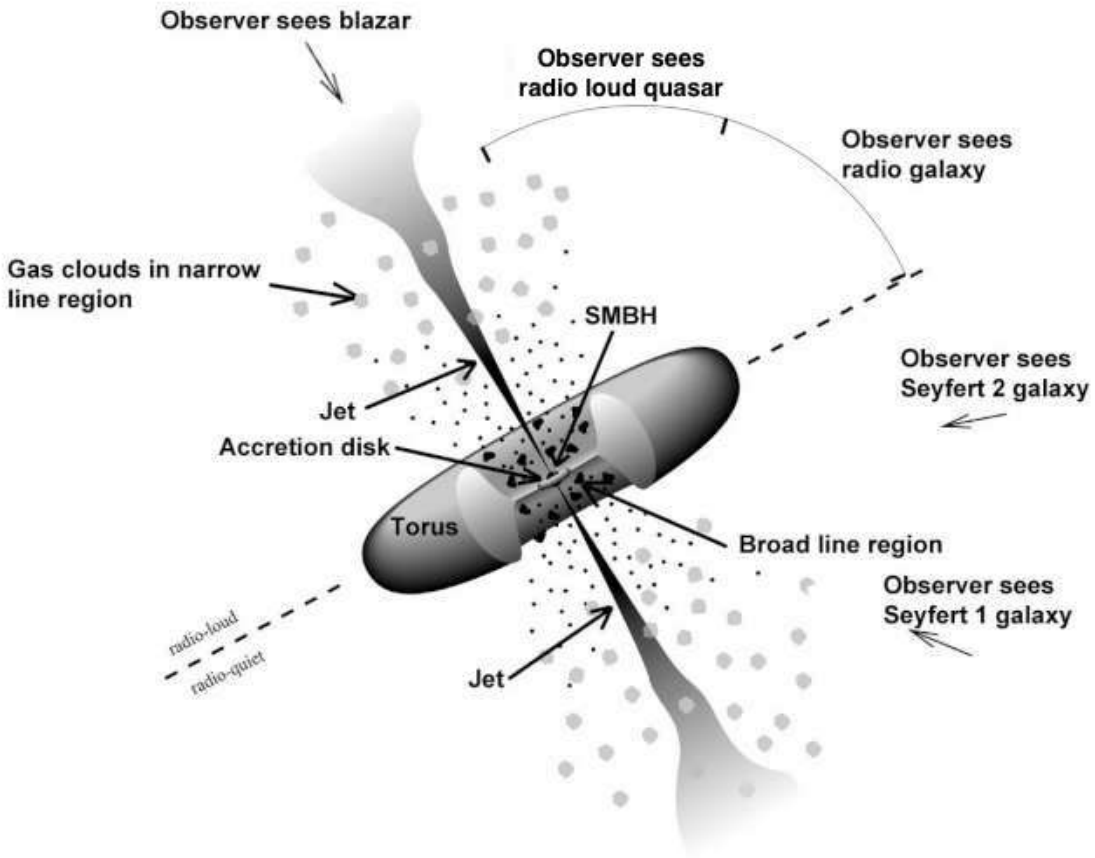

Figure 1.1: Diagram depicting the main features included in unified models of AGN. Image Credit: Robert Findlay. Figure reproduced with permission from Condon & Ransom (2016), permission conveyed through Copyright Clearance Center, Inc.

are AGN. AGN are compact galactic cores with enormous luminosities due to active accretion of material onto a central SMBH. To unify the wide variety of AGN-related phenomena observed across the electromagnetic spectrum, astronomers have reduced theories of the components of AGN to the following key elements that make up the unified model of AGN (Urry & Padovani, 1995):

- a central SMBH surrounded by an accretion disk,
- a region of fast-moving material which creates velocity-broadened spectral features known as a broad-line region,

- a dusty torus, which obscures some observable features depending on the observer's orientation,
- regions of slow-moving material which emit ordinary, narrow spectral features, known as the narrow-line region,
- a jet of relativistic particles launched from the central SMBH.

These features, along with sub-classes based on observed properties, are diagrammed in Figure 1.1. One such class, quasars, includes the brightest AGN that can be seen across the electromagnetic spectrum. The bright nucleus vastly outshines the rest of the galaxy and appears nearly point-like in images, giving rise to the original name of this class, "quasi-stellar objects." Quasars are distant AGN that are viewed from a line of sight that is nearly aligned with the jet axis, and therefore appear distinct from stars by exhibiting spectral features from the visible broad-line region, synchrotron emission from the AGN jet, and redshifts that equate to large luminosity distances (i.e. distance measured according to the flux-luminosity relationship).

1.1.1.1 Quasar Variability

When observed in the optical regime of the electromagnetic spectrum, most quasars exhibit significant (to the order of 20%) stochastic flux variability (Hook et al., 1994). Potential causes for these random variations in flux include accretion-disk instabilities, numerous supernova and rapidly evolving remnants in the nuclear region, and microlensing of the core by an orbiting distribution of compact objects (Kawaguchi et al., 1998; Aretxaga et al., 1997; Hawkins, 1993; Trèvese & Vagnetti,

2002).

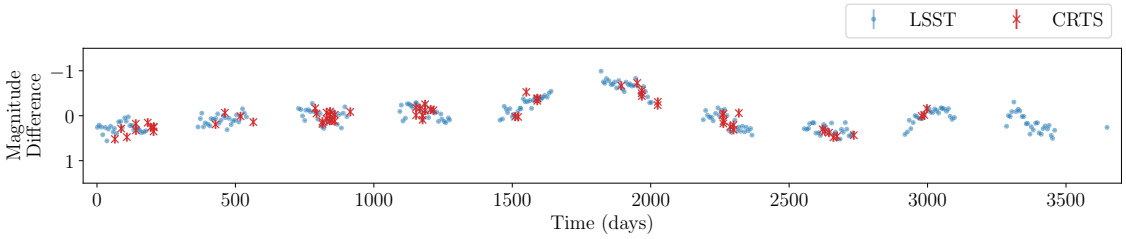

Figure 1.2: Simulated photometric data of an AGN with significant DRW variability.

The most common model used to describe this variability has been dubbed a "damped random walk," (DRW; Kelly et al., 2009; MacLeod et al., 2010; Kozłowski et al., 2010) which includes a damping term to return the time series towards the average value in addition to a pure random walk. This creates a characteristic noise pattern that is "red," or time-dependent, on short timescales but "white", or gaussian-distributed, on long timescales. The DRW noise is modeled by a power spectral density function

$$P(f) = \frac{4\sigma^2\tau}{1 + (2\pi\tau f)^2}, \tag{1.1}$$

where σ^2 is the variance of the data, τ is the characteristic timescale of the DRW process, and f is the Fourier-space frequency.

Quasars are a key object of interest of numerous time-domain photometric surveys, including the Sloan Digital Sky Survey (SDSS), Catalina Real-Time Transient Survey (CRTS), the Palomar Transient Factory (PTF), and Zwicky Transient Facility (ZTF), among many others. Using the resultant data sets, significant efforts have been made to characterize the variability of the quasar population (e.g. MacLeod et al., 2010). However, due to the selection effects of each individual survey and the

wavelength-dependent nature of quasar variability, the underlying distributions of the variability parameters are difficult to infer (Kozłowski et al., 2010) for the entire quasar population. Future surveys such as the Legacy Survey of Space and Time (LSST) may be able to probe new areas of parameter space due their high cadence observations across a range of wavelengths, and will provide invaluable data for use in the characterization of this effect (LSST Science Collaborations et al., 2017).

1.1.2 Supermassive Black Hole Binaries

As will be described in section 1.3, SMBHBs are among the most luminous nanohertz GW sources, a key focus of this thesis. Here, I describe their formation and key attributes. As each massive galaxy is thought to harbor a central SMBH, when these galactic hosts undergo a major merger, an opportunity exists for the formation of SMBHBs. The evolution of these titanic pairs is a process that takes many millions of years, with aspects still open to debate. Through interactions with massive components of the merged host galaxies (e.g. gas, stars), dynamical friction will first cause the SMBHs to sink to the center of the merger remnant. Through continued interactions with the dense environment of a galactic core, the now gravitationally-bound binary's orbit will continue to shrink and evolve (Begelman et al., 1980).

There are a few quantities that we will define to describe the basic attributes of an SMBHB; later in subsubsection 1.3.1.3, we will expand these parameters to include a more complete description of the binary. First are the masses of the two

SMBHs, with the larger, m_1, being dubbed the primary, and the smaller, m_2, the secondary. The sum of these, $m_1 + m_2$, is the total mass of the system, sometimes written as M_{tot}. The SMBHB mass ratio q, is, logically, defined as m_2/m_1. This implies that for an SMBHB consisting of with two nearly equal-mass SMBHs, q will be ~ 1; unequal-mass systems will have $q < 1$. Finally, to describe the orbit of the SMBHB system, we can define various quantities including the orbital separation (a) and orbital period ($P_{\rm orb}$). It is important to note that this orbital period will be different in the rest frame of the binary and that of the observer at cosmological distances; the observed redshifted value of P, in the rest frame, $P_r = P/(1 + z)$.

If the SMBHB can evolve to sufficiently small separations (< 0.1 pc), GW emission will continue the binary's evolution towards an eventual coalescence. However, the hotly-debated "final-parsec problem" raises the question of whether or not there will be sufficient material to drive evolution to these very close separations. As the binary evolves through dynamical interactions, the surrounding environment will lose material through scattering processes. If there are not sufficient amounts of material to drive the dynamical interactions necessary to evolve the orbit until GWs can take over, the binary will "stall" its evolution, perhaps indefinitely (e.g. Colpi, 2014). This possibility has enormous implications for multi-messenger astrophysics, the main focus of this thesis, which will be discussed in more detail in section 1.5. If SMBHBs cannot overcome the final parsec problem, they will never emit detectable nanohertz GWs. Furthermore, the complex processes that may help drive an SMBHB to sub-parsec separations may dramatically influence the resultant electromagnetic and GW signatures of the binary. As shown in Figure 1.3, these

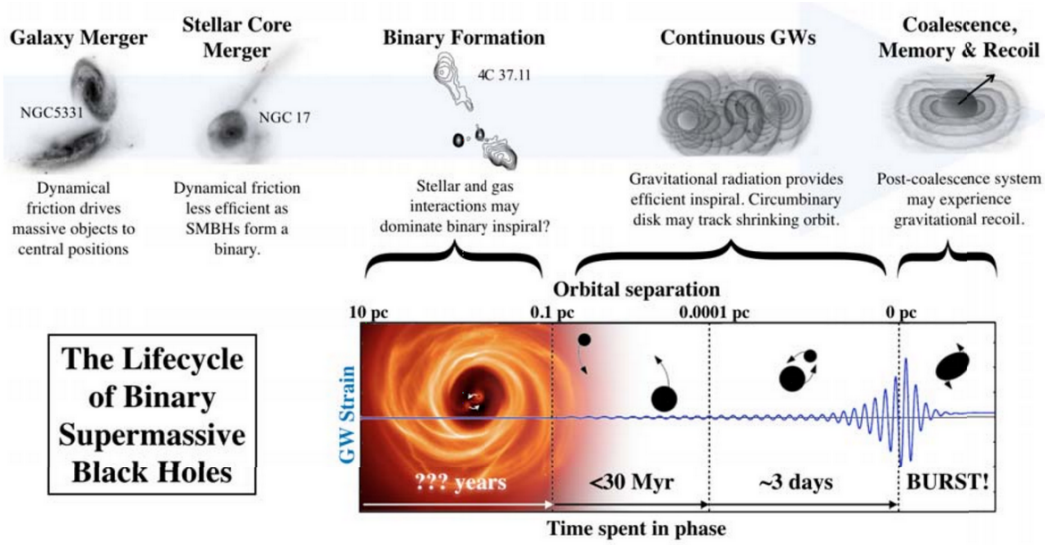

Figure 1.3: After a major galaxy merger, there is a long lifecycle of SMBHBs before their eventual coalescence. Key stages will have effects on both the electromagnetic emission from the SMBHB and the host galaxy, as well as any GW emission. Image credits: Galaxies, Hubble/STSci; 4C37.11, Rodriguez et al. (2006); Simulation visuals, C. Henze/NASA; Circumbinary accretion disk, C. Cuadra. Figure reproduced with permission from Burke-Spolaor et al. (2019) under the Creative Commons Attribution 4.0 International License (CC-BY 4.0).

two messengers will each display unique signatures that indicate the evolutionary progression of the binary.

The presence of an SMBHB and its evolving dynamics can induce changes to the typical structure of an AGN described in subsection 1.1.1. Clearly, now two SMBHs are present, rather than a single central engine surrounded by a single accretion disk. Instead, in most models of SMBHBs where AGN activity is present, gas is expected to settle in a circumbinary disk (Barnes, 2002), with a central cavity that has been cleared by the binary. Quasi-periodically, gas will stream through the cavity to form mini-disks around each of the SMBHs (Farris et al., 2014), with a higher proportion of this gas accreting onto the secondary black hole due to its location nearer to the edge of the cavity.

1.2 Electromagnetic Signatures of SMBHBs

Astronomers have sought the confident discovery of an SMBHB for many years, and a few hundred candidates have been proposed using a variety of electromagnetic observations. These range from direct imaging of structures containing an SMBH to indirect spectroscopic evidence of orbital motion of the proposed binary. Each method, summarized below, probes a distinct range of orbital separations of the two SMBHs, creating a unique suite of techniques necessary to investigate the SMBHB population.

1.2.1 Imaging of Dual and Binary AGN

The most straightforward SMBHB detection technique is direct imaging of two distinct regions that can each be linked to an SMBH. This technique spans the electromagnetic spectrum, with candidates discovered through emission from radio to X-rays (Saade et al., 2020a). However, this method can only probe well separated SMBH pairs. If the SMBHs are so widely separated so as to not yet be gravitationally bound, but are separated by less than 1 kpc, they are referred to as "dual SMBHs" or "dual AGN." However, some techniques, such as Very Long Baseline Interferometry (VLBI), are able to probe more closely separated binaries with separations less than a few parsecs. These more closely separated pairs can be referred to as "binary" AGN or SMBHs, the most famous of which is 0402+379 (Rodriguez et al., 2006). Despite the incredible resolutions made possible with VLBI, these sources are still much too widely separated to produce significant GW

emission.

1.2.2 Radio Jet Morphologies

The jets launched from the central region of an AGN typically follow a predictable geometry, with the jet axis lying approximately perpendicular to the plane of the central accretion disk (Donea & Biermann, 2002). However, many AGN jets display significant helical structure, indicating a precession of the jet orientation. This has been theorized to be a potential SMBHB indicator, with the orbital motion of the binary causing periodic changes in the accretion disk orientation, which therefore affects the direction of the jet. As the relativistic particles are launched away from the base of the jet, this periodically shifting direction results in a helical structure. However, these observations are also limited by the resolution of VLBI telescopes, and therefore unable to investigate effects due to closely separated SMBHBs.

There are numerous SMBHB candidates (Qian et al., 2019) that have been discovered through precision tracking of these evolving jets. However, there is also the possibility that this jet precession is possible without the presence of an SMBHB due to the warping of the accretion disk due to orbital instabilities.

1.2.3 Emission Line Profiles

The high velocities of gas orbiting very close to the central SMBH create broad spectral line profiles that can be observed depending on the orientation of the AGN.

In the presence of an SMBHB, these broad lines may appear shifted in frequency when compared to the narrow lines emitted by slow-moving gas further from the central region of the AGN as the active SMBH orbits either toward or away from the observer. In the case of a binary where both SMBHs are active, a double broad line feature (Begelman et al., 1980) may be observed on either side of the narrow line. While numerous sub-parsec candidates have been detected (Eracleous et al., 2012), broad emission lines are truncated at very small separations, leaving the need for alternative methods to probe the most closely-separated binaries.

1.2.4 Light Curve Features

The continually changing system of an orbiting SMBHB creates opportunities for dramatic features in long-term photometric observations of the host galaxy. One of the longest-standing SMBHB candidates, OJ 287 (Lehto & Valtonen, 1996), is one such example. This eccentric, unequal-mass system is observed as a blazar, and contains one active $10^{10} M_{\odot}$ SMBH. Quasi-periodic flares observed in the decades-long light curve point to the existence of a second, smaller ($10^{8} M_{\odot}$) black hole, which, during its orbit, plunges through the accretion disk of the primary SMBH twice per orbit. This rapid change in the accretion disk vastly increases the temperature, causing a sharp increase in brightness followed by a gradual decline.

Nearly edge-on systems are also expected to produce sharp gravitational-lensing spikes in photometric light curves. As one black hole moves behind its companion as seen by the observer, its emission can be self-lensed by its binary

companion to drastically amplify the observed brightness (Hu et al., 2020) of the host.

Finally, longer timescale periodicities provide an excellent way to find SMBHB candidates in extended time-domain surveys of quasars, such as PG 1302–102 (Graham et al., 2015a). These periodicities may be induced by either periodically-varying accretion rates onto the component SMBHs (MacFadyen & Milosavljević, 2008) as perturbations of the circumbinary disk lead to periodic streaming of gas into the cavity. The induced periodic variations may be at the orbital period of the binary, particularly for unequal-mass systems. However, for nearly equal-mass binaries, the period of the fluctuations may occur at several times the orbital period (D'Orazio et al., 2013).

Periodic Doppler boosting of the emission from the mini-disks surrounding the SMBHs due to their relativistic velocities within the binary orbit (De Rosa et al., 2019) will also induce periodicities into quasar light curves. Emission will be boosted according to the Doppler factor

$$\mathcal{D} = \sqrt{\frac{1+\beta_{\parallel}}{1-\beta_{\parallel}}}, \tag{1.2}$$

where $\beta_{\parallel} = v_{\parallel}/c$, the line of sight component of the emitting SMBH's velocity, normalized by the speed of light. For an SMBHB in a circular orbit, the periodic

observed flux will become

$$\begin{aligned} F_\nu &= F^{\rm sec}_{\nu,0}\left[1+(3-\alpha_\nu)\,v_{2\parallel}/c\right] \\ &= F^{\rm sec}_{\nu,0}\left[1+(3-\alpha_\nu)\,v_2\sin i\sin\left(\omega t+\Phi_0\right)/c\right] \end{aligned}, \tag{1.3}$$

where α_ν is the source spectral index, v_2 and $F^{\rm sec}_{\nu,0}$ are the velocity and flux of the secondary mini-disk, i is the inclination of the orbital plane, ω is the orbital velocity, and Φ_0 is the initial orbital phase. For a full derivation, see Charisi et al. (2022).

This method will be more likely to detect unequal-mass binaries, due to the preferential accretion of gas onto the smaller SMBH in such a binary, which is also moving faster, therefore inducing a higher Doppler boost. In this case, the period of the Doppler boost effect can be related directly back to the orbital period of the binary (D'Orazio et al., 2015; Charisi et al., 2018). However, it is important to note that such a periodic feature will be present in addition to the intrinsic quasar variability discussed in subsubsection 1.1.1.1, as shown in Figure 1.4, necessitating careful analyses to disentangle a stochastic red DRW process, which can appear quasi-periodic, from a true periodicity.

Depending on the orientation of the binary and the makeup of the system, these features may appear in various combinations. These three methods are some of the few that are able to probe sub-parsec SMBHBs, the same class which also emit nanohertz GWs; therefore, this technique will be related the most directly to the following sections of this thesis.

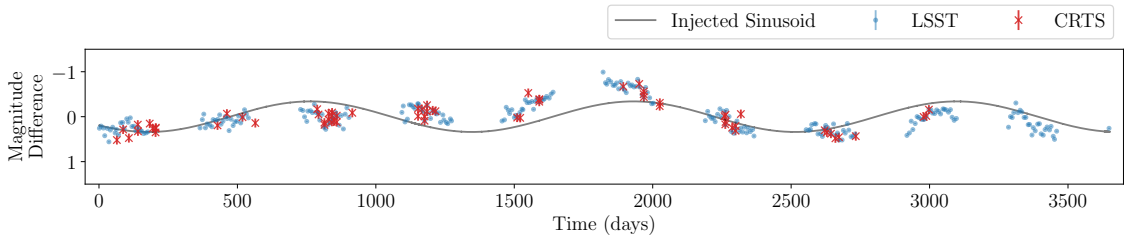

Figure 1.4: Simulated photometric data of an AGN with significant DRW variability and an injected periodicity.

1.2.5 Other Signatures

While the previous sections focus on a few electromagnetic indicators of SMBHBs that represent key stages or features, there are numerous others that have been suggested in the literature. These include features in the continuum spectra of the host, including low soft X-ray and ultraviolet luminosity (Tanaka et al., 2012; Tanaka & Haiman, 2013), clipped spectral features surrounding the FeKα spectral line (Tanaka & Haiman, 2013), and a thermal Compton hard X-ray spectrum (d'Ascoli et al., 2018). There are also suggestions that numerous tidal disruption events in the same host galaxy may suggest the presence of an SMBHB (Wegg & Nate Bode, 2011), as well as characteristic drops in the light curve of such an event (Liu et al., 2014). These systems may also induce velocity shifts which may be detectable through careful observations of H_2O megamasers in AGN (Pesce et al., 2018). Finally, it is important to note that many more signatures may arise from the theorized circumbinary disk or mini-disks that may be present within an active SMBHB system, or from magnetized jets that accompany one or both of the black holes (e.g. Bogdanovic et al., 2021).

1.3 Nanohertz Gravitational Waves

In 2015, the window to the gravitational-wave universe was opened when the first detection of GWs from a black hole binary was made by the Laser Interferometer Gravitational-wave Observatory (LIGO) (Abbott et al., 2016). However, as illustrated in Figure 1.5, only a small portion of the GW spectrum can be investigated by ground-based interferometers. While LIGO and similar observatories can probe stellar-mass compact binaries, more efforts are needed to see the GWs emitted by their larger, low-frequency cousins. Space-based interferometers, such as the proposed Laser Interferometer Space Antenna (LISA) mission (Amaro-Seoane et al., 2017), will use much longer interferometric baselines to detect the inspirals of more massive binaries in addition to other longer-period sources such as galactic neutron star and white dwarf binaries. To investigate even lower frequencies in the nanohertz regime, emitted by SMBHBs, GW astronomers have constructed a galaxy-sized GW detector called a pulsar timing array (PTA), which will be be discussed in more detail in subsection 1.3.1 and is the main focus of the work in this thesis.

While each of these detectors work in different ways to sense the respective GWs that they are each sensitive to, the subtle shifts in space-time they are seeking are created with the same basic principles. Commonly analogized to ripples on a still pond, GWs are, essentially, a wave solution to a slight perturbation to the flat space-time described by the Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$. Therefore,

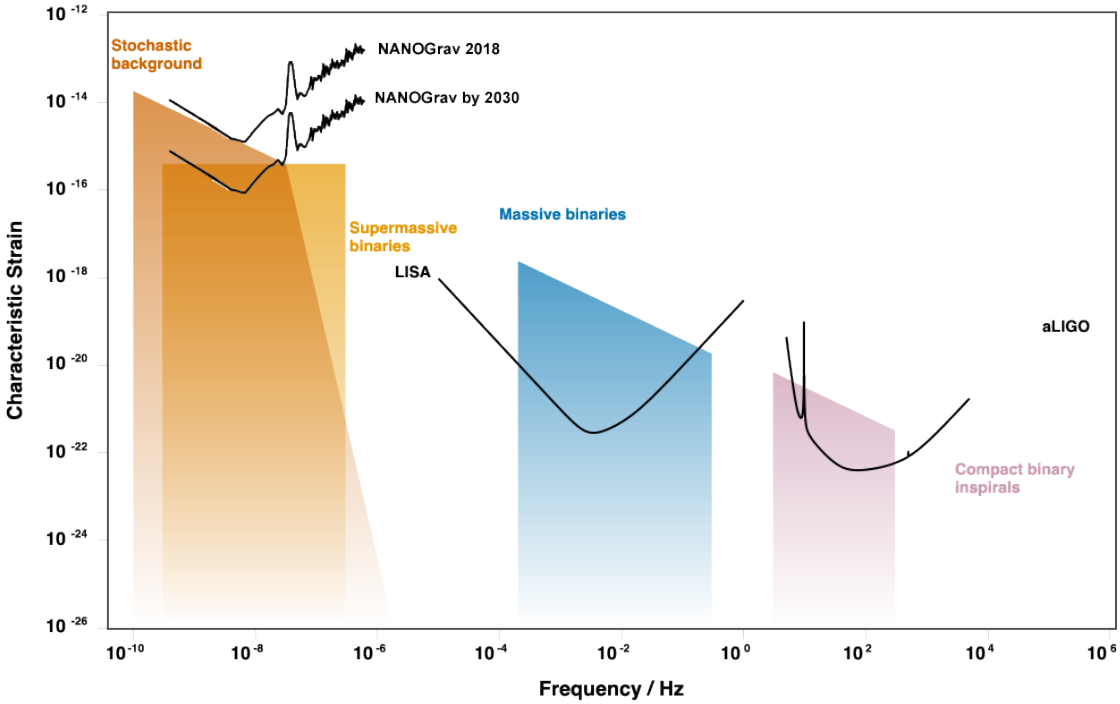

Figure 1.5: The GW spectrum, with example sensitivities of ground-based interferometers, space-based interferometers, and PTAs. Colored regions indicate typical characteristic strains and GW frequency ranges of typical GW sources. Created using `http://gwplotter.com/` (Moore et al., 2015).

the total metric can be written as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{1.4}$$

Gravitational radiation is emitted as the result of a changing gravitational tidal field. Due to conservation of mass and momentum, the existence of such a field requires any system that emits GWs to have a non-zero quadrupole moment, or second time-derivative of moment of inertia tensor. With this requirement, evolving binary systems of compact objects, including neutron stars and black holes, make an ideal example.

GWs are observed through their influence on the massive objects they affect. By distorting space-time, GWs will alter length-scales between test masses. This effect is quantified by the strain, which can be defined as

$$h \simeq \frac{\Delta L}{L}, \tag{1.5}$$

where L is the initial length-scale, and ΔL is the difference between two measurements of L. In General Relativity, GWs are emitted in two polarizations, "plus" and "cross," named for the patterns they create while stretching and squeezing a circular ring of test masses as shown in Figure 1.6. Therefore, the total strain can be written as

$$h = A_+ h_+ + A_\times h_\times, \tag{1.6}$$

where $A_{+,\times}$ represents an amplitude of each polarization state.

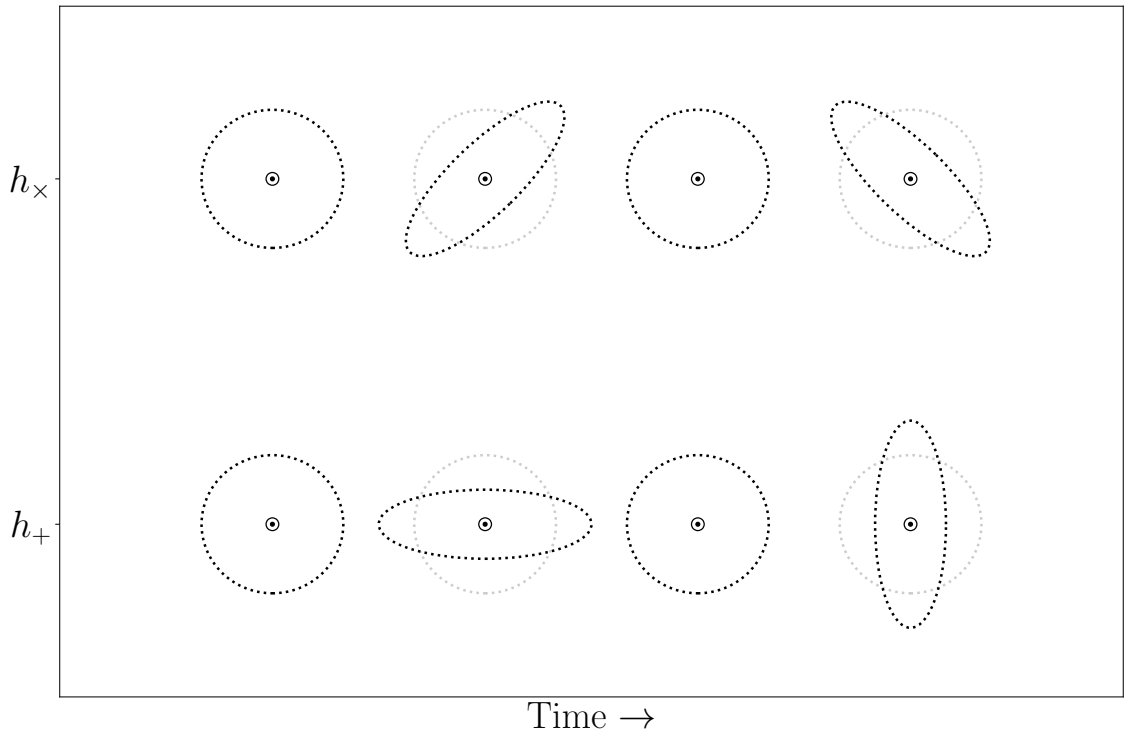

Figure 1.6: As a GW propagates towards the observer out of the page, the induced deformation will be only perpendicular to the direction of propagation. For a ring of particles, the induced deformation over time is shown for both the plus (bottom) and cross (top) polarizations.

In some cases, it is convenient to represent this strain as a characteristic strain h_c that is integrated over some timescale or, for stochastic sources, an integrated power spectral density. For further information on GW theory and how nanohertz GWs are observed, see Maggiore (2007) or Taylor (2021).

1.3.1 Detection Methods

1.3.1.1 Pulsar Timing

Neutron stars, remnants of massive stars with densities comparable to an atomic nucleus, are some of the most extreme objects known to astronomy. Pulsars are a unique class of neutron stars which emit a beam of radiation from the star's magnetic poles which are oriented so as to be observable from the Earth. As the pulsar rotates, these beams of emission sweep across our line of sight, resulting in a narrow "pulse" in high resolution time domain observations. By fitting a pulse template to measure the times-of-arrival (TOAs) of these pulses, astronomers have discovered that these natural objects are some of the most precise clocks in the universe.

Millisecond pulsars (MSPs) are unique in being the most stable rotators, so their TOAs are able to be measured to incredible precision. These "recycled" pulsars have gained angular momentum through the accretion of material from a binary companion, and therefore have spin periods (P) $\lesssim$ 30 ms and period derivatives ($\dot{P}$) of about 10^{-20} s s^{-1} (Manchester et al., 2005). To model the observed TOAs of an MSP, any effects that can delay the TOAs, such as spin period derivatives,

the motion of the pulsar, any binary companions, the Earth, and the remaining principal delays, must be compiled. The principal terms in this process can be summarized as

$$t = t_{\rm t} - t_0 + \Delta_{\rm clock} - \Delta_{\rm DM} + \Delta_{\rm R\odot} + \Delta_{\rm E\odot} + \Delta_{\rm S\odot}, \tag{1.7}$$

where the topocentric time t_t is adjusted to that measured from the solar system barycenter t (Condon & Ransom, 2016; Lorimer & Kramer, 2012). After adjusting to a reference epoch t_0, the remaining delays include a clock correction $\Delta_{\rm clock}$, the dispersion delay caused by the interstellar medium $\Delta_{\rm DM}$, the Roemer delay $\Delta_{\rm R\odot}$ caused by the time necessary to cross the Earth's orbit, the Einstein delay $\Delta_{\rm E\odot}$ accounting for the time dilation induced by the moving pulsar and observatories and any induced gravitational redshifts, and the Shapiro delay $\Delta_{\rm S\odot}$ accounting for the extra travel time for a pulse to travel through curved space-time. These are each calculated for the solar system, and additional terms will be necessary for pulsars in binary systems to describe the same delays induced by a binary companion. The resultant timing model is then compared with observed TOAs to create "residuals" as

$$\rho_i = {\rm TOA}_i - t_i, \tag{1.8}$$

where for each pulse, the residual ρ is equal to the difference between the observed TOA and the modeled value t. A perfect timing model of a noiseless pulsar will result in residual values of zero for all observations of the pulsar as in the top panel

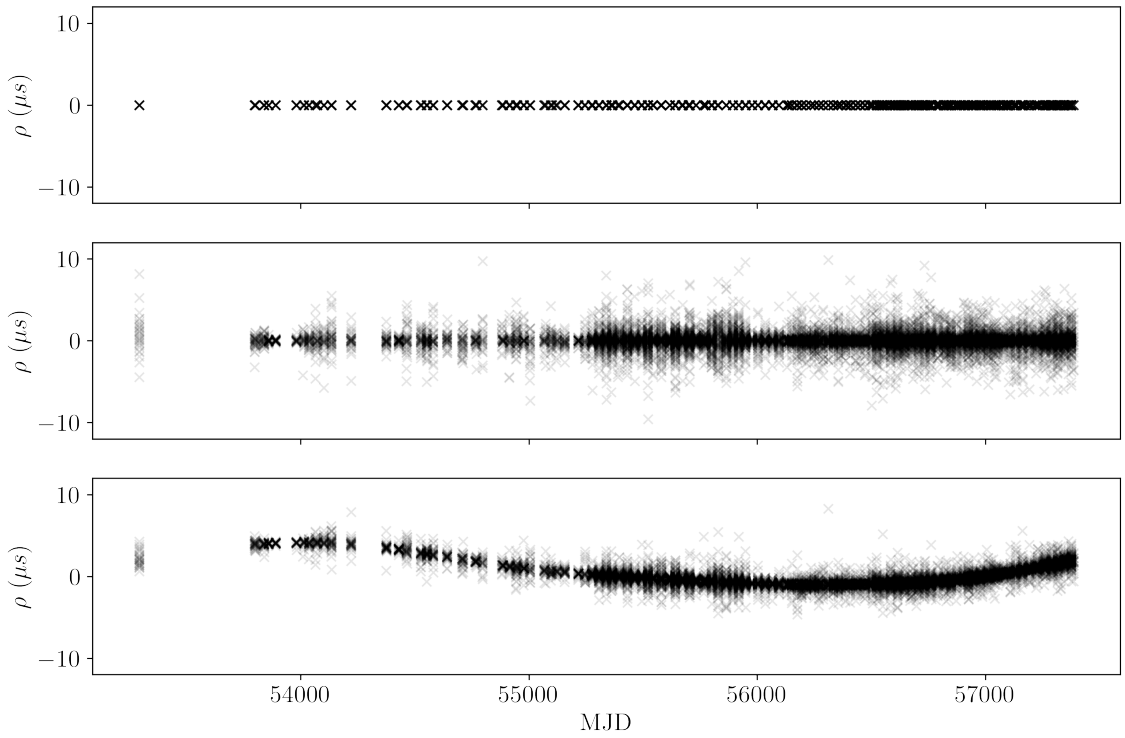

Figure 1.7: Pulsar timing residuals that are idealized (top), contain only white noise (middle), and contain white noise and red noise (bottom) over a ten year observation baseline in Modified Julian Days (MJD).

of Figure 1.7, and for the best MSPs, these values approach nanosecond precision. Any unaccounted-for signals, such as nanohertz GWs, must be teased out from these residuals.

However, even the best-timed pulsars have some amount of noise present in their timing residuals, which can be separated in to two different types: white noise and red noise. Examples of how these two noise processes effect pulsar timing residuals are shown in Figure 1.7. While white noise is uncorrelated, red noise is correlated in time, with a power spectral density modeled with a power-law function as

$$P(f) = \frac{A_{\rm red}^2}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma}. \tag{1.9}$$

Here, $f_{\rm yr} \equiv 1/(1\text{year})$, $A_{\rm red}$ is the red noise amplitude, and γ is a power-law spectral index. White noise is induced by processes including template-fitting TOA uncertainty and pulse jitter intrinsic to the pulsar (Lam et al., 2019). Red noise is induced by multiple processes, including spin noise, pulse profile variations, clock (Tiburzi et al., 2016) and ephemeris (Vallisneri et al., 2020) errors, and unmodeled dispersion measure changes (Jones et al., 2017).

1.3.1.2 Pulsar Timing Arrays

A true GW signal will be present in observations of many pulsars, with distinctive correlations related to the pulsars' locations relative to the Earth and the GW source. Therefore, unambiguously detecting a GW signal requires observations of many pulsars known as a pulsar timing array (PTA) over the span of decades to unambiguously determine whether an emerging signal is due to a GW or an effect local to a single pulsar. The methods used to search for two types of nanohertz GWs visible to PTAs are summarized in subsubsection 1.3.1.3 and subsubsection 1.3.1.4.

The observation strategies used by a PTA will determine the range of GW frequencies to which it is sensitive. The lowest GW frequency a PTA can detect is around $1/T_{obs}$, where T_{obs} is the total time spanned by all of the observations. The highest frequency which can be detected is equal to $1/t_{obs}$, where t_{obs} is the average observation cadence, or time between observations (Aggarwal et al., 2019).

Currently, there exist multiple distinct PTA collaborations; these include the North American Nanohertz Observatory for Gravitational waves (NANOGrav), the

European Pulsar Timing Array (EPTA), the Parkes Pulsar Timing Array (PPTA), the Indian Pulsar Timing Array (InPTA), and the Chinese Pulsar Timing Array (CPTA). Due to the restrictions on which pulsars each of these regional PTAs can observe (due to the portions of the sky visible to each continent), it is extremely beneficial for PTAs to make use of data from multiple telescopes. Therefore, these collaborations have joined together in a consortium known as the International Pulsar Timing Array (IPTA), which, through regular data combinations (Perera et al., 2019), will reach greater sensitivities than any lone PTA.

1.3.1.3 Continuous Gravitational Waves

As described in subsection 1.1.2, SMBHBs evolve very slowly over their long lifetimes, and therefore, nanohertz GWs emitted by individual SMBHBs can be considered "continuous waves" (CWs) (Aggarwal et al., 2019; Arzoumanian et al., 2014; Ellis, 2013). These signals can be uniquely identified by eight of the following nine properties of the binary system, including:

- position on the sky (θ, ϕ);
- GW frequency (f_{GW});
- orbital phase at some reference time (Φ_0);
- GW polarization angle (ψ);
- orbital inclination (i);
- chirp mass $(\mathcal{M})$;
- luminosity distance (d_{L});

- strain amplitude (h),

some of which were defined in subsection 1.1.2. The chirp mass is a quantity similar to a reduced mass that, as will be apparent later in this section, a convenient form to relate various GW amplitudes and amounts of energy loss. It can be defined as

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} = \left[\frac{q}{(1+q)^2}\right]^{3/5} M_{\rm tot}, \tag{1.10}$$

where m_1 and m_2 are the masses of the two black holes, and q is the mass ratio of the system. This is again the redshifted value in the frame of the observer; in the rest frame of the binary, $\mathcal{M}_r = \mathcal{M}/(1+z)$. However, as PTAs are currently only sensitive to the local universe where $z \lesssim 1$, this adjustment has only a minor effect. The GW frequency f_{GW} of a circular SMBHB can be related to the orbital period $P_{\rm orb}$ by $f_{GW} = 2/P_{\rm orb}$. A careful reader will notice that while eight parameters are required to define a CW signal, nine are listed above. This is because one is redundant; the strain amplitude of a CW can be related to the SMBHB chirp mass, GW frequency, and luminosity distance as

$$h_0 = \frac{2\mathcal{M}^{5/3}(\pi f_{\rm GW})^{2/3}}{d_L}. \tag{1.11}$$

The quadrupolar nature of GW emission creates a signal that is not equally strong in all directions. Therefore, a CW will induce a larger response in the residuals of some pulsars than others. This effect is described by the antenna pattern response functions of a detector, $F^{+,\times}$, which are determined by the sky locations of the

SMHHB and the pulsar as well as the GW polarization angle ψ (Taylor et al., 2016). First, we construct a set of right-handed basis vectors

$$\begin{aligned}
\hat{n} =&(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta),\\
\hat{p} =&(\cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi,\\
&\cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi, -\cos\psi\sin\theta),\\
\hat{q} =&(\sin\psi\cos\theta\cos\phi + \cos\psi\sin\phi,\\
&\sin\psi\cos\theta\sin\phi - \cos\psi\cos\phi, -\sin\psi\sin\theta)
\end{aligned} \tag{1.12}$$

including $\hat{n}$, which is the opposite of the vector pointing to the SMBHB sky location, $\hat{\Omega}$. Then, we can define

$$F^A(\hat{\Omega}) \equiv \frac{1}{2}\frac{\hat{u}^a\hat{u}^b}{1+\hat{\Omega}\cdot\hat{u}}e^A_{ab}(\hat{\Omega}), \tag{1.13}$$

where $e^{+,\times}_{ab}$ are the GW basis tensors

$$\begin{aligned}
e^+_{ab} &= \hat{p}_a\hat{p}_b - \hat{q}_a\hat{q}_b\\
e^\times_{ab} &= \hat{p}_a\hat{q}_b + \hat{q}_a\hat{p}_b.
\end{aligned} \tag{1.14}$$

An example set of antenna pattern response functions for an SMBHB located at RA 12h Dec +45d is shown in the first two panels of Figure 1.8. While many pulsars will be sensitive to a CW emanating from a source located anywhere on the sky, others may lie in a region where the induced signal is near-zero due to the antenna pattern response function.

For SMBHBs with quasi-circular orbits, CWs will be emitted at a GW fre-

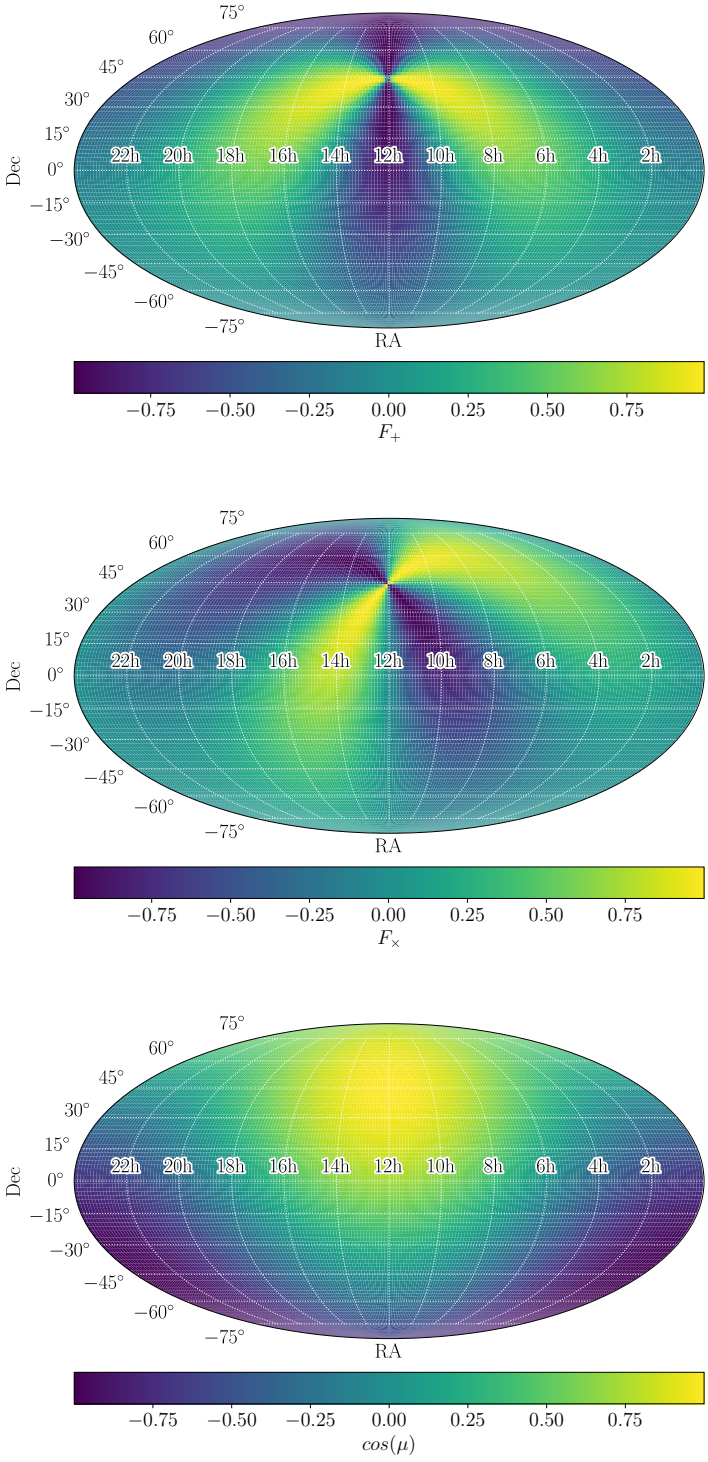

Figure 1.8: Antenna pattern response functions F_+ (top) and $F_\times$ (middle) for an SMBHB located at RA 12h Dec +45d. The cosine of the angle between the source sky location and the pulsar line of sight (μ) is used to calculate the pulsar time t_p (bottom).

quency corresponding to the orbital frequency of the binary (to be exact, $f_{GW} = 2f_{orb}$) which may slowly evolve over time as the binary loses energy. This evolution is described by

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{96}{5}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/3}\omega(t)^{11/3}, \tag{1.15}$$

where $\omega(t) = \pi f_{GW}(t)$. This can be integrated to yield

$$f(\tau) = \frac{1}{\pi}\left(\frac{G\mathcal{M}}{c^3}\right)^{-5/8}\left(\frac{5}{256}\frac{1}{\tau}\right)^{3/8}, \tag{1.16}$$

where τ is the time until the binary's coalescence (Hazboun et al., 2019b). Therefore, we can estimate the magnitude of the binary's frequency evolution over the course of a given timescale T to be

$$\Delta f \simeq \frac{1}{\pi}\left(\frac{G\mathcal{M}}{c^3}\right)^{-5/8}\left(\frac{5}{256}\frac{1}{\tau}\right)^{3/8}\frac{3}{8}\frac{T}{\tau}. \tag{1.17}$$

For a given PTA experiment which spans approximately a decade, and a typical SMBHB that would emit GWs within the PTA band, this change in frequency is much smaller than the frequency resolution of a PTA. Therefore, for most searches, GWs from a circular SMBHB can be considered to be emitted at a single discrete frequency.

However, the binary evolution is critical to consider for some aspects of GW searches with PTAs. Unlike a "small" detector such as an interferometer located on the Earth, the significant light travel times between the Earth and the numerous pulsars in the PTA result in a large difference between the form of a CW at the

time a wavefront passes either a pulsar or the Earth. This results in two distinct, coupled terms being present in the signal a CW is modelled to induce in each pulsar's residuals, helpfully dubbed the "Earth term" and "pulsar term" (Aggarwal et al., 2019) that each describe the effect of the CW at the location of the Earth and the pulsar in space, respectively.

If the locations of both the Earth and the pulsar are known, the pulsar term can be modeled by geometrically calculating the point in time at which the pulsar is "seeing" the binary as

$$t_p = t - L(1 + \hat{\Omega} \cdot \hat{p}), \tag{1.18}$$

where L is the distance to the pulsar. This time is then used to relate the GW frequency of the pulsar term to that of the Earth term. Since there is a longer light travel time associated with the pulsar term, this frequency is always lower than the Earth term's. However, pulsar distances are difficult to measure, and instead can be calculated through various methods, including measurements of timing parallaxes and estimates from dispersion measure models. However, due to the uncertainties associated with pulsar distance measurements, which are often much larger than a GW wavelength, the pulsar distances must be used as a free parameter in any CW search that includes the pulsar term. Improved measurements of pulsar distances with longer timing baselines and more accurate techniques, such as VLBI, will dramatically shrink these prior ranges, as discussed in chapter 4. For more discussion of this point, see section 1.4 and chapter 4.

Combining these effects, the signal induced into a pulsar's residuals by an

SMBHB can be written as

$$s(t,\hat{\Omega}) = F^{+}(\hat{\Omega})\Delta s_{+}(t) + F^{\times}(\hat{\Omega})\Delta s_{\times}(t), \tag{1.19}$$

where

$$\Delta s_{+,\times}(t) = s_{+,\times}\left(t_p\right) - s_{+,\times}(t). \tag{1.20}$$

The GW signal $s_{+,\times}(t)$ for a given time t can be written as

$$\begin{aligned} s_{+}(t) =& \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \left[-\sin 2\Phi(t)\left(1+\cos^2 i\right)\right] \\ s_{\times}(t) =& \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \left[-2\cos 2\Phi(t)\cos i\right] \end{aligned}, \tag{1.21}$$

with the time dependent phase and frequency defined as

$$\begin{aligned} \Phi(t) &= \Phi_0 + \tfrac{1}{32}\mathcal{M}^{-5/3}\left[\omega_0^{-5/3} - \omega(t)^{-5/3}\right] \\ \omega(t) &= \omega_0\left(1 - \tfrac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}t\right)^{-3/8} \end{aligned}, \tag{1.22}$$

respectively. Figure 1.9 shows an example of a CW waveform in two pulsars in different locations. The top panel shows only the Earth term of the signal, with the different signal amplitudes related to the antenna pattern response functions calculated for these SMBHB-pulsar pairs. The lower panel shows the signal with both the Earth and pulsar terms, with both distinct frequencies visible.

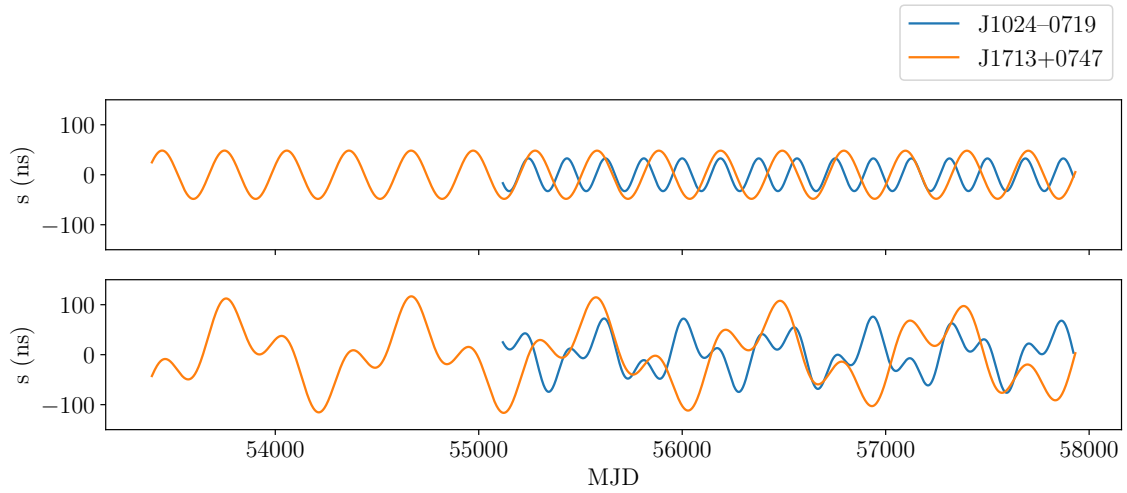

Figure 1.9: Simulated CW waveforms in two pulsars for an SMBHB located at 1200+4500, with $\mathcal{M} = 1.5 \times 10^9 \mathrm{M}_\odot$, $d_L = 85$ Mpc, and $f_{GW} = 60$ nHz including only the Earth term (top) and the Earth and pulsar terms (bottom).

1.3.1.4 The Stochastic Gravitational Wave Background

While only the brightest GWs emitted by SMBHBs could be detected as CWs, there is an entire population of SMBHBs hosted in galaxies throughout the universe constantly emitting GWs too faint to be detected individually. However, these SMBHBs together create a pervasive hum of GW emission known as the stochastic GW background (GWB).

The GWB will appear as a spatially-correlated common red noise (CRN) process in all of the pulsars in the PTA with a specific spectral shape (Romano & Cornish, 2017), with higher characteristic strains (h_c) at lower frequencies. This is often parameterized as

$$h_c(f) = A_{\mathrm{GWB}} \left(\frac{f}{f_{\mathrm{yr}}} \right)^{\alpha}, \tag{1.23}$$

where A_{GWB} is the amplitude of the GWB measured at f_{yr}, and α is the spectral index. A population of inspiraling SMBHBs, with many sources emitting at low

frequencies but stronger emission at high frequencies, will have an $\alpha = -2/3$. Unlike red noise intrinsic to each pulsar, which may each have a unique spectral index, the GWB spectral shape will be the same in each pulsar in the PTA.

In a PTA search for the GWB, this information is transformed into the timing-residual cross-power spectral density

$$S_{ab}(f) = \Gamma_{ab} \frac{A^2_{\rm GWB}}{12\pi^2} \left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} f_{\rm yr}^{-3}, \tag{1.24}$$

where $\gamma = 3 - 2\alpha$, and therefore, for a nominal $\alpha = -2/3$ (Rajagopal & Romani, 1995) for a GWB produced by SMBHBs, we expect $\gamma = 13/3$. Γ_{ab} is an overlap reduction function that describes the degree of correlation expected between two pulsars in the array, which varies according to the angle between their locations on the sky, θ_{ab}. This term also differentiates this function, $S_{ab}(f)$, from the pulsar red noise power spectral density function in Equation 1.9. For an isotropic GWB, Γ_{ab} is given by Hellings & Downs (1983) as the "Hellings-Down Curve"

$$\Gamma_{ab} = \frac{3}{2}\frac{(1-\cos(\theta_{ab}))}{2}\ln\left(\frac{(1-\cos(\theta_{ab}))}{2}\right) - \frac{1}{4}\frac{(1-\cos(\theta_{ab}))}{2} + \frac{1}{2} + \frac{1}{2}\delta_{ab}, \tag{1.25}$$

and is plotted in Figure 1.10. Note that for small separations, Γ_{ab} will approach a value of 1/2. This is due to the fact that the correlation induced by the GWB is present because of the Earth term of the signal; while the GWB affects the pulsars as well, those signals will not be correlated according to their position relative to one another. However, as a GWB signal in one pulsar will inherently be correlated

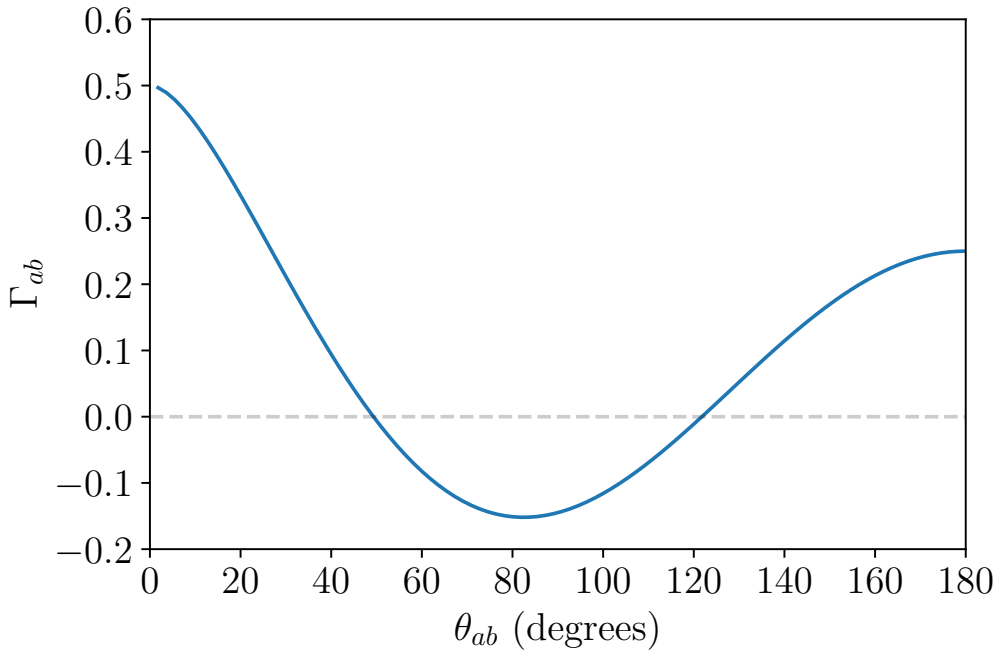

Figure 1.10: Overlap reduction function Γ_{ab} for Hellings-Downs correlations plotted as a function of the angle between pulsars a and b (θ_{ab}).

with itself, the Kroneker delta δ_{ab} term accounts for the possibility that two pulsars at the same sky location are in fact the same object, in which case $\Gamma_{ab} = 1$. For a full description, see Burke-Spolaor et al. (2019) and Taylor (2021), as well as Arzoumanian et al. (2018a) and Arzoumanian et al. (2020a) for recent applications.

While this thesis focuses primarily on the detection of CWs and their electromagnetic counterparts, the presence of the GWB must be a consideration for all types of GW searches by PTAs. As PTAs start to approach the sensitivities required to detect the GWB, the particular realization of the GWB that exists in the universe may appear similar to a CW in the discrete data sets. Much like a periodicity needs careful treatment to be differentiated from a DRW process in an AGN light curve as described in subsection 1.2.4, care is needed to ensure that a GWB, which is expected to be the first detection made by PTAs, is not falsely identified as a CW

(Baker et al., 2019).

1.4 Computational Methods and Bayesian Inference

While the data used to search for SMBHBs using electromagnetic signatures of GW emission varies widely, the computational techniques used for each type of project in this thesis are remarkably similar. As described thoroughly in Zhu & Thrane (2020), both PTA searches for CWs and SMBHB searches in quasar light curves fundamentally aim to determine whether a periodicity exists within an unevenly-sampled time series that contains red noise. Therefore, similar methods are easily applied to both types of analysis.

To investigate this possibility, we make use of Bayesian statistics, which, rather than questioning "how likely it is that we observe data $\mathbf{D}$ given our theorized model M," investigates the probability that the underlying parameters $\mathbf{\Theta}_M$ are equal to a value given the observed data. These various probabilities can be described with Bayes Theorem

$$P\left(\mathbf{\Theta}_M \mid \mathbf{D}, M\right) = \frac{P\left(\mathbf{D} \mid \mathbf{\Theta}_M, M\right) P\left(\mathbf{\Theta}_M \mid M\right)}{P(\mathbf{D} \mid M)}, \tag{1.26}$$

with the following components:

1. the likelihood $P\left(\mathbf{D} \mid \mathbf{\Theta}_M, M\right)$, which describes the probability of both having a set of parameters $\mathbf{\Theta}_M$ that describe the data and observing the particular set of data $\mathbf{D}$

2. the prior, $P(\boldsymbol{\Theta}_M \mid M)$, which describes the probability of having a given set of values $\boldsymbol{\Theta}_M$ for our model M, without knowledge of the data $\mathbf{D}$

3. the evidence, $P(\mathbf{D} \mid M) = \mathcal{Z}$, which quantifies how well the model M fits the data $\mathbf{D}$ after integrating over all possible values of the parameters $\boldsymbol{\theta}_M$

4. the posterior, $P(\boldsymbol{\Theta}_M \mid \mathbf{D}, M)$, which describes the probability of having a given set of values $\boldsymbol{\Theta}_M$ for our model M given the data $\mathbf{D}$.

Throughout this work, we make use of Markov-Chain Monte Carlo (MCMC) analyses, which use random number draws to investigate many-dimensional parameter spaces. After randomly sampling the priors of each model parameter, the algorithm will then calculate the likelihood of the chosen model parameters. The sampler will then take a step by drawing new values for each parameter and again calculating the likelihood for the new set, and the difference in the value of the likelihoods for the two steps will determine if the later step is accepted or rejected. After continuing this process for many steps to create a chain of samples, the accumulated samples made up of accepted steps will equal the posterior distribution for each model parameter if the chain has converged.

Perfect convergence, in theory, could require an infinite number of MCMC iterations; however, numerous methods, such as computing effective independent samples and calculating autocorrelation lengths of MCMC chains, can be used to determine sufficient levels of convergence. These tests are particularly critical for analyses with many parameters and models with highly complex likelihood surfaces, such as PTA searches for CWs. For further discussion of Bayesian analyses using

MCMCS, see Speagle (2019), and for specific applications, see Aggarwal et al. (2019) and chapter 2 of this thesis.

The resulting posterior distributions can then be used to estimate the preferred values and uncertainties of the model parameters. If these resulting credible regions are one-sided, upper limits can be calculated for any parameter in the model by integrating the posterior distribution until 95% of it is enclosed.

Numerous methods exist to perform a Bayesian model selection as conducting such an analysis. First among these is the Bayes factor $\mathcal{B}_{12} = \mathcal{Z}_1/\mathcal{Z}_2$, which, given the data and strength of the model, tells us how much a given model is preferred over the other. As the full Bayesian evidence is difficult to calculate, Bayes factors are often estimated. One such technique uses the Savage-Dickey approximation (Dickey, 1971) to calculate Bayes factors for nested models by comparing the prior to the posterior distribution at the relevant parameter value. For example, for a CW search, a recovered CW strain value consistent with zero would be identical to a model with only noise and no CW. In this case,

$$\mathcal{B}_{10} \equiv \frac{\mathcal{Z}_1}{\mathcal{Z}_0} = \frac{p(h_0 = 0|M_1)}{p(h_0 = 0|\mathcal{D}, M_1)}. \tag{1.27}$$

The strength of model preference given by a Bayes factor can be summarized according to the rules of thumb in Table 1.1 (Taylor, 2021; Jeffreys, 1939).

For a fuller comparison, a metric such as the Bayes Information Criterion (BIC) may be used. After performing an MCMC analysis for each model being compared, the BIC adjusts the maximized log-likelihood values by the number of

Bayes factor, $\mathcal{B}$	$\ln(\mathcal{B})$	Strength of evidence
$< 1:1$	< 0	Negative (supports M_0)
$1:1$ to $3:1$	$0 - 1.1$	Barely worth mentioning
$3:1$ to $10:1$	$1.1 - 2.3$	Substantial
$10:1$ to $30:1$	$2.3 - 3.4$	Strong
$30:1$ to $100:1$	$3.4 - 4.6$	Very strong
$> 100:1$	> 4.6	Decisive

Table 1.1: Bayes factors can be interpreted using loose rules of thumb which indicate the strength of the preference for one model over another in a Bayesian model selection.

parameters in the model as

$$\mathrm{BIC} = k \ln(n) - 2 \ln(\widehat{L}), \tag{1.28}$$

where k is equal to the number of free parameters, n is equal to the number of data points in the light curve, and $\widehat{L}$ is the maximum likelihood value (Liddle, 2007).

$$\Delta\mathrm{BIC} = \mathrm{BIC}_0 - \mathrm{BIC}_1. \tag{1.29}$$

A lower value of ΔBIC indicates more support for the hypothesized model M_1 over M_0. The strength of model preference given by a ΔBIC can be summarized according to the rules of thumb in Table 1.2 (Kass & Raftery, 1995).

ΔBIC	Strength of evidence
> 0	Negative (supports M_0)
0 to -2	Barely worth mentioning
-2 to -6	Substantial
< -6	Strong

Table 1.2: ΔBIC values can be interpreted using loose rules of thumb which indicate the strength of the preference for one model over another in a Bayesian model selection.

For many PTA GW analyses, the python package `enterprise`[1] (Ellis et al., 2019) is used to construct Bayesian priors and calculate the PTA likelihood. For all chapters in this thesis, the python package `PTMCMCSampler`[2] (Ellis & van Haasteren, 2017) is used for all MCMC analyses, due to the easy construction of custom sampling techniques to explore high-dimensional parameter spaces.

1.5 Multi-Messenger Astrophysics

Unlike the effects described in section 1.2, as well as the vast majority of astronomical observations, GWs are not carried by photons. Instead, they are their own unique "messenger," which, through extensive observations, observers can use to learn about the universe entirely independently from light.

However, while these two messengers can be used to observe the same objects independently, this does not imply that they *should*. By combining data to conduct multi-messenger observations, astronomers have the opportunity to learn more about the universe than with either messenger alone. For example, the first multi-messenger analysis of a GW source was a detection of merging neutron stars by LIGO and the resulting kilonova by numerous electromagnetic telescopes worldwide (Abbott et al., 2017). While both observations were spectacular in their own right, the combination provided an unprecedented confirmation of numerous theories of kilonova emission mechanisms and heavy metal nucleosynthesis. This result was made possible by by an unparalleled level of cooperation between worldwide tele-

[1] https://github.com/nanograv/enterprise
[2] https://github.com/jellis18/PTMCMCSampler

scope operations to create a target-of-opportunity observation scheme which makes rapid follow up of LIGO alerts possible.

Low-frequency GWs, including those observable by PTAs, have yet to be detected, with a multi-messenger detection of an SMBHB being a key goal for the field in the coming years (Astro2020). However, as described above, GWs from SMBHBs are not short-duration events like the chirping mergers detectable by LIGO. Instead of a rapid dash to coordinate observations of an electromagnetic counterpart of a brief, cataclysmic event, these observations rely on vast efforts regularly undertaken over years to decades.

Significant efforts already have been undertaken to prepare for multi-messenger analyses of SMBHBs. In chapter 2 of this work, it is shown that multi-messenger methods will lead to improved SMBHB mass limitations, even without a GW detection. Simulations have also shown that targeting a specific SMBHB candidate in a PTA search leads to vastly improved parameter estimation results (Liu & Vigeland, 2021).

As previously mentioned, one of the best possibilities for a nanohertz multi-messenger detection is through the collaboration of PTAs and photometric time-domain surveys of quasars due to the significant overlap in relevant source populations. Numerous candidates have been found electromagnetically (Graham et al., 2015b), and have been shown to be excellent PTA targets in the coming years (Xin et al., 2020b). With the advance of next generation surveys with high sensitivities, long durations, and fast cadences, more are sure to follow, so long as efficient data processing methods can be developed to identify the most likely candidates.

In addition to the increased sensitivity multi-messenger techniques can bring to GW analyses, ongoing multi-messenger observations of an SMBHB will provide insights that are impossible with one messenger alone. These include the nature of SMBHBs' interactions with their environments and host galaxies, as well their co-evolution (Shankar et al., 2016). These observations will also closely study accretion in the presence of an SMBHB, and test models of induced variability and other SMBHB identification methods. Finally, independent distance measurements from electromagnetic and GW observations can turn an SMBHB into a standard siren to probe cosmological models (Holz & Hughes, 2005). For a detailed overview, see Kelley et al. (2019a).

Chapter 2

Multi-Messenger Gravitational-wave Searches with Pulsar Timing Arrays: Application to 3C66B Using the NANOGrav 11-year Data Set

2.1 Abstract

When galaxies merge, the supermassive black holes in their centers may form binaries and, during the process of merger, emit low-frequency gravitational radiation in the process. In this paper we consider the galaxy 3C 66B, which was used as the target of the first multi-messenger search for gravitational waves. Due to the observed periodicities present in the photometric and astrometric data of the source, it has been theorized to contain a supermassive black hole binary. Its apparent 1.05-year orbital period would place the gravitational wave emission directly

Published in ApJ as Arzoumanian et al. (2020b).

Contributing authors: Zaven Arzoumanian, Paul T. Baker, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, Bence Bécsy, Maria Charisi, Shami Chatterjee, James M. Cordes, Neil J. Cornish, Fronefield Crawford, H. Thankful Cromartie, Kathryn Crowter, Megan E. DeCesar, Paul B. Demorest, Timothy Dolch, Rodney D. Elliott, Justin A. Ellis, Robert D. Ferdman, Elizabeth C. Ferrara, Emmanuel Fonseca, Nathan Garver-Daniels, Peter A. Gentile, Deborah C. Good, Jeffrey S. Hazboun, Kristina Islo, Ross J. Jennings, Megan L. Jones, Andrew R. Kaiser, David L. Kaplan, Luke Zoltan Kelley, Joey Shapiro Key, Michael T. Lam, T. Joseph W. Lazio, Lina Levin, Jing Luo, Ryan S. Lynch, Dustin R. Madison, Maura A. McLaughlin, Chiara M. F. Mingarelli, Cherry Ng, David J. Nice, Timothy T. Pennucci, Nihan S. Pol, Scott M. Ransom, Paul S. Ray, Brent J. Shapiro-Albert, Xavier Siemens, Joseph Simon, Renée Spiewak, Ingrid H. Stairs, Daniel R. Stinebring, Kevin Stovall, Joseph K. Swiggum, Stephen R. Taylor, Michele Vallisneri, Sarah J. Vigeland, Weiwei Zhu

in the pulsar timing band. Since the first pulsar timing array study of 3C 66B, revised models of the source have been published, and timing array sensitivities and techniques have improved dramatically. With these advances, we further constrain the chirp mass of the potential supermassive black hole binary in 3C 66B to less than $(1.65 \pm 0.02) \times 10^9\ M_\odot$ using data from the NANOGrav 11-year data set. This upper limit provides a factor of 1.6 improvement over previous limits, and a factor of 4.3 over the first search done. Nevertheless, the most recent orbital model for the source is still consistent with our limit from pulsar timing array data. In addition, we are able to quantify the improvement made by the inclusion of source properties gleaned from electromagnetic data over 'blind' pulsar timing array searches. With these methods, it is apparent that it is not necessary to obtain exact a priori knowledge of the period of a binary to gain meaningful astrophysical inferences.

2.2 Introduction

Continuous gravitational waves (GWs), defined by single-source cyclic GW emission, are expected to arise from the supermassive black hole binaries (SMBHBs) that form during a galaxy merger. When a SMBHB evolves such that it emits GWs in the microhertz to nanohertz GW band (orbital periods of weeks to several decades), a sufficiently massive and/or nearby SMBHB may be detectable by pulsar timing arrays (PTAs; e.g., Aggarwal et al. 2019) (hereafter A19).

While GWs from individual sources in the PTA regime have been sought after in multiple works (Arzoumanian et al., 2014; Feng et al., 2019; Jenet et al.,

2004) through a variety of methods, none have been detected. However, numerous advances have been made in the field of pulsar timing. As PTA experiments gain longer time baselines and higher cadences and the numbers of millisecond pulsars grows, sensitivity to GW sources increases. The notable ongoing PTA programs in the world include the European PTA, Parkes PTA, and the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) (e. g. Desvignes et al., 2016; Manchester et al., 2013; Arzoumanian et al., 2018b, respectively). Altogether, these PTAs time approximately 100 pulsars to high precision with the goal of GW detection, among other endeavors (e. g. Caballero et al., 2018; Hobbs et al., 2020).

In addition, sophisticated GW detection methods have been developed to detect quadrupolar continuous-wave signals in the data of coordinated timing arrays (e. g. Babak et al., 2016; Zhu et al., 2014; Aggarwal et al., 2019). However, past analyses that used the most up-to-date methods have used 'blind' detection methods; that is, the software did not consider any binary model information gained from electromagnetic data to directly benefit the search. In comparison, most works that do target limits on specific sources using electromagnetic information have used smaller data sets consisting of a single pulsar with as a periodogram approach (Jenet et al., 2004; Feng et al., 2019), or have been derived from the stochastic GW background (Zhu et al., 2019) rather than the full GW analysis pipeline. Here, we have combined these methods in the first search of this type, where we used the entire NANOGrav array of pulsars and full GW search analysis, while incorporating electromagnetic data to conduct a more informed search for GWs from our test source, 3C 66B.

Since the report of a hypothesized orbital motion in the core of the galaxy

3C 66B by Sudou et al. (2003) (hereafter S03), it has been an ideal test case for searches for GWs from SMBHBs. Using long-baseline interferometry, the authors found apparent elliptical motions in 3C66B's radio core, modeling this motion as the gyration of the jet nozzle due to an orbit-induced precession of the smaller black hole's jet. S03 proposed a period and chirp mass for the binary of 1.05 ± 0.03 years and $1.3 \times 10^{10}\ M_{\odot}$, respectively. Given the relatively small redshift of the galaxy ($z = 0.02126$), a binary with those properties would be emitting gravitational radiation well within the sensitivity of pulsar timing arrays (PTAs).

As such, 3C 66B has long been a prime candidate for continuous GW detection. It was the first object targeted for continuous wave detection, as reported by Jenet et al. (2004) (hereafter J04), in which seven years of Arecibo timing data from PSR B1855+09 (Kaspi et al., 1994) was used to search the Fourier domain timing residuals (commonly referred to as a Lomb-Scargle periodogram), using harmonic summing (Press et al., 1992), for a GW signal consistent with the binary period modeled by S03. With these methods, they did not see evidence of a significant signal, and were able to place an upper limit of $7 \times 10^{9}\ M_{\odot}$ on the chirp mass of the system at a binary eccentricity of $e = 0$.

Since the study of J04, Iguchi et al. (2010) (hereafter I10), reported a 93-day variability in the active galactic nucleus's millimeter light that was interpreted as likely due to doppler boosting of a relativistic outflow that is modulated by orbital motion (its period differs due to geometric effects). The new model assumed the 1.05-year orbital period from S03, but predicted an updated chirp mass of $7.9 \times 10^{8}\ M_{\odot}$, almost a full order of magnitude lower than the upper limit set by

J04.

3C 66B was also one of the objects targeted by Zhu et al. (2019), which used a novel approach to test 3C66B indirectly by using the source to predict the GW background strength implied by this source's existence. They concluded that the I10 model produced GW backgrounds that were larger than are currently probed by pulsar timing arrays, implying that the source was not likely to be a binary with parameters as proposed by I10.

The work reported here presents a Bayesian cross-validation framework in which we use 3C66B's binary parameter measurements as priors for our continuous wave search. Our search has resulted in the most stringent direct GW derived limit to date on the chirp mass of 3C66B's candidate SMBHB. We also test, more generically within our search framework, what sensitivity improvements can be gained by knowing the GW frequency of a target to increasingly good precision.

Therefore, we have quantified the improvement made by searching for GWs from a specific source, including cases where the orbital period is only known with large error or not known at all.

Note that because J04 used only one pulsar in their study, they would have been unable to perform a formal experiment to detect 3C 66B, as the use of one pulsar precludes the ability to demonstrate the quadrupolar signature that is unique to the influence of gravitational waves. Thus, our study here is the first formal targeted detection experiment for 3C 66B using a pulsar timing array.

This paper is laid out as follows: in section 2.3, we describe our data, mathematical model, and software pipeline. In section 2.4 we report the detection Bayes

factor and chirp mass upper limit for 3C 66B, as well as results for new test methods. In section 2.5 and section 2.6, we present our conclusions as well as discuss implications for future detection prospects of this and other SMBHBs.

2.3 Analysis Methods

2.3.1 Pulsar Timing and Electromagnetic Data

We make use of the NANOGrav 11-year Data Set (Arzoumanian et al., 2018b), which provides high precision timing of 45 millisecond pulsars. Only the 34 pulsars with baselines of at least 3 years are used for GW detection analyses (Arzoumanian et al., 2018a). We describe slight differences in the use of the data set in this work as compared to other papers in section 2.3. However, the majority of the data are treated similarly to A19. Due to the 11-year timing baseline, the data set is most sensitive to binaries with orbital periods of less than a decade.

The electromagnetic data we incorporate into our models are mainly derived from S03 and I10, as well as the location from the NASA/IPAC Extragalactic Database (NED)[1]. These values are summarized in Table Table 2.1. The right ascension, declination, and luminosity distance are taken as constants throughout the analysis, as the PTA sensitivity to sky location and distance is much lower than any associated errors. For consistency with earlier work, we take the luminosity distance of 3C 66B to be 85 Mpc, as in S03. Therefore, all calculations use

[1]The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration, and can be accessed at https://ned.ipac.caltech.edu/

Table 2.1: GW Model Values and Uncertainties

Parameter	Value	Reference
Chirp Mass ($\mathcal{M}$)	$7.9^{+3.8}_{-4.5} \times 10^8 M_\odot$	I10
GW frequency (f_{GW})	60.4 ± 1.73 nHz	S03
Redshift (z)	0.02126	Huchra et al. (1999)
RA	02h 23m 11.4112s	Fey et al. (2004)
Dec	+42d 59m 31.384s	Fey et al. (2004)
GW strain (h)	$7.3^{+6.8}_{-5.8} \times 10^{-15}$	S03; I10

$H_0 = 75$ km s^{-1} Mpc^{-1}. Note that minor differences in the distance due to different reports of redshift or H_0 cause only a small fractional variation in the results. If the fractional change in the luminosity distance is defined as

$$d_{85} \equiv \left(\frac{d_L}{85 \text{ Mpc}} \right), \tag{2.1}$$

any GW strain limit can be converted to the reader's preferred distance by multiplying the strain by d_{85}, and $\mathcal{M}$ limits by multiplying by $d_{85}^{3/5}$.

2.3.2 Signal Model

We use the methods presented in A19 for the generation of expected pulsar timing residuals influenced by a signal from a continuous GW from a circular SMBHB. While we will not present the full derivation, we will summarize below the relevant equations needed to follow our analysis on the NANOGrav data and refer the reader to A19 for more detail. Note that throughout this section, equations are written in natural units (where $G = c = 1$).

Pulsar timing residuals describe the deviation of an observed pulse arrival time from that predicted from a model based on spin, astrometric, interstellar delay, and if needed, binary parameters of the pulsar. These are the basic data product that we use to search for GWs, which will not be included in the pulsar's timing model. A vector of timing residuals (δt) that is fit without a GW for each pulsar is modeled as

$$\delta t = M\epsilon + n_{\rm white} + n_{\rm red} + s \ , \tag{2.2}$$

where M is the design matrix, which describes the timing model, and ϵ is a vector of the linearized timing model parameter offsets from the best fit solution. In other words, the timing model, which was originally derived without the presence of a GW, must now be adjusted. We write a vector describing the white noise in the data as $n_{\rm white}$, and the same for the red noise, $n_{\rm red}$, which is correlated over long timescales. The noise terms are described in more detail in subsection 2.3.3.

The signal s can be derived as follows. For a GW source whose sky location is described by polar and azimuthal angles θ and ϕ, the strain induced by the emitted GWs is written in terms of two polarizations as

$$h_{ab}(t,\hat{\Omega}) = e^{+}_{ab}(\hat{\Omega})h_{+}(t,\hat{\Omega}) + e^{\times}_{ab}(\hat{\Omega})h_{\times}(t,\hat{\Omega}) \ , \tag{2.3}$$

where $h_{+,\times}$ are the polarization amplitudes and $e^{+,\times}_{ab}$ are the polarization tensors, which we write in the solar system barycenter (SSB) frame as

$$e^{+}_{ab}(\hat{\Omega}) = \hat{m}_a \hat{m}_b - \hat{n}_a \hat{n}_b \tag{2.4}$$

$$e^{\times}_{ab}(\hat{\Omega}) = \hat{m}_a \hat{n}_b + \hat{n}_a \hat{m}_b, \tag{2.5}$$

(Wahlquist, 1987). In these equations, we define $\hat{\Omega}$ as a unit vector pointing from the GW source to the SSB, written as

$$\hat{\Omega} = -\sin\theta\cos\phi\hat{x} - \sin\theta\sin\phi\hat{y} - \cos\theta\hat{z} \ . \tag{2.6}$$

We define the vectors $\hat{m}$ and $\hat{n}$ as

$$\hat{m} = \sin\phi\hat{x} - \cos\phi\hat{y}, \tag{2.7}$$

$$\hat{n} = -\cos\theta\cos\phi\hat{x} - \cos\theta\sin\phi\hat{y} + \sin\theta\hat{z} \ . \tag{2.8}$$

The pulsar's response to the GW source is described by the antenna pattern functions (Sesana & Vecchio 2010; Ellis et al. 2012; Taylor et al. 2016 and references therein)

$$F^{+}(\hat{\Omega}) = \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}}, \tag{2.9}$$

$$F^{\times}(\hat{\Omega}) = \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})}{1+\hat{\Omega}\cdot\hat{p}} \ , \tag{2.10}$$

where $\hat{p}$ is a unit vector pointing from the Earth to the pulsar.

Finally, we write the signal s induced by the GW, as seen in pulsar's residuals, as

$$s(t, \hat{\Omega}) = F^{+}(\hat{\Omega})\Delta s_{+}(t) + F^{\times}(\hat{\Omega})\Delta s_{\times}(t). \tag{2.11}$$

Here, $\Delta s_{+,\times}$ represents the difference between the signal induced at the Earth (the Earth term) and that at the pulsar (the pulsar term), and can be written as

$$\Delta s_{+,\times}(t) = s_{+,\times}\left(t_p\right) - s_{+,\times}(t) \tag{2.12}$$

where t is the time at which the GW passes the SSB and t_p is the time the GW passes the pulsar.[2] These times can be related from geometry by

$$t_p = t - L(1 + \hat{\Omega} \cdot \hat{p}) \tag{2.13}$$

where L is the distance to the pulsar.

For a circular binary at zeroth post-Newtonian order, $s_{+,\times}$ is given by (Wahlquist, 1987; Lee et al., 2011; Corbin & Cornish, 2010)

[2]This definition is occasionally written as the negative of the right side of the equation here, e.g., $s_{+,\times}(t) - s_{+,\times}\left(t_p\right)$ as in A19. This is resolved with a change of convention in the definition of the GW antenna pattern, as we have done here; thus all results are consistent between these works.

$$
\begin{aligned}
s_+(t) =& \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \left[-\sin 2\Phi(t) \left(1 + \cos^2 i\right) \cos 2\psi \right. \\
& \left. - 2 \cos 2\Phi(t) \cos i \sin 2\psi \right], \\
s_\times(t) =& \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \left[-\sin 2\Phi(t) \left(1 + \cos^2 i\right) \sin 2\psi \right. \\
& \left. + 2 \cos 2\Phi(t) \cos i \cos 2\psi \right],
\end{aligned}
\tag{2.14}
$$

where i is the inclination angle of the SMBHB, ψ is the GW polarization angle, d_L is the luminosity distance to the source, and $\mathcal{M}$ is the chirp mass, which is related to the two black hole masses as

$$
\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \tag{2.15}
$$

It is important to note that $\mathcal{M}$ and ω, in this case, refer to the observed redshifted values.

For a circular binary, we relate the orbital angular frequency to the GW frequency with $\omega_0 = \pi f_{\rm GW}$, where $\omega_0 = \omega\left(t_0\right)$. For this work, as in A19 we define t_0 as the last MJD in the 11-year data set (MJD 57387). The orbital phase and frequency of the SMBHB are given by

$$
\Phi(t) = \Phi_0 + \frac{1}{32} \mathcal{M}^{-5/3} \left[\omega_0^{-5/3} - \omega(t)^{-5/3} \right], \tag{2.16}
$$

$$
\omega(t) = \omega_0 \left(1 - \frac{256}{5} \mathcal{M}^{5/3} \omega_0^{8/3} t \right)^{-3/8}, \tag{2.17}
$$

where Φ_0 and ω_0 are the initial orbital phase and frequency. As in A19, we use the

full expression for $\omega(t)$ to maintain consistency across runs, as this form is needed to model the signal at the higher frequencies sampled in some runs, as described in subsubsection 2.3.4.3.

2.3.3 Software and Analyses

In this work, we make use of NANOGrav's GW detection package, `enterprise`[3], an open-source code written fully in Python that contains a built-in interface with the pulsar timing data and noise models required to perform Bayesian GW analysis (Arzoumanian et al., 2018a, limits and detection). Basic algorithms for Bayesian continuous wave analysis are described in detail in a number of past works (e. g. Ellis, 2013; Ellis & Cornish, 2016)

Using `enterprise`, we can use a priori constraints on a binary system, which come from electromagnetic observation (for instance, the period of 3C66B) to set priors on GW parameters that are derived from the binary model. Within `enterprise` we can easily add these priors to the timing model and noise model to obtain a full model of the signal. We then perform Markov-Chain Monte Carlo (MCMC) methods implemented in `PTMCMCSampler`[4] to find the posterior distribution for each of the free parameters. For 'blind' continuous wave (CW) searches as in A19, we typically set uninformative priors, which are uniform across the allowed range of values, for the binary system's parameters, such as sky location, frequency, mass, and distance to the source. Thus, the methods here could be considered a "targeted"

[3] https://github.com/nanograv/enterprise
[4] https://github.com/jellis18/PTMCMCSampler

search by our use of informed priors.

For instance, in the simplest treatment of 3C 66B, a specific binary model has been hypothesized, with measurements and associated unknowns in the mass, mass ratio, and orbital frequency (e. g. S03; I10). We can use these electromagnetically constrained parameters, in addition to knowledge of the location of this object on the sky, to restrict our priors.[5]

Assuming a SMBHB with a circular orbit, a continuous GW signal can be characterized by eight of the following nine parameters:

$$\{\theta, \phi, f_{\rm GW}, \Phi_0, \psi, i, \mathcal{M}, d_{\rm L}, h_0\}, \tag{2.18}$$

which represent the GW source's:

- position on the sky (θ, ϕ);
- GW frequency, related to the orbital frequency at some reference time $(f_{\rm GW})$;
- orbital phase at some reference time (Φ_0);
- GW polarization angle (ψ);
- orbital inclination (i);
- chirp mass $(\mathcal{M})$;
- luminosity distance $(d_{\rm L})$;

[5]Note that our restricted priors might not always be Gaussian; in some cases, electromagnetic observations of a source may produce a model that contains greater complexity than Gaussian error bars. In such cases, non-Gaussian priors must be used. The functionality exists in `enterprise` for studies that would require such a setup. As an example, if cyclic flux variability is observed, the period of variability might represent the fundamental orbital frequency, a harmonic, or even a resonance, requiring a multi-valued prior. In our analysis, the reported errors on binary masses from I10 were asymmetric, and thus for some analyses, our chirp mass prior required an asymmetric distribution.

- strain amplitude (h_0), which is related to the chirp mass, GW frequency, and luminosity distance .

The ninth parameter is redundant, as the strain amplitude h_0 can be defined by

$$h(t) = F_+ h_+ + F_\times h_\times = A h_0 \cos\left(\Phi(t) - \Phi_0\right) \tag{2.19}$$

(Sathyaprakash & Schutz, 2009), where

$$\begin{aligned} A &= \left(A_+^2 + A_\times^2\right)^{1/2} \\ A_+ &= \frac{1}{2} F_+ \left(1 + \cos i^2\right) \\ A_\times &= F_\times \cos i, \end{aligned}$$

and can be related to other physical parameters by

$$h_0 = \frac{2\mathcal{M}^{5/3}(\pi f_{\rm GW})^{2/3}}{d_L}. \tag{2.20}$$

Since the strain is entirely determined by $\mathcal{M}$, $f_{\rm GW}$, and d_L, a limit on h_0 based on a PTA search can be translated into constraints on these source parameters. Since the uncertainties on θ, ϕ, and d_L are much smaller than the PTA sky localization accuracy, by targeting a specific source with a known position and redshift, we can set these parameters as constant values, and therefore reduce the number of search parameters to five.

In all runs, there is also a set of free parameters associated with each pulsar included in the PTA which are varied in the analysis. First of these is the pulsar distance, which has a Gaussian prior in all cases. In pulsars where the distance is reported in Verbiest et al. (2012), the Gaussian is defined using the recognized distance and the associated error. For the remaining pulsars, the Gaussian prior is set to a fiducial 1.0±0.2 kpc, which is consistent with the distribution of distances and uncertainties obtained from Verbiest et al. (2012). Although this range does not necessarily encompass the actual distances to most of these pulsars, it works as a proxy value, and the choice of this value does not affect our results. As in A19, this assumption can be seen to hold in the posteriors for these pulsars, as the prior is returned in all cases, meaning this analysis cannot inform on the distances for these pulsars. This is expected, as these pulsars are largely those with shorter observation baselines, which are influencing the PTA to a smaller degree. The recovered pulsar distances also affect the GW frequency difference between the Earth and the pulsar, which therefore will be related to the chirp mass. When a wide range of chirp mass values are allowed by the data, the uncertainty in the pulsar distances is not significant to the final result of the search. Additionally, for small chirp masses, for even a large change in the distance to the pulsar, the change induced in the GW frequency at the pulsar is well below the resolution limit of the PTA ($1/T_{obs}$). This angular frequency at the pulsar can be calculated as

$$\omega_{p,0} = \omega_0 \left(1 + \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3} d_p(1 + \hat{\Omega}\cdot\hat{p})\right)^{-3/8}, \tag{2.21}$$

where d_p is the distance from the Earth to the pulsar.

Also included is the GW phase at the pulsar. While this quantity could be calculated geometrically from the other parameters, including it as a search parameter mitigates potential issues sampling the complex parameter space, which arise due to the large uncertainty on the distances to the pulsars compared to the GW wavelength.

As is standard for these types of analyses, (e.g., Arzoumanian et al. 2018a; Aggarwal et al. 2019) the white noise of each pulsar (described as EFAC, EQUAD, and ECORR) is held fixed. The power spectral density of the pulsar intrinsic red noise is modeled as

$$P = \frac{A_{\rm red}^2}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma} {\rm yr}^3, \tag{2.22}$$

where $A_{\rm red}$ (the red noise amplitude) and γ (the red noise spectral index) are also allowed to vary in each pulsar in our Markov-Chain Monte Carlo simulation. Here, $f_{\rm yr}$ is 1/(1yr) in Hz. To assist the sampler, empirical distributions of the red noise parameters were made from single pulsar noise run posteriors and used to create jump proposals. These determine how steps in the MCMC are taken through generating proposed samples, and were added to significantly improve sampling and decrease burn-in time for our analyses. For a more detailed description, see Appendix A of A19.

Our treatment of the red noise in one pulsar, J0613−0200, required additional noise modeling. As described in A19, this pulsar possesses extra unmodeled noise processes that, in the 11-year continuous wave search, presented as an increase in

strain upper limit at a frequency of 15 nHz. In this work, this manifested as poor sampling in the CW parameters, particularly in $f_{\rm GW}$. Because of this poor sampling, the $f_{\rm GW}$ parameter would periodically get stuck near this frequency. Due to this pulsar's location relative to 3C 66B, which places it among the ten pulsars with the highest antenna pattern response amplitudes, it is important to find a robust solution to these issues rather than remove the pulsar from the analysis. To mitigate this effect, we applied more sophisticated noise modeling techniques to allow the red noise to deviate from the typical power-law, with corresponding jump proposals to assist sampling. The noise model that was chosen is a t-process spectrum, which allows for 'fuzziness' in the typical power-law spectrum by scaling the power spectral density by a variable factor for each frequency. This model is created by generalizing the typical Gaussian process prior to a Student's t-distribution. This process will be discussed in more depth in Simon & Hazboun (in prep), and, due to increasingly complex data, will likely become more typical in future analyses.

Even with this model, poor sampling in the $f_{\rm GW}$ parameter still occurred, and can be attributed to unmodeled noise due to changes in the dispersion measure of pulsar J1713+0747, caused by variations in the interstellar medium along the line of sight (Lam et al., 2018; Hazboun et al., 2020). While this pulsar is NANOGrav's most sensitive in general, it is not particularly sensitive to 3C 66B, as shown in Figure 2.1, and thus excluding it did not significantly effect the upper limit on target 3C 66B. As such, this pulsar was removed from our search.

The above procedure is used for all `enterprise` runs as described in detail in the next subsection.

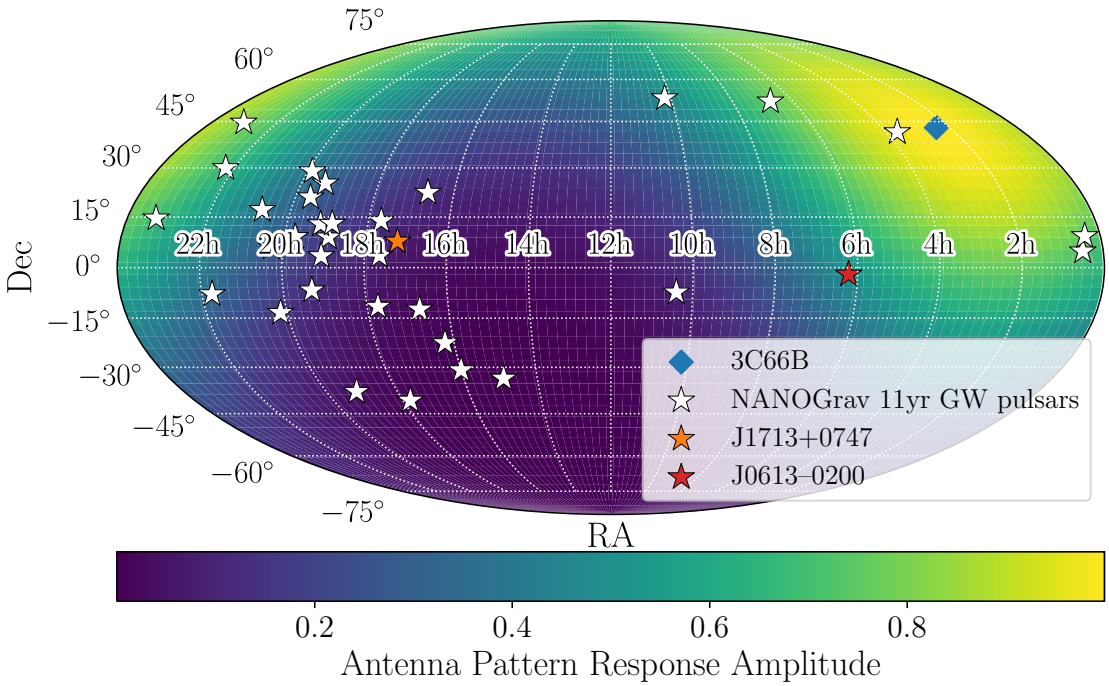

Figure 2.1: Sky map depicting the antenna pattern response amplitude ($F_\times^2 + F_+^2$) due to a GW located at the sky position of 3C 66B. Also plotted are the locations of the 34 pulsars used in GW analyses of the NANOGrav 11-year data set, with the two pulsars in need of special attention noted with separate colors.

2.3.4 Four Distinct Tests

We constructed several separate set-ups for `enterprise` for the purpose of testing distinct hypotheses. The purpose of each of these, and the difference in procedures within `enterprise`, is described below.

2.3.4.1 Detection

To determine if a CW from 3C 66B is detected, we conduct an `enterprise` search using a single frequency, with a value corresponding to the 1.05-year orbital period for a circular binary, making the final set of search parameters

$$\{\Phi_0, \psi, i, \mathcal{M}\}. \tag{2.23}$$

Due to the frequency resolution of the PTA, which is defined by the timing baseline, it is reasonable to set a parameter with errors of this magnitude (Table 2.1) to a constant value. However, we will explore the relaxation of this assumption in later sections. Note that the I10 and S03 models make assumptions about the electromagnetic data which may or may not be correct; our model simply tests the presence of a SMBHB in this system at a period of 1.05 years.

The detection prior on $\mathcal{M}$ is log-uniform in the range 10^7 to 10^{10} $M_\odot$, and is sampled in log-space. This prior is convenient for calculating Bayes factors as a measure of detection significance, using the Savage-Dickey formula (Dickey, 1971),

$$\mathcal{B}_{10} \equiv \frac{\text{evidence}[\mathcal{H}_1]}{\text{evidence}[\mathcal{H}_0]} = \frac{p(h_0 = 0|\mathcal{H}_1)}{p(h_0 = 0|\mathcal{D}, \mathcal{H}_1)}, \tag{2.24}$$

Here, $\mathcal{H}_1$ is the model with a GW signal plus individual pulsar red noise, and $\mathcal{H}_0$ is the model with individual pulsar red noise only. The prior and posterior volumes at $h_0 = 0$ are $p(h_0 = 0|\mathcal{H}_1)$ and $p(h_0 = 0|\mathcal{D}, \mathcal{H}_1)$, respectively. We are able to apply the Savage-Dickey formula because these models are nested ($\mathcal{H}_0$ is $\mathcal{H}_1$ where $h_0 = 0$), and $p(h_0 = 0|\mathcal{D}, \mathcal{H}_1)$ is approximated as the fraction of quasi-independent samples in the lowest-amplitude bin of a histogram of h_0. The error in the Bayes factor is computed as

$$\sigma = \frac{\mathcal{B}_{10}}{\sqrt{n}}, \tag{2.25}$$

where n is the number of samples in the lowest amplitude bin. This process is done once the samples in GW strain are calculated from the directly sampled parameters. In the detection analyses, the red noise amplitude is sampled with a matching prior (log-uniform in $A_{\rm red}$). All other GW parameters are searched with a uniform prior.

2.3.4.2 Upper Limits

To set an upper limit on the chirp mass of 3C 66B, we again conduct an `enterprise` search using a single frequency, with a value corresponding to the 1.05-year orbital period, making the final parameter set as in the previous section (Equation 2.23). However, in contrast with the case for detection, the upper limit prior on $\mathcal{M}$ is uniform (rather than log-uniform) meaning the prior set on the $\log_{10}\mathcal{M}$ exponentially increases over the range $\{7, 10\}$. This is done as an astrophysically reasonable prior, as we expect SMBHBs to lie anywhere in this mass range, while still allowing for efficient sampling. Additionally, this prior choice allows the derived

upper limit to be as conservative as possible by allowing a higher proportion of high chirp mass samples, and be independent from the choice of lower prior bound. In the upper limit analyses, the red noise amplitude is sampled with a matching prior (uniform in $A_{\rm red}$). Upper limits are taken to be the value of the 95th percentile of the posterior distribution. Following the approach of Arzoumanian et al. (2018a), we calculate the error on upper limit calculations as

$$\sigma = \frac{\sqrt{x(1-x)/N_s}}{p\left(h_0 = h_0^{95\%}|\mathcal{D}\right)}, \tag{2.26}$$

where $x = 0.95$ and N_s is the number of effective samples in the chain, which is estimated by dividing the total number of samples by the autocorrelation length of the chain.

2.3.4.3 Frequency Prior Testing

In addition to the tests described above of the S03 and I10 models, where the GW frequency is fixed to discrete values as in other continuous wave searches (A19; Arzoumanian et al. 2014), it is also crucial to test frequencies within the confidence region of these values. For this aim, we have developed methods to directly sample in $f_{\rm GW}$. These include specialized parameter groupings and jump proposals to help the sampler move through the more complex parameter space. Using these techniques, we are able to obtain an upper limit from the $\mathcal{M}$ posteriors for a variety of frequency priors from various `enterprise` setups.

When searching over GW frequency, a log-uniform chirp mass prior is used,

and the samples are re-weighted during upper limit calculations to modify the prior choice from a uniform-in-log distributions of masses to a uniform-in-linear distributions of masses, the latter of which is more common in upper-limit analyses by virtue of insensitivity to the lower sampling boundary. This both assists with sampling and maintains a consistent prior on the GW strain, which is not directly sampled. To match the $\mathcal{M}$ prior, a log-uniform prior is used on $A_{\rm red}$. Since we are no longer fixing f_{GW} to a single value, our final parameter set for these searches was

$$\{\Phi_0, \psi, i, \mathcal{M}, f_{\rm GW}\}, \tag{2.27}$$

In addition, we also chose to limit our GW frequency prior to a range of 1–100 nHz, rather than the 1–300 nHz used in A19. Besides the PTA's insensitivity at these high frequencies, we expect a source to remain in these frequency bins for very little time, with residence timescales as small as months, so their detection prospects are minimal (Burke-Spolaor et al., 2019; Hazboun et al., 2019b).

Using the three priors shown in Table 2.2, we are able to find re-weighted upper limits for a variety of scenarios. These include:

1. The GW frequency is known, and set to a single value
2. The GW frequency is known with large errors, and the error region is searched over
3. The GW frequency is not known or has significant uncertainty, and the entire PTA sensitivity band is searched over.

Then, we examined the change in re-weighted chirp mass upper limit as a function

Table 2.2: Frequency Prior Testing Weighted Upper Limits

Scenario	$f_{\rm GW}$ Prior	Weighted $\mathcal{M}$ Upper Limit ($10^9 M_\odot$)
1	Constant	1.57± 0.02
2	10σ	1.54 ± 0.01
3	Log-Uniform	8.68 ±0.07

of frequency prior width. In addition to allowing for possible errors in the orbital period measured by S03, these widened priors allow us to test the feasibility of this process on a less constrained source. Additionally, if there was any significant frequency evolution in the source, a signal would still have the chance to be detected in either of these setups. In addition to a single value and a uniform prior across the PTA sensitivity bandwidth, we also use 10 times the uncertainty on the predicted frequency as an example of a search with significant uncertainty. We also bin the samples of the widest $f_{\rm GW}$ search to interpolate between these three individual prior widths. The results of this examination are described in section 2.4, and are summarized in Figure 2.5

2.3.4.4 Test of a Specific Binary Model

To directly test the consistency of the model presented in I10 with the NANOGrav data, we create priors for an `enterprise` run corresponding to the values presented (see the first line of Table 2.3). For $f_{\rm GW}$, we are able to use a Gaussian prior, where the error on the measured value from I10 directly corresponds to the standard deviation of the prior. However, $\mathcal{M}$ has uneven error bars, so a more complicated

prior is needed. Here, we fit a skewed normal distribution to the reported value and error, and construct a skewed normal prior based on this distribution, and also keep a log-uniform prior on $A_{\rm red}$. Therefore, the final parameter set for this search was

$$\{\Phi_0, \psi, i, \mathcal{M}, f_{\rm GW}\}. \tag{2.28}$$

To analyze the amount of information gained between the prior and posterior models, we employed the Kullback–Leibler divergence (Kullback & Leibler, 1951). We calculate this information gain in bits between the posterior $p(x|d)$ and the prior $p(x)$ as

$$D_{\rm KL}(P\|Q) = \int_{-\infty}^{\infty} p(x|d) \log\left(\frac{p(x|d)}{p(x)}\right) dx. \tag{2.29}$$

This is done for the distributions for both $\mathcal{M}$ and $f_{\rm GW}$. To maintain consistency between forms of the posterior and the prior, we fit a skewed normal distribution to both posteriors to directly compare to the prior.

2.4 Results

The results discussed in this section can be reproduced, and the MCMC data examined, using code provided for the reader's convenience.[6]

2.4.1 Detection

Using the setup for a detection run as described in subsubsection 2.3.4.1, we find no evidence for a GW signal from 3C 66B. We calculate a Savage-Dickey Bayes

[6] https://github.com/nanograv/11yr_3c66b

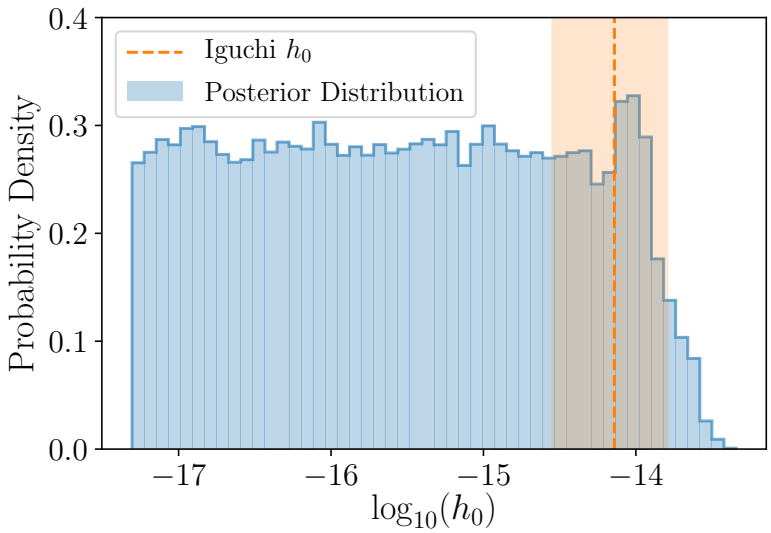

Figure 2.2: Posterior for the detection analysis described in subsection 2.4.1 (blue). The vertical orange region describes the area of parameter space where a signal with the parameters found by I10 would lie. While the upper end of the parameter space is ruled out, there is clearly no value that is preferred by the sampler.

factor of $\mathcal{B}_{10} = 0.74 \pm 0.02$. Therefore, there is no evidence for the detection of a GW signal in the data. The posterior for this run is plotted in Figure 2.2.

2.4.2 Upper Limits

As no GW signal is detected from 3C 66B, we set upper limits on the chirp mass using the procedure described in subsubsection 2.3.4.2. Using the constant-value frequency prior at 60.4 nHz (corresponding to the 1.05-year orbital period), we set a 95% upper limit of $(1.65 \pm 0.02) \times 10^9\ M_\odot$ for $\mathcal{M}$ of the SMBHB in 3C 66B. This value corresponds to a strain of $(2.47 \pm 0.05) \times 10^{-14}$. To compare, the expected strain of the model in I10 is $(7.2^{+6.8}_{-5.8}) \times 10^{-15}$. As can be seen in Figure 2.3, while we achieve a factor of 4.3 improvement over the limit set by J04, we cannot rule out the I10 model. The posterior distribution of samples does include a peak at about

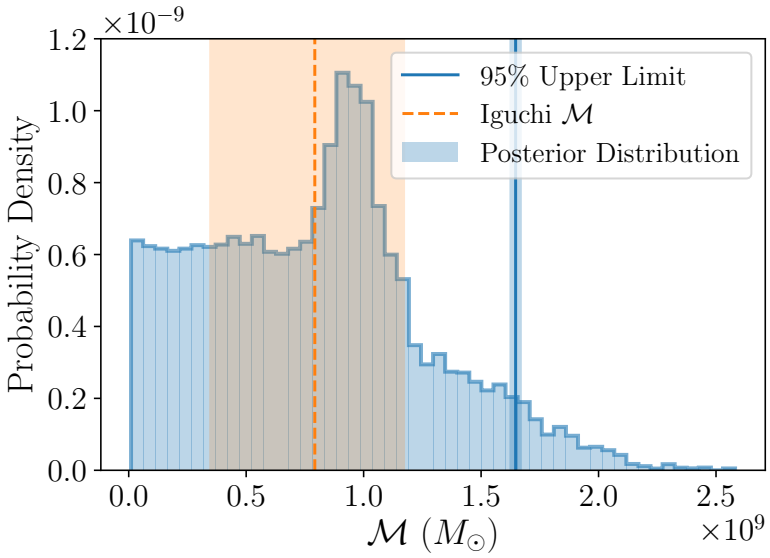

Figure 2.3: The chirp mass posterior histogram is plotted in blue, with a vertical line depicting the 95% upper limit. Shown in orange is the chirp mass upper limit of I10, with the shaded region representing the error on the value. With these methods, the I10 mass estimate is impossible to rule out. We also note that the peak in the posterior at $1 \times 10^9 M_\odot$ is not statistically significant.

1×10^9 $M_\odot$, which is within the error region for the chirp mass calculated from I10. However, this peak is not statistically significant, and is able to be traced to a single pulsar, J1909−3744. By examining the posterior distributions constructed from samples corresponding to this peak, we find structure in the GW phase posterior at J1909−3744 that does not occur for any other pulsar. This likely occurs due to covariances between the model and sinusoidal behavior caused by noise processes in the data as a real GW signal would be recovered by more than one pulsar. Therefore, this peak in the posterior is not indicative of a signal, and our upper limit can be considered robust. We will note that the upper limit listed can be calculated for the reader's preferred distance using the transformation described in subsection 2.3.1.

2.4.3 Frequency Prior Testing

As described in subsubsection 2.3.4.3, we also performed tests to quantify how much our upper limits might improve if we have constrained (through electromagnetic observation) the orbital frequency of the target. While for 3C 66B the orbital frequency is assumed to be known to within small errors, for other targets, a frequency may not be known or be only poorly constrained. This test provides a sense of how well the period must be constrained to provide effective sensitivity gains for a GW search.

Using the three scenarios described above, we are able to characterize the change in re-weighted upper limit between the setups. The result of the log-uniform prior search over the entire frequency band is summarized with Figure 2.4. The white area represents the area of $\mathcal{M}$-$f_{\rm GW}$ parameter space ruled out in this analysis. From the uniformity of the samples over the parameter space, it is clear there are no sampling issues. This is due to the improved sampling methods described in subsection 2.3.3. The weighted 95% upper limit is plotted for each frequency bin, allowing us to quantify for which frequencies we are the most sensitive to 3C 66B. We note that for the very lowest frequencies, the upper limit is dependent upon the choice of prior, as the search cannot rule out any of the prior range.

In addition to the three runs described above, it was also possible to infer the upper limit that would be derived from a run with a frequency prior width between those of the three separate runs. To do this, we bin the samples in the scenario 3 (widest $f_{\rm GW}$ prior) run to keep only a certain range of frequencies and recalculate

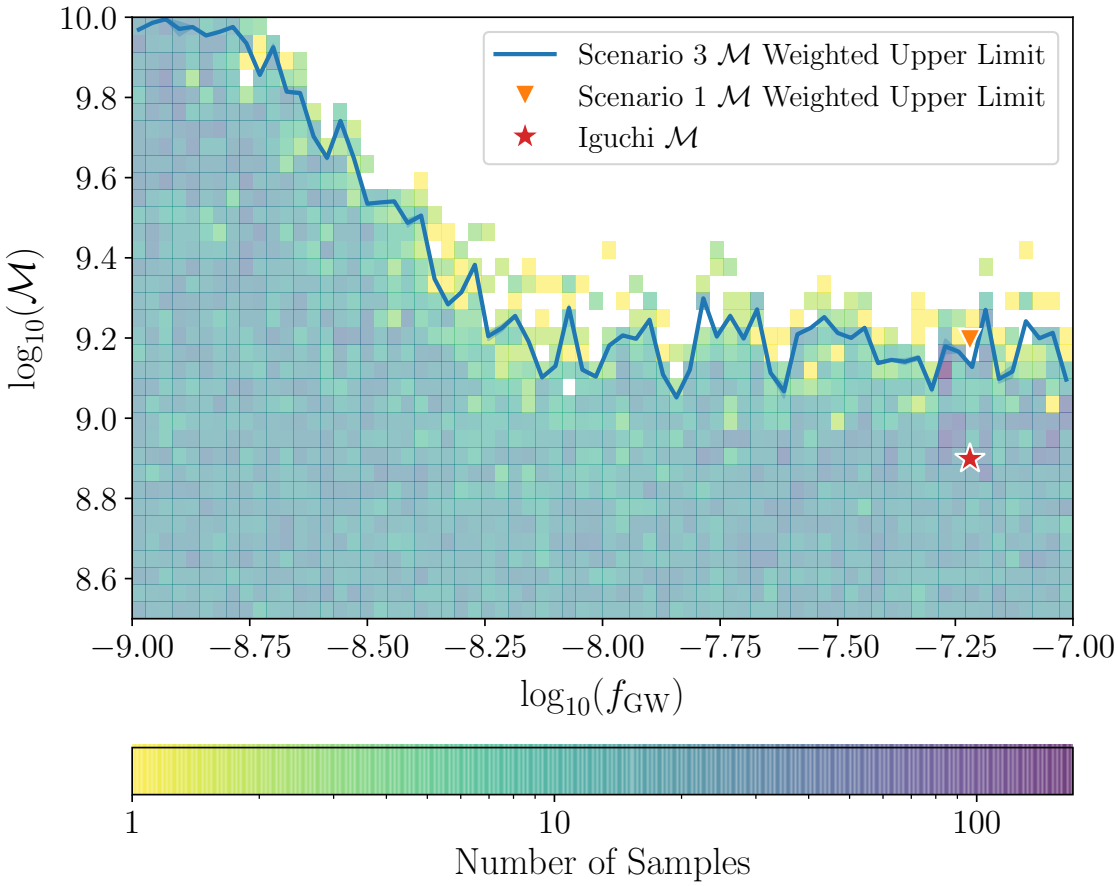

Figure 2.4: 2D histogram of samples in the log-uniform prior setup. Also plotted is the weighted 95% upper limit for each frequency bin (blue) from the scenario 3 setup. The white area indicates the section of parameter space ruled out by our search. It is clear from the uniform distribution of samples across all frequency and mass channels that all sampling issues have been resolved. This uniform distribution also makes clear that there is no indication of a signal at the distance and sky location of 3C 66B. We only plot the upper half of the parameter space in $\mathcal{M}$ to resolve more detail. Below $\log_{10}\mathcal{M} = 8.5$, all sampling is uniformly distributed, identically to the upper half of the figure. For comparison, the scenario 1 weighted upper limit (orange triange) and I10 chirp mass estimate (red star) are also shown.

the weighted upper limit for this subset. These bins increase symmetrically in log space about the value of $f_{\rm GW}$ reported by S03, from a log space full-width of 0 dex (a constant) until the upper bound reaches $f_{\rm GW} = 100$ nHz. After this, only the lower bound expands to reach a full log space full-width of 2 dex (essentially, 2 orders of magnitude in linear space). The weighted upper limits calculated from these binned samples are plotted in Figure 2.5.

Also plotted in Figure 2.5 are the upper limits from the three individual runs. From the consistency of these points with the calculated curve, it is clear that this technique is robust. Additionally, this shows the feasibility of searching over $f_{\rm GW}$, as the results are consistent with those calculated for both an individual frequency and a small range.

As can be seen in Figure 2.5 and Table 2.2, there is nearly an order of magnitude difference in the upper limits derived from frequency varied runs of different prior widths. Because the upper limits at the very lowest frequencies are dependent upon the prior choice, the difference seen here is a lower limit. However, from the curve calculated from binned samples, we see that this increase does not begin until about one order of magnitude in frequency space about the I10 value is included. This implies that a targeted search such as this is worthwhile even without exact orbital information, as long as the frequency is known to within an order of magnitude.

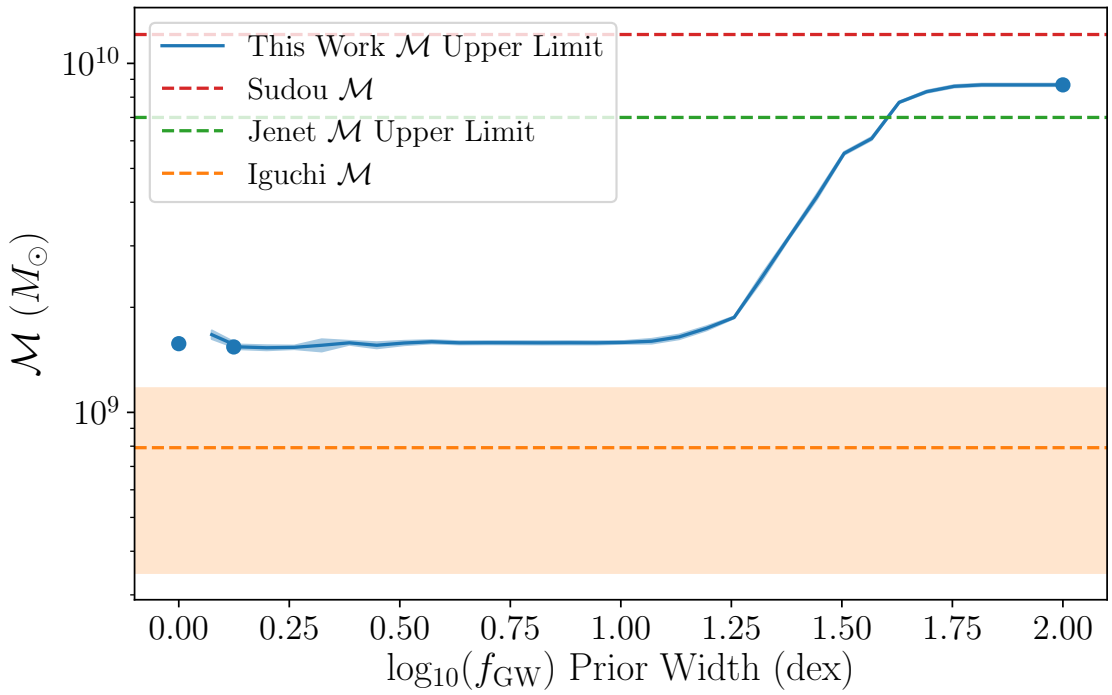

Figure 2.5: Chirp mass upper limits plotted with respect to frequency prior width (blue). Also shown as horizontal lines are previous upper limits set by S03 (red), J04 (green), and I10 (orange), from top to bottom. Shaded regions describe error bars on the quoted limit. It is clear that none of these upper limits rule out that of I10. However, this figure accentuates the fact that when a period is known to less than 1 order of magnitude of precision, the limits on the target's mass improve by nearly one order of magnitude; that is, while the tightest prior produces the lowest upper limit, moderately wide priors also produce similar results, indicating that perfect orbital models would not be necessary to perform such a search on other systems. It is not until the prior spans approximately an order of magnitude that sensitivity is lost. Also plotted for comparison are the weighted upper limits for each of the three separate runs.

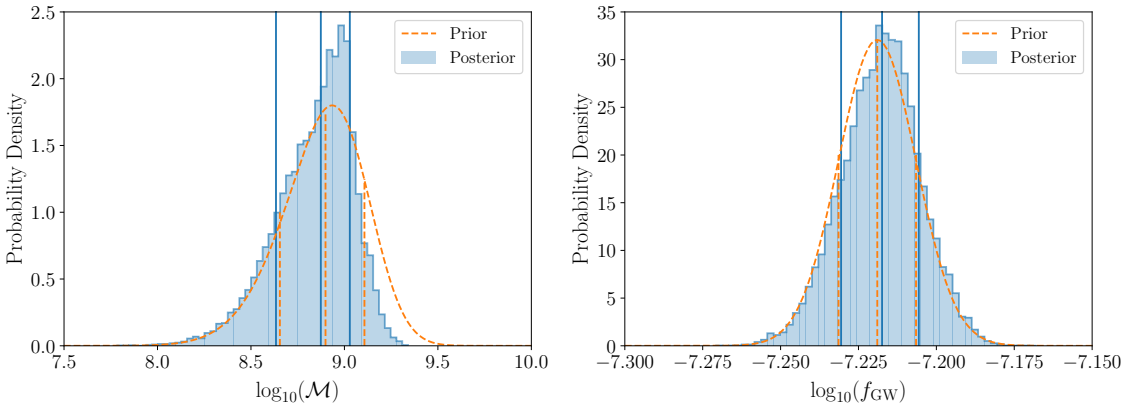

Figure 2.6: Posteriors (blue) and priors (orange) for the direct test of the model presented in I10. Vertical bars mark the 16.86, 50, and 84.13 percentiles of each, to represent the 1σ error bars.

2.4.4 Test of a Specific Binary Model

To directly test our sensitivity to a GW from the model of 3C 66B proposed in I10, we directly test priors as described in subsubsection 2.3.4.4. In Figure 2.6, we can compare the prior and posterior for both $f_{\rm GW}$ and $\mathcal{M}$. These distributions are quantified in Table 2.3, where the error on the posterior values are calculated with the percentiles of the posterior distribution corresponding to 1σ error bars. The values of $f_{\rm GW}$ are consistent with those of the prior, but for $\mathcal{M}$, we are able to significantly lower the upper bound on the value, effectively ruling out part of the high mass region of the model.

Additionally, we report the information gained between the posterior and the prior as described in subsubsection 2.3.4.4. The differences in the distributions for $f_{\rm GW}$ produce a KL divergence of 0.0096, while those of the $\mathcal{M}$ distributions produce a KL divergence of 0.2597. While neither of these values is large, it is clear that much more information is gained about the chirp mass of 3C 66B from this model

Table 2.3: Model Testing Prior and Posterior Values

	log(Frequency)	log(Chirp Mass)
Iguchi (Prior)	-7.219 ± 0.012	$8.90^{+0.21}_{-0.24}$
This Work (Posterior)	$-7.217^{+0.012}_{-0.013}$	$8.87^{+0.16}_{-0.24}$

test than the GW frequency.

2.5 Discussion

To provide context for the upper limit on 3C 66B set in this work, we can compare to the limits set in A19, which do not have the benefit of electromagnetic constraints (i. e. a 'blind' search). This comparison will allow us to estimate the improvement in sensitivity gained by including electromagnetic data over a typical blind search. By comparing our strain upper limit of $(2.47 \pm 0.05) \times 10^{-14}$ to the sensitivity curve in Figure 3 of A19, where the strain upper limit at the nearest searched frequency is 5.3×10^{-14} nHz, we observe that we have gained a factor of 2.1 in sensitivity by holding the source position fixed in our search. Note that a much greater improvement comes from knowing the binary candidate's period, as demonstrated in Figure 2.4 and Figure 2.5.

With the framework developed in Hazboun et al. (2019b) we can construct detection sensitivity curves to estimate the PTA that will be required to detect or rule out the mass model presented in I10. The `hasasia` (Hazboun et al., 2019a) package[7] allows us to construct these detection sensitivity curves using a straight

[7] https://hasasia.readthedocs.io/

forward matched filter statistic and to simulate PTA data with control over the number of pulsars, observing cadence, timing precision, and data length. Using this software to estimate an idealized signal-to-noise ratio (S/N) (see Eqn (79) in Hazboun et al. (2019b)), assuming the parameters in I10 and using the pulsar noise parameters in Arzoumanian et al. (2018b) we obtain $S/N = 0.87$. We used this software to extend the baseline of the the existing 11-year NANOGrav data set by adding new data to the existing pulsars with a timing precision and cadence that matches recent data. We also augmented the PTA, adding new pulsars with timing precisions and cadences similar to those already in the array; we added pulsars for each projected year at a rate comparable to the current growth-rate of NANOGrav, which has been approximately 7 pulsars per year for the past three data sets.

We find that NANOGrav should be able to detect or rule out the existence of a SMBHB in 3C 66B with the I10 mass within five to eight years from the end of the data set considered here. However, while `hasasia` allows us to calculate the PTA's sensitivity to a CW at a specific sky location, it is unable to set other parameters (such as luminosity distance) as known due to electromagnetic information about the GW source as is done in this work. As is discussed above, including source parameters that are electromagnetically derived to reduce the parameter space of the GW search allows for an increased sensitivity. Because of this, using electromagnetic information will likely allow us to accelerate this estimated timeline. To more reliably estimate this timescale, detailed simulation work will be necessary to quantify the improvement made by including electromagnetic information over typical searches.

Because the sensitivity of the array depends heavily on the observing baseline of each pulsar, the inclusion of additional data can help tremendously. Data of this sort are accessible through the IPTA (Perera et al., 2019), and followup analyses of 3C 66B by the international community could prove fruitful. This timeline to the PTA sensitivity required to confirm or deny 3C 66B as a SMBHB will be reduced with the more rapid addition of pulsars to the array, e.g., by adding more than 7 per year. This improvement will be accelerated if the newly included pulsars are near the sky location of 3C 66B, as, currently, there are few pulsars in the array that are very sensitive to 3C 66B. To accomplish this, pulsar searches should be undertaken near the sky locations of potential PTA targets to begin improving our sensitivity more rapidly. Some pulsars in this area of the sky can be included through use of data provided by the IPTA (Perera et al., 2019), showing once again that an international effort to detect 3C 66B could be worthwhile.

In addition to the results for GWs from 3C 66B, our work has many implications for detection prospects of other binary candidates. As discussed in subsection 2.4.3 and shown in Figure 2.5, for 3C 66B, it was not until we widened our prior to span an order of magnitude in frequency space on either side of the target frequency that sensitivity was lost. For similar candidates, particularly those at similarly high orbital frequencies, we presume that this result will hold. Therefore, as long as the sky location and luminosity distance of a potential target are known, a search of this type is worth attempting if at least an estimate of an orbital period can be obtained. We will caution that this improvement will differ depending on the sky location of the source, and that the amount of frequency-space that can be

effectively searched with this method will be larger for higher-frequency sources. As can be seen in Figure 2.4, it is the inclusion of samples at low frequencies that raise the upper limit. However, typical errors on binary periods are quite a bit smaller than the limit suggested here, meaning that this method should prove useful for most binary candidates. This method will also account for any frequency error due to unaccounted for frequency evolution of the SMBHB, which, in the case of a detection, would provide important constraints for evolutionary models.

2.6 Conclusions

In this work, we present a new method for performing multi-messenger searches for individual SMBHBs, using 3C 66B as a test case. 3C 66B was first identified as a binary candidate by S03, and was first visited by PTAs in J04, which ruled out the proposed binary model. In the intervening 15 years, a revised model was published by I10 and PTA data and analysis methods have greatly improved. We used the NANOGrav 11-year data set, as well as the collaboration's flagship GW detection package, `enterprise`, to search for GWs from 3C 66B. Here, we are able to limit 3C 66B's chirp mass, at 95% confidence, to $(1.65 \pm 0.02) \times 10^9\ M_\odot$, a factor of 4.3 smaller than the limit set in J04. However, we are unable to rule out the existence of a binary corresponding to the revised model proposed in I10.

In addition to directly placing a limit on the chirp mass of 3C 66B for the published orbital period, we are able to quantify how much this multi-messenger approach increases our sensitivity over a typical 'blind' PTA search. We have con-

ducted a search on real data that includes GW frequency as a free parameter, and from this analysis, we learn that by including frequency constraints from electromagnetic binary source measurements to restrict the prior, we can gain approximately an order of magnitude in sensitivity when compared to a frequency-blind search spanning the whole PTA band. However, this drop in sensitivity does not occur until a relatively wide range of frequencies is searched over, meaning that this approach will be useful even for candidates with relatively poor constraints on their orbital periods.

Chapter 3

Quasars with Periodic Variability: Capabilities and Limitations of Bayesian Searches for Supermassive Black Hole Binaries in Time-Domain Surveys

3.1 Abstract

Supermassive black hole binaries (SMBHBs) are an inevitable consequence of galaxy mergers. At sub-parsec separations, they are practically impossible to resolve and the most promising technique is to search for quasars with periodic variability. However, searches for quasar periodicity in time-domain data are challenging due to the stochastic variability of quasars. In this paper, we use Bayesian methods to disentangle periodic supermassive black hole binary (SMBHB) signals from intrinsic damped random walk (DRW) variability in AGN light curves. We simulated a wide variety of realistic DRW and DRW+sine light curves. Their observed properties (cadence, gaps, photometric uncertaintly) are modeled after the Catalina Real-time Transient Survey (CRTS) and expected properties of the upcoming Legacy Survey of Space and Time (LSST) from the Vera C. Rubin Observatory. Through a careful

Submitted to ApJ.
Contributing authors: Maria Charisi, Stephen R. Taylor, Sarah Burke-Spolaor

analysis of parameter estimation and Bayesian model selection, we investigate the range of parameter space for which binary systems can be detected. We also examine which DRW signals can mimic periodicity and be falsely classified as binary candidates. We found that periodic signals are more easily detectable if the period is short, the amplitude of the signal is large, and the contribution of the DRW noise is weak. We saw similar detection rates both in the CRTS and LSST-like simulations. On the other hand, the false detection rate depends on the quality of the data and is minimal in LSST, with every set of DRW parameters being equally capable of producing false positives in CRTS. Our idealized simulations provide an excellent way to uncover the intrinsic limitations in quasar periodicity searches and set the stage for future searches for supermassive black hole binaries.

3.2 Introduction

Supermassive black hole binaries (SMBHBs) should form frequently in the aftermath of galaxy mergers (Haehnelt & Kauffmann, 2002). However, the evolution from this initial stage to the formation of a bound binary and the final coalescence is complex. After the galaxy merger, the SMBHs hosted in the cores of their parent galaxies sink to the center of the created galactic remnant through dynamical friction. At scales of a few parsecs, stellar scatterings and interactions with ambient gas continue shrinking the binary orbit. If these processes remove sufficient energy and angular momentum so that the binary efficiently overcomes the "final-parsec problem", then gravitational waves (GWs) dominate the binary decay and drive the

binary to the final merger (Begelman et al., 1980; Colpi, 2014; De Rosa et al., 2019).

The most massive binaries (total mass of $10^8 - 10^{10} M_\odot$) emit GWs at low frequencies (few to hundreds of nanohertz). GWs in this frequency band can be detected by Pulsar Timing Arrays (PTAs), and offer one of the only direct probes to SMBHBs at close (roughly milli-parsec) separations (Burke-Spolaor et al., 2019; Taylor et al., 2019). Electromagnetic observations can also infer the existence of a SMBHB, and provide a unique probe of the binary's environment (Bogdanovic et al., 2021). The detection of GWs along with associated electromagnetic counterparts will mark the beginning of multi-messenger astrophysics in the low-frequency regime (Kelley et al., 2019a). In fact, multi-messenger techniques are already being developed. Incorporating information from SMBHB candidates in GW searches allows us to place tighter constraints on the SMBHB chirp mass (Arzoumanian et al., 2020b), and can boost the detectability of the candidate in a typical "blind" search (Liu & Vigeland, 2021).

Closely-separated SMBHBs in the GW regime may be detected as Active Galactic Nuclei (AGNs) or quasars with periodic variability (Haiman et al., 2009). Previous studies have demonstrated a link between AGN and galaxy mergers; this follows naturally from the idea that the mergers bring significant amounts of gas to the central regions of the post-merger galaxies, which may actively accrete onto the SMBHs, triggering AGN activity (Goulding et al., 2018). Similarly, binaries are expected to be surrounded by significant amounts of gas, which can give rise to bright quasar-like electromagnetic emission (Armitage & Natarajan, 2002; Tanaka et al., 2012; Bogdanovic et al., 2021).

Specific predictions for periodic variability in binary AGN has been demonstrated in multiple hydrodynamical simulations of binaries embedded in gaseous disks (MacFadyen & Milosavljević, 2008; D'Orazio et al., 2013; Roedig et al., 2012; Farris et al., 2014). The consensus of these simulations is that the binary carves out a central cavity, i.e. a region of low-density gas. As the binary orbit perturbs the edge of this cavity (especially the secondary SMBH, which moves closer to the edge), it pulls streams of gas inwards. Periodic accretion onto the SMBHs from these streams may produce periodic brightness fluctuations. Another mechanism that produces periodic variability is relativistic Doppler boosting (D'Orazio et al., 2015; Tang et al., 2018). Some of the gas that penetrates the cavity ends up bound to the SMBHs forming mini-disks which orbit with relativistic speeds. The emission from these mini-disks may be periodically boosted (and dimmed), even if the rest-frame luminosity is constant. This signature is prominent for unequal-mass binaries orbiting close to edge-on, where the emission of the faster-moving secondary —which is also typically brighter—dominates the variability.

In recent years, vast photometric databases of time-domain surveys have provided light curves for large samples of AGNs, which are ideal for searches of SMBHBs. Numerous candidates have been identified from systematic searches in optical surveys, such as the Catalina Real-time Transient Survey (CRTS; Graham et al. 2015b), the Palomar Transient Factory (PTF; Charisi et al. 2016), the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Liu et al. 2019), and the Dark Energy Survey (DES; Chen et al. 2020). However, AGN also have intrinsic stochastic variability which makes periodicity identification quite difficult. AGN

variability is successfully modeled by a "damped random walk" (DRW) model, which takes the form of a red-noise process at high frequencies, but a white-noise process at low frequencies (MacLeod et al., 2010; Kozłowski et al., 2010). This intrinsic noise is impressively capable at mimicking periodicity, particularly in sparsely sampled or short-baseline time series (Vaughan et al., 2016). So far, studies have focused on additional signatures for the binary nature of candidates, such as multi-wavelength Doppler boost (D'Orazio et al., 2015; Charisi et al., 2018; Xin et al., 2020a), periodicity with multiple components (Charisi et al., 2015), X-ray spectral excess (Saade et al., 2020b) and distorted radio jets (Kun et al., 2015; Mohan et al., 2016).

However, multi-wavelength follow-up monitoring of candidates is demanding and such studies will be impractical (if not impossible) in the upcoming generation of surveys like the Legacy Survey of Space and Time (LSST) of the Vera C. Rubin Observatory (LSST Science Collaboration et al., 2009). LSST is expected to observe over 20 million of quasars, delivering an unprecedented data set for quasar periodicity searches both in terms of quality and quantity. If we extrapolate the detection rate of SMBHB candidates in the current time-domain surveys ($\sim 1/1000$) to LSST, we expect several thousands of candidates. However, we know that these samples likely contain many false detections (as demonstrated by their tension with the GW background limits when extrapolated to a full binary population; Sesana et al. 2018). On the other hand, theoretical models predict that hundreds of genuine binaries should be detectable in LSST (Kelley et al., 2019b; Xin & Haiman, 2021; Kelley et al., 2021). Because of this, the time is ripe to develop a careful model selection in order to reliably identify binary candidates.

In this work, we explore the capabilities and limitations in identifying quasars with periodic variability in the data sets of the upcoming decade. We simulate idealized AGN light curves that contain DRW noise with realistic parameters, while a subset of those contains sinusoidal variations on top of the DRW noise. We construct a pipeline that employs Bayesian model selection and parameter estimation to identify periodic signals (i.e. the binary candidates) in our sample, and constrain their parameters. Finally, we quantify our ability to select genuine binaries and the degree of contamination with false detections.

This paper is laid out as follows. In section 3.3, we describe the methodology for creating simulated light curves, as well as the Bayesian parameter estimation and model selection methods. In section 3.4, we examine the efficacy of our Bayesian pipeline, as well as present a statistical analysis of this efficacy across the simulated SMBHB population. In section 3.6 we present the conclusions we can draw from our analysis. These include the following key findings:

- Our method can recover orbital periods extremely accurately (even very long or short values), provided the signal is of sufficient strength. The detectability of periodicity also depends on the amplitude of the sinusoid and the contribution of the DRW noise.
- While a DRW process can mask some sinusoids in current surveys, the false positive rate is very low for LSST, and thus it is expected to deliver reliable candidates.
- Particular combinations of DRW and sinusoidal parameters are more likely to mask a signal than others. This will help inform future analyses as we attempt

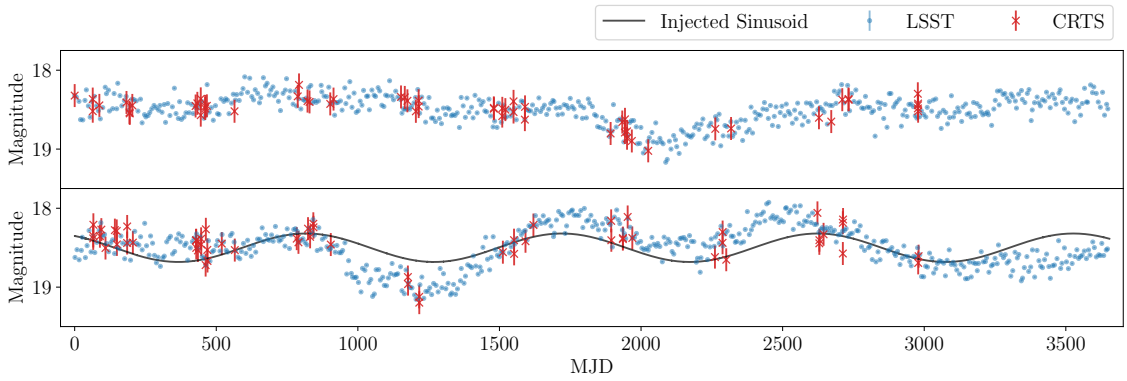

Figure 3.1: An example of simulated light curves containing a DRW process (top panel) and a DRW process plus a sinusoid (bottom panel). The sinusoid is shown in the solid black curve, while the simulated data for a CRTS-like and LSST-like survey are shown as red X's and blue points, respectively. Depending on the DRW and sinusoid parameters, it is possible for these two models to produce deceptively similar results.

to confront the massive data volume that will be produced by LSST.

Finally, in section 3.5, we discuss caveats of our method, future improvements, and the prospects of multi-messenger observations of binaries. This work presents a necessary first step in preparation for the flood of SMBHB candidates in the upcoming Rubin era.

3.3 Methods

As mentioned above, identifying periodicity in quasars is challenging because of the intrinsic stochastic variability of quasars, the relatively short observation baselines compared to the potential binary periods, and the noisy, irregular data. Our goal is to explore the variety of binary signals (e.g., range of periods, amplitudes) that can be detected in current and upcoming time domain surveys. We also aim to assess the expected false positive rate in systematic searches for quasar periodicity.

For this, we simulate typical quasar light curves with realistic DRW noise properties as well as SMBHB light curves which include sinusoidal signals with a variety of periods and amplitudes on top of DRW noise. We chose to model the binary signal with a pure sinusoid both for simplicity, and because previous searches for quasars with periodic variability have focused on quasi-sinusoidal signals. We construct a periodicity detection pipeline that employs a Bayesian model parameter estimation and selection between a DRW and DRW+sine model, and apply it in a wide range of simulated light curves. Below we describe the light curve simulations and the periodicity detection method.

3.3.1 Simulated Data

To ensure that our analysis was realistic, we constructed our simulated light curves with properties that reflect the observational capabilities of ongoing and planned time-domain surveys. Each survey has a distinct observing strategy (depending on their primary scientific objectives), which defines the average cadence (frequency of observations), and observation baseline (length of light curve). Additionally, each survey has a limiting depth, which depends on the size of the telescope used and the exposure time of the typical observation. This defines the photometric uncertainty, which is typically a function of apparent magnitude; dimmer sources have larger photometric errors and vice versa. To limit the complexity of our analysis, we did not incorporate the magnitude dependence of the photometric errors, and used an average photometric error for all the simulated light curves. However,

a future study should address this and other limitations, as discussed in section 3.5. To assess the detectability of SMBHBs in current surveys, we used CRTS-like light curves, whereas for future surveys, we used idealized LSST light curves as described below.

In order to construct the CRTS-like light curves, we extracted 10,000 AGN light curves spread across the sky from the online database.[1] We examined the sampling pattern in this set, which turns out to be similar for most light curves. In particular, for each night the source was observed, the light curves contain clusters of four successive data points, then the next set of observations is taken about one week to one month later for as long as the source is observable (for about six months). Subsequently, there is a significant gap of no data for about six months, e.g., when the source is obstructed by the sun or below the horizon of the specific telescope, and then the pattern roughly repeats. For the sampling of the simulated light curves, we use one typical light curve from that sample as a template. This light curve has an average span of $\sim$20 days between successive nights of observations, gaps of $\sim$200 days, and a total observation baseline of $\sim$8 years. This gives an effective cadence (observation baseline divided by the number of data points) of 46 days. Since very short term variations are not relevant for our study, this calculation did not include multiple observations within the same night. By examining the distribution of photometric errors in these 10,000 light curves, we found that the peak occurs at ~ 0.1 mag, and thus set the average photometric uncertainty in the CRTS-like light curves at this value.

[1] http://nesssi.cacr.caltech.edu/DataRelease/

Survey	Mean Cadence (days)	Photometric Error (mag)	Baseline (years)
CRTS	46	0.1	8
LSST	7	0.01	10

Table 3.1: Average parameters for each survey used to create simulated data sets.

For LSST, the nominal duration is set to ten years, but the observing strategy is not finalized yet (LSST Science Collaborations et al., 2017). The majority of time will be spent on the deep-wide-fast survey mode, which will cover the 18,000 deg^2 footprint with a regular cadence. We set our simulation cadence at a conservative value of seven days, while actual observations may repeat every five or even three nights. We note, however, that LSST will rotate between six filters, and successive observations will provide data in different photometric bands. We do not take this into account in our simulations, but we discuss this caveat further in section 3.5. Since the observations will not repeat in exactly seven day increments, we create a linear grid of time stamps separated by seven days and add Gaussian noise with a standard deviation of one day. In our idealized LSST-like light curves, we did not include gaps between observations, although we recognize that they are unavoidable and the observed light curves will have gaps of a few months each year (but see also section 3.5). Finally, we set the average photometric error at 0.01 mag. See Table 3.1 for a summary of the parameters of the simulated light curves for each survey.

With the observed properties of the time series, we proceeded to simulate DRW and DRW+sine light curves following the steps from Charisi et al. (2016).

The power power spectral density function (PSD) of DRW is

$$P(f) = \frac{4\sigma^2\tau}{1 + (2\pi\tau f)^2}, \tag{3.1}$$

where σ^2 is the variance of the light curve data points, τ is a characteristic DRW timescale, and f is the Fourier-space frequency. With the inverse Fourier transform of the PSD, we generate evenly sampled light curves (with $\Delta t = 1\,d$) using the prescription from Timmer & Koenig (1995), included in the python package `astroML` (Vanderplas et al., 2012; Ivezić et al., 2014). We downsample the data to match the desired sampling pattern of the survey setup described above and in Table 3.1. Next, we add Gaussian errors with zero mean and standard deviation equal to the average photometric uncertainty of the respective survey (Table 3.1).

For the set of simulations that also include SMBHB signals, we inject a sinusoid on top of the DRW light curve. This signal has the form

$$\mathbf{s}(t) = A \sin\left(\frac{2\pi}{P}(t - t_0)\right) \tag{3.2}$$

where A is the amplitude in magnitudes, P is the period of the sinusoid, and t_0 is a reference time. Both the period and the amplitude of the sinusoid can be linked to the parameters of the binary; the observed period is typically the redshifted orbital period of the binary, and if the periodicity is produced by relativistic Doppler boost, the amplitude A depends on the line of sight velocity of the secondary SMBH. Example time series with a simulated DRW-only process, and DRW+sine, can be

seen in Figure 3.1, where with blue (and red) data points we show the LSST-like (and CRTS-like) light curve.

We generated DRW and DRW+sine light curves for a wide variety of these five input parameters ($P,\ A,\ t_0,\ \sigma$, and τ). First, in order to test the Bayesian pipeline's ability to recover the model parameters, we choose values across an extreme range of τ corresponding to those used in Kozłowski (2017). Injected values are randomly selected from the range $\tau = [10^{-3}T, 15T]$, where T=10 yr is the nominal LSST observation baseline. However, for the model selection analysis, we restrict the values of τ to a realistic distribution derived from those presented in MacLeod et al. (2010). For σ, we draw values from a log-uniform distribution ranging from $[-1.6, -0.25]$, corresponding to a range of greater than an order of magnitude in σ, to encompass a wide range of DRW variability amplitudes similar to the range presented in MacLeod et al. (2010).

The periods of the injected sinusoids range from 30 days to 10 years. The maximum value is set by the LSST baseline, so that at least one full orbital cycle is observed. This wide range of periods covers all the potential SMBHBs that have GW frequencies detectable by PTAs. However, it does not include very high-frequency SMBHBs possibly detectable by the Laser Interferometer Space Antenna (LISA; Xin & Haiman 2021), which are expected to have periods of only a few days ($P < 1-2$ d). In section 3.5, we explore whether such short period binaries need a distinct strategy for detection, such as accounting for filter alternation and combining the multi-band data in a single light curve. Previous studies have required that at least 1.5 cycles (or more) of the periodicity be observed within the available baselines. We relax this

requirement to assess the ability to recover binaries in this regime and the resulting contamination with false positives. This is significant, since binaries evolve slower at large separations, and long-period binaries are expected to be more common. The reference time is set to any time between 0 and the period of the specific realization, which corresponds to a phase range of $[0, 2\pi]$. The amplitude is set to a value in the range $[0.05, 0.5]$ mag. These distributions of simulated values are summarized in Table 3.2.

3.3.2 Likelihood and MCMC Methods

For the DRW process defined in Equation 3.1, the covariance matrix S that determines the correlation between two data points at times t_i and t_j is given by

$$S_{ij} = \sigma^2 \exp\left(-\frac{|t_i - t_j|}{\tau}\right) \tag{3.3}$$

where σ^2 and τ are the same values defined above. The full covariance matrix is $C = S + N$, where $N = \mathrm{diag}(\sigma_{\mathrm{err}}^2)$ is the noise covariance matrix with σ_{err} the survey's photometric error. The DRW likelihood function marginalized over the mean of the light curve is given by

$$\begin{aligned} P(\mathbf{y} \mid \mathbf{p}) \propto |C|^{-1/2} \left|L^T C^{-1} L\right|^{-1/2} \\ \times \exp\left(-\frac{\mathbf{y}^T C_{\perp}^{-1} \mathbf{y}}{2}\right), \end{aligned} \tag{3.4}$$

with $\mathbf{y}$ the vector of the data (observed magnitudes) and L a vector of ones with a length equal to the number of data points, and

$$C_{\perp}^{-1} = C^{-1} - C^{-1}L\left(L^{T}C^{-1}L\right)^{-1}L^{T}C^{-1}. \tag{3.5}$$

For a detailed derivation we refer the reader to Kozłowski et al. (2010). The likelihood function for the DRW+sine model is given by

$$\begin{aligned} P(\mathbf{y} \mid \mathbf{p}) \propto & |C|^{-1/2}\left|L^{T}C^{-1}L\right|^{-1/2} \\ & \times \exp\left(-\frac{(\mathbf{y}-\mathbf{s})^{T}C_{\perp}^{-1}(\mathbf{y}-\mathbf{s})}{2}\right), \end{aligned} \tag{3.6}$$

with $\mathbf{s}$ a vector of the sinusoid $s(t) = A\sin(2\pi/P(t-t_0))$ sampled at the observed times.

We utilize Markov-Chain Monte Carlo (MCMC) methods for both parameter estimation and model selection. In particular, we sample the likelihood using a parallel tempering MCMC sampler called `PTMCMCSampler` (Ellis & van Haasteren, 2017). This sampler was developed for GW searches in PTA data sets, which employ similar Gaussian likelihoods for the analysis of the PTA time series. The main advantages of using this sampler is that it allows for easy implementation of custom jump proposals within the MCMC sampler and the pipeline developed here can easily be extended to multi-messenger searches of SMBHBs with joint PTA and time-domain data (see section 3.5).

In general, we use relatively uninformative priors for our MCMC analyses (either uniform or log-uniform), as summarized in Table 3.2. The priors typically

Parameter	Simulation Distribution	Prior
$\log_{10}\sigma$	Log-Uniform[–1.6, –0.25]	Log-Uniform[–1.6, –0.25]
$\log_{10}\tau$ (Wide Range)	Log-Uniform[0.56, 4.73]	Log-Uniform[0.56, 4.73]
$\log_{10}\tau$ (Realistic Distribution)	SkewNorm(3.0, 0.5, –1.4)	Log-Uniform[0.56, 4.73]
$\log_{10}P$	Log-Uniform[1.5, 3.5]	Log-Uniform[1.5, 3.5]
A	Uniform[0.05, 0.5]	Uniform[0.05, 0.5]
t_0	Uniform[0, 3650]	Uniform[0.05, 0.5]

Table 3.2: Simulation ranges for each of our five parameters, and prior shapes and ranges for our MCMC. Note that for model selection analyses, we simulate the realistic distribution of τ values derived from MacLeod et al. (2010).

mirror the distributions of simulated parameters described in subsection 3.3.1. We chose flat priors to avoid introducing potential biases, and to emulate an uninformed systematic search. More informative priors could be imposed for the DRW parameters; for example, MacLeod et al. (2010) found that σ and τ are correlated with properties of the AGN (e.g., the SMBBH mass, the luminosity, etc). Since in our simulated light curves we did not vary luminosity-related parameters (e.g., the observed magnitude) a fairly unrestricted search is more appropriate.

For each simulated light curve, we performed the MCMC analysis for two models (DRW and DRW+sine): The first uses the DRW likelihood from Equation 3.4 to search over only the two DRW parameters, σ and τ, and the second uses Equation 3.6 that also searches over the sinusoid parameters. To ensure convergence, each MCMC was conducted for 10,000 MCMC iterations, each resulting chain was thinned by a factor of 10, and the first 25% of the chain was trimmed to ensure burn-in. From these MCMC searches, we estimated the values of the two (or five) parameters that are most likely given each simulated light curve. The posterior distributions provided both median values and uncertainties for the parameters. From

the MCMC samples we also determined the value for set of parameters that maximized the likelihood. For each simulation, we also calculated a signal-to-noise ratio (SNR), where

$$\mathrm{SNR}^2 = \mathbf{s}^T \cdot C^{-1} \cdot \mathbf{s}. \tag{3.7}$$

Here, $\mathbf{s}$ is the vector containing the input signal and sampled at the simulated timestamps, and C^{-1} is the inverse of the DRW covariance matrix (Equation 3.3).

We used the outcome of the two MCMC searches (DRW, and DRW+sine) to perform Bayesian model selection using the Bayes Information Criterion (BIC)

$$\mathrm{BIC} = k \ln(n) - 2 \ln(\widehat{L}), \tag{3.8}$$

where k is equal to the number of free parameters, n is equal to the number of data points in the light curve, and $\widehat{L}$ is the maximum likelihood value (Liddle, 2007). The BIC provides a simple metric through which to compare our two models, and avoids overfitting the data by accounting for the number of parameters in the model. When selecting among multiple models, the one with the smallest BIC is usually preferred. Here we selected the preferred model by comparing the BIC values for the DRW-only search to that of the DRW+sine search by introducing

$$\Delta\mathrm{BIC} = \mathrm{BIC}_{\mathrm{DRW}} - \mathrm{BIC}_{\mathrm{DRW+sine}}. \tag{3.9}$$

A lower value of ΔBIC indicates more support for the DRW+sine model. In general, evidence for the DRW+sine model can be considered positive for $-2 > \Delta\mathrm{BIC} > -6$,

and strong for $\Delta\text{BIC} < -6$ (Kass & Raftery, 1995). Here, we defined our threshold to claim a detection of a sinuosidal signal as $\Delta\text{BIC} = -2$. Using this threshold, we sorted each result into one of four categories:

- **True Positive**: A sinusoid was injected and the DRW+sine model was preferred.
- **False Negative**: A sinusoid was injected, yet the DRW-only model was preferred.
- **False Positive**: No sinusoid was injected, yet the DRW+sine model was preferred
- **True Negative**: No sinusoid was injected, and the DRW-only model was preferred.

In an idealized search we would have only true positives/negatives and no false positives/negatives, but typically one needs to compromise and balance the rate of detection of true signals with the contamination of a few false positives. One of the main goals of this analysis is to constrain these rates for current and future survey capabilities. We note, however, that these rates refer to our specific method of Bayesian model selection and cannot be extended to existing samples of SMBHB candidates, since these candidates were selected with completely different methods, as we explain in section 3.5.

3.4 Results

We assessed our ability to identify periodicity in AGN light curves by simulating DRW and DRW+sine light curves and performing Bayesian model selection. First, we tested how the MCMC algorithm performs in constraining the parameters of each model independently. Subsequently, we determine the performance of the model selection method by calculating the true and false-positive rates, and characterizing their dependence on the signal and noise parameters.

3.4.1 Parameter estimation

3.4.1.1 DRW model

For both our CRTS-like and LSST-like simulations, we simulated 500 DRW light curves with properties as described in subsection 3.3.1 and conducted the MCMC analysis for the DRW likelihood (Equation 3.4). In Figure 3.2 we show the median values of σ and τ as a function of the respective input values for the LSST light curves. We note that the parameter estimation for the DRW model shows very similar trends for the CRTS-like light curves.

We saw that low values of τ are recovered accurately, while high values were poorly constrained. This is a known limitation in DRW studies. For instance, Kozłowski (2017) found that for τ to be well-recovered, the baseline of the light curve must be at least ten times greater than τ ($\tau \leq 10T$). In that study, the authors demonstrated this effect with simulated light curves for a fixed parameter σ. Here, by varying the values of σ for each simulation, we demonstrate that this

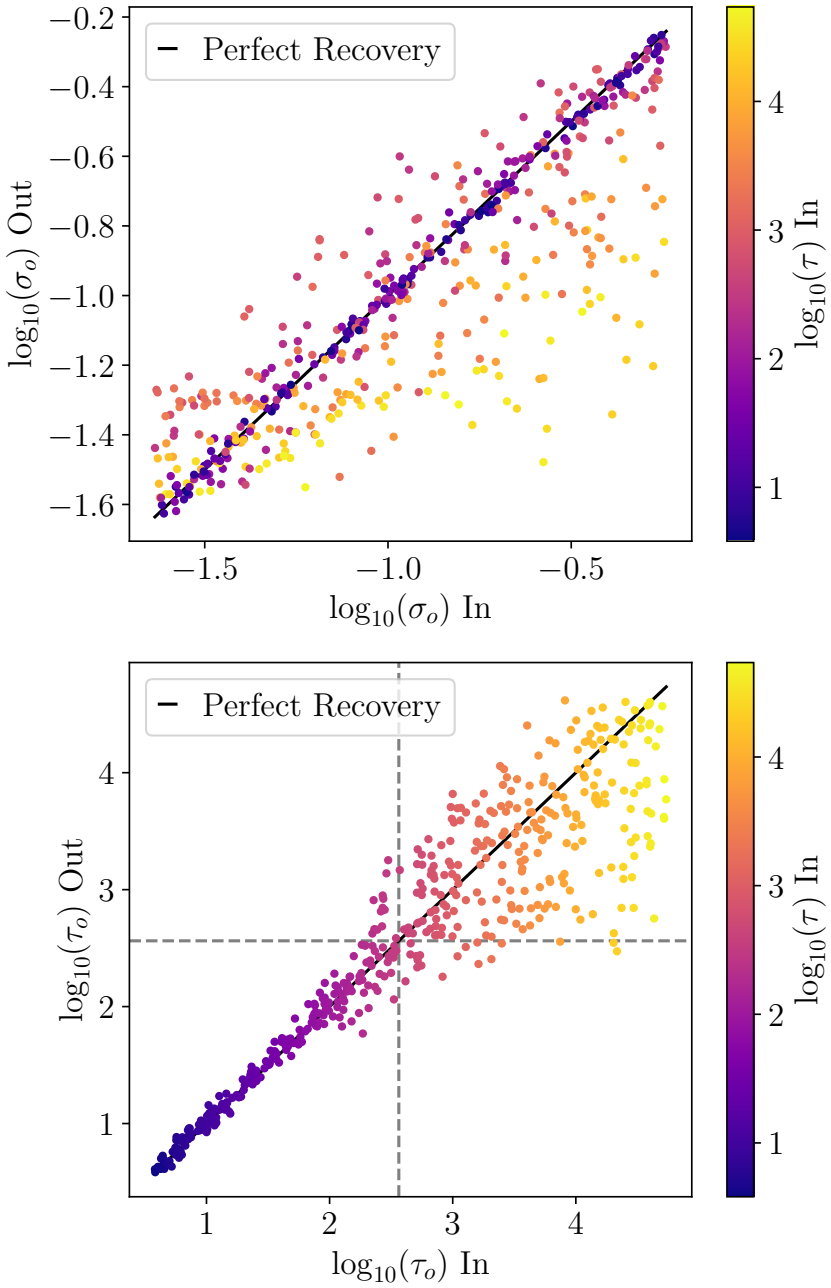

Figure 3.2: Parameter estimation capability of our MCMC methods for LSST light curves with a DRW process, colorized by the simulated value of τ. For low values of τ, both parameters are recoverable as expected. However, for very high values of τ, both are unlikely to be constrained accurately.

limitation affects the recovery of σ as well. In Figure 3.2 we colorized the data points by the input value of τ. We observed that for light curves with large values of τ, where τ is not constrained (orange-yellow points), the algorithm fails to recover the input value of σ. On the other hand, for light curves with small values of τ (purple points), the recovery of both σ and τ is very accurate.

3.4.1.2 DRW+Sine Model

Once we confirmed that the DRW parameters can be recovered by the MCMC methodology (within already known limitations), we expanded our search to also include the sinusoidal signal representing an SMBHB. We repeated the 500 simulations of both CRTS- and LSST-like DRW light curves, and added a randomly generated sinusoid to the data. This was then searched with a 5-parameter MCMC using the DRW+sine likelihood from Equation 3.6.

Figure 3.3 summarizes the recovery capability of the DRW+sine model in LSST-like simulations, color-coded by the SNR of the input signal. We note that even though it is preferable to sample the likelihood in terms of a reference time t_0, we present results converted to an initial phase $\phi_0 = 2\pi t_0/P$ to avoid potential biases or correlations with the period. In general, for LSST-like simulations, we recovered strong sinusoids (SNR>5) extremely accurately, with 75% of these signals having both their periods and amplitudes recovered accurately (i.e., within the 90% credible region of the posterior distribution). 64% of these strong signals had all parameters recovered accurately, and 80% of signals with SNR>5 had both DRW

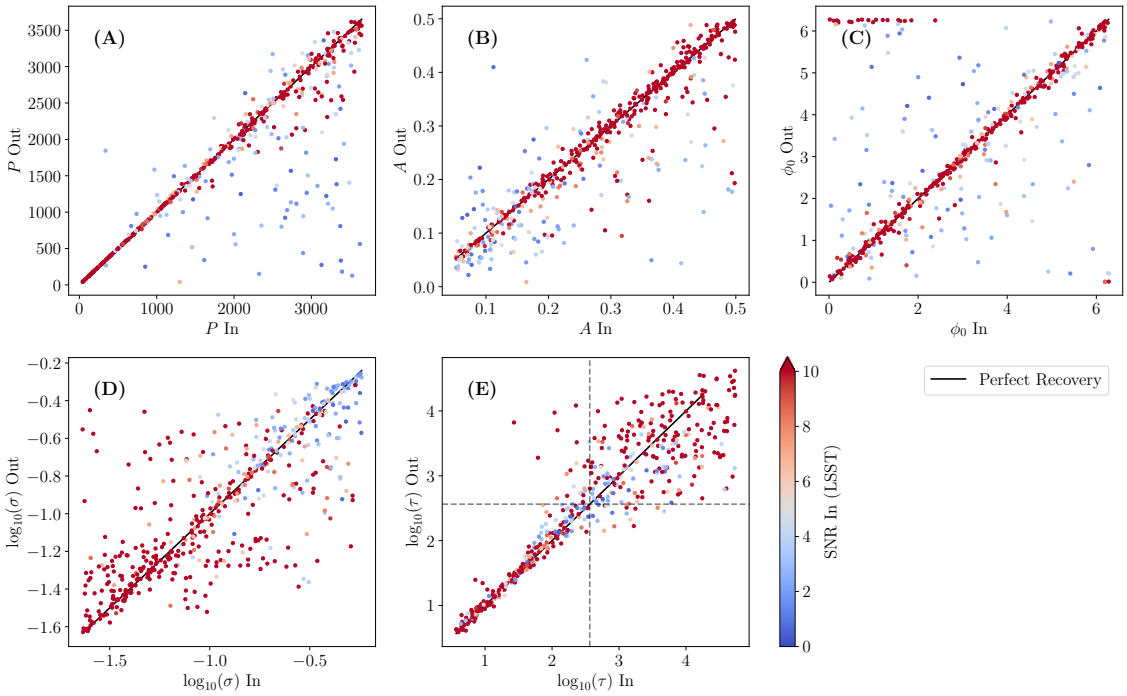

Figure 3.3: Recovery capabilites for the three sinusoid parameters (P, A, and ϕ_0) and two DRW parameters (σ, and τ), as demonstrated by our LSST-like light curve simulations. Color axis represents the SNR of the input signal, with red points being strong signals (SNR>5). The sinusoids in these simulations were able to be recovered extremely accurately, and the DRW parameters were recovered to the extent we expect from noise-only simulations.

parameters recovered accurately. These proportions are nearly identical for the CRTS-like simulations, but for a smaller fraction of light curves having a sinusoid with SNR>5. This method is successful at recovering sinusoids with a wide range of injected parameters. It is important to note that the algorithm accurately recovered periods from 30 days to 10 years, and it was not required for all light curves to cover more than two cycles of the sinusoid for their parameters to be recoverable, as may be expected based on analyses by Vaughan et al. (2016). We further explore the longer period regime in subsection 3.4.2.

The DRW parameters σ and τ were recovered with the same accuracy as in the DRW-only search, even in the presence of the sinusoid. We also saw the same limitations in recovering long τ and resulting limitations in recovering σ for this subset of light curves. However, our inability to constrain the DRW parameters in certain light curves was not propagated to the recovery of the parameters of the periodic signal. Additionally, the highest σ values are at near the maximum of the observed quasar population, and will be fairly rare in reality (MacLeod et al., 2010).

3.4.2 Model Selection

Next, we used a Bayesian model selection, described in detail in section 3.3, to select quasars with periodic variability. With simulated DRW and DRW+sine light curves, we traced the algorithm's effectiveness. We considered two distinct surveys (CRTS and LSST, reflecting current and future capabilities of time-domain surveys) to explore how the light curve quality and properties affect the detection rates of

this method.

First, we simulated 1,500 DRW light curves, added a randomly generated sinusoid and then applied our model selection scheme. In Figure 3.4 we show the true positive rate of periodic signals in the presence of DRW noise, as a function of the input parameters P, A, σ, τ, and ϕ_0. Here, we define the true positive rate as the number of detected periodic signals (true positives), divided by the total number of simulated DRW+sine signals (condition positives). In each bin, the associated uncertainty of the rate is calculated with a binomial proportion confidence interval (Newcombe, 1998), where the rate can be considered as

$$\frac{n_S}{n} \pm \frac{z}{n\sqrt{n}}\sqrt{n_S n_F}, \tag{3.10}$$

where n is the number of trials with n_S successes and n_F failures, and z is the $1-\alpha/2$ quantile of a normal distribution (for a 95% confidence interval, $\alpha = 1 - 0.95$). We observed that our ability to detect periodicity depends both on the parameters of the sinusoid and the intrinsic DRW variability. As expected, the true positive rate increased for high sinusoidal amplitudes and was independent of the initial phase. The true positive rate was highest for short sinusoid periods, however, it was non-zero even for periods equal to the observation baseline, which is an unexpected improvement from Vaughan et al. (2016), which showed a requirement of > 2 cycles for a sinusoid to be differentiated from a stochastic process. The true positive rate decreased for increasing input σ; therefore, when the noise contribution became more significant, it hindered the periodicity detection, as expected. We also saw

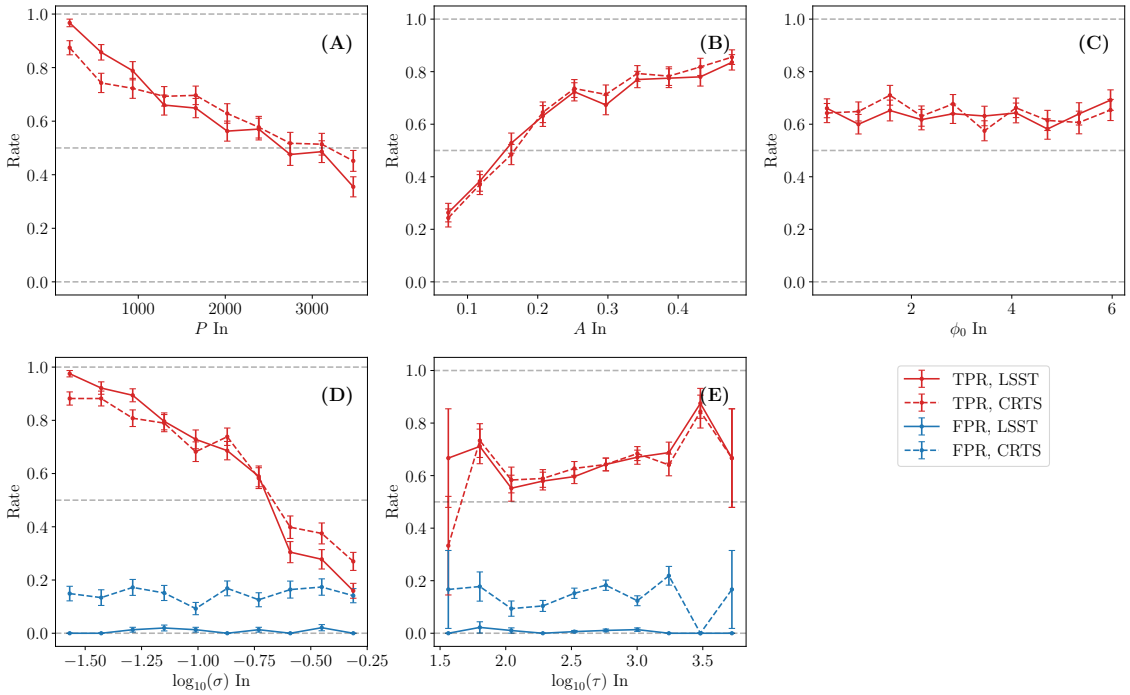

Figure 3.4: True positive rates (red) and false positive rates (blue) for LSST-like (solid lines) and CRTS-like (dashed lines) simulations, shown as a function of the input values of each parameter in the simulations. Note that false positive rates are only shown as a function of the two DRW parameters, as there are no input sinusoids present in the false positives. The rates in each parameter bin are shown with associated uncertainties.

in subsection 3.4.1 that high values of σ (or equivalently low SNR) resulted in an innaccurate estimation of the parameters. On the other hand, τ did not seem to have a significant effect on the detection rate, despite the inability to constrain large values of τ, with the true positive rate slightly increasing for longer τ. Surprisingly, the overall true positive rate varied only slightly between the two surveys.

Next, we explored how the periodicity detection rate varies as a function of the periodic parameters normalized by the noise parameters. In Figure 3.5 we present the input ratios of A/σ against P/τ, colorized by the resulting ΔBIC. In the side panels, we track the fraction of recovered sinusoids (true positive rate) as a function

of either A/σ (for the vertical panel) or P/τ (for the horizontal panel), again with the associated binomial uncertainty marked in each bin.

As can be expected, the fraction of binaries recovered was highly correlated with A/σ. This value can be considered similar to an SNR; we saw that even though it was not absolutely necessary that $A > \sigma$ for a periodic signal to be detected, the detection rate dropped to $\sim$50% when the amplitude of the sinusoid was comparable to the standard deviation of the DRW noise. The recovery fraction also clearly depends on the value of P/τ, albeit less strongly than with A/σ. That is, even without considering A/σ, the periodic signal is more likely to be detected (i.e., ΔBIC is lower) for smaller ratios of P/τ. In terms of detectability, we see that all binary signals were identified for small values of P/τ, whereas the true positive rate is $\sim$75% when P and τ are comparable and is further reduced to 50% for larger values. This is consistent with our findings in Figure 3.4, where we see that detectability increases for small periods and for larger values of τ, although the latter correlation is weaker. The correlation of the true positive rate with the period seen in Figure 3.4 is fairly intuitive; a relatively weak signal can be confidently detected if the period is short and enough cycles are repeated within the data. However, it is somewhat less obvious what drives the correlation with P/τ. One potential explanation is that it may be easier to detect a periodic signal if the two characteristic timescales (P and τ) of the light curve are fairly distinct. Otherwise, if the values are similar, they may be misidentified by the model selection process (e.g., see subsection 3.5.3).

As a counterpoint to the previous analysis, we subsequently simulated 1,500 DRW-only signals, ran our model selection pipeline, and calculated the false positive

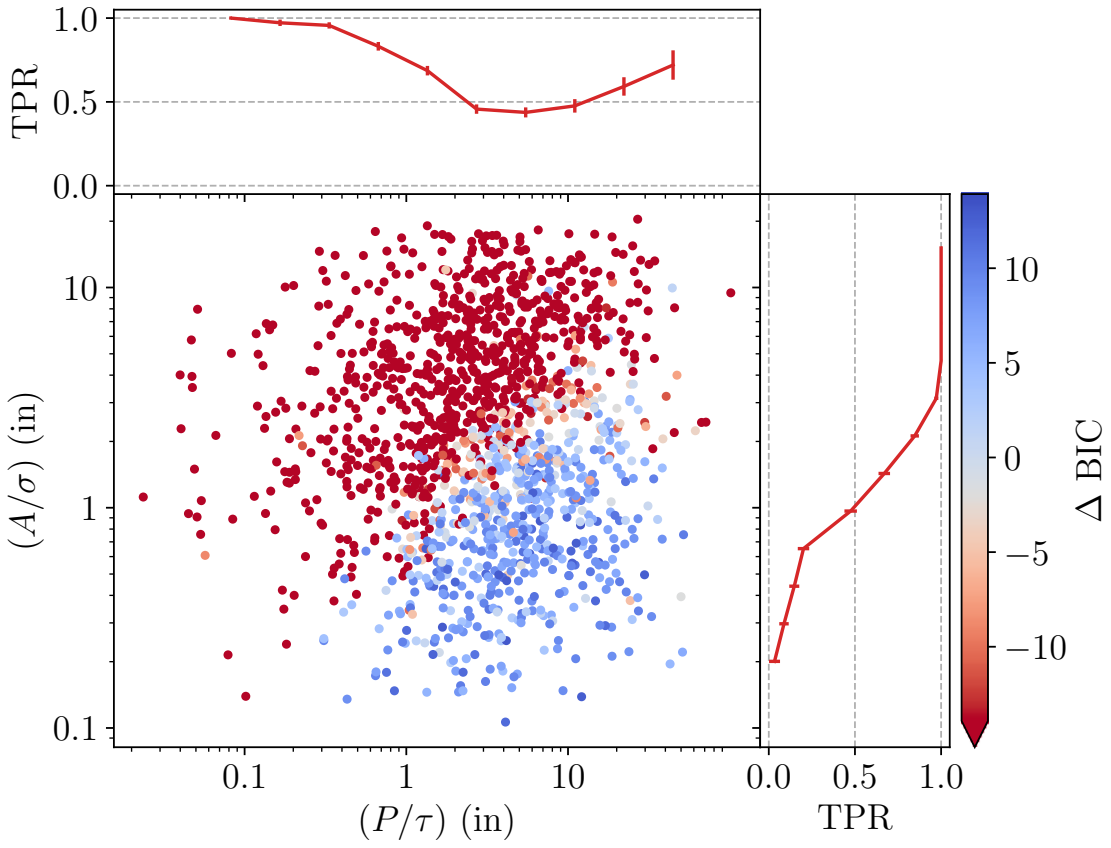

Figure 3.5: True positive (red) and false negative (blue) signals tend to lie in particular areas of parameter space in idealized LSST-like searches, when quantified by the ratios of A/σ and P/τ. The net true positive rates integrated over A/σ and P/τ are shown in the upper and right panels, respectively, with associated uncertainties.

rate for our same detection method. This represents a scenario in which only DRW processes are occurring, and either and an SMBHB is not present in the target, or it is not influencing the AGN light curve. Here, the false positive rate is defined as the number of DRW-only signals identified as periodic (false positives), divided by the total number of DRW simulations (condition negatives). In Figure 3.4, we show the false positive rate for both surveys as a function of input σ and τ with blue curves, and again using the associated uncertainties calculated with Equation 3.10. We see that the false positive rate is significantly higher in CRTS, reflecting the lower measurement precision and sampling rate of the light curves, whereas in LSST the false positives are almost negligible. The overall false positive rate for CRTS-like simulations is 14%, while for CRTS-like simulations, it is 0.8%, an improvement of over an order of magnitude. Interestingly, the false positive rate does not show any significant trend with σ or τ. This indicates that any combination of the DRW parameters is equally likely to produce a signal that can be misidentified as a sinusoid. We also observed that the false positive rate did not increase for large recovered sinusoid periods, as was suggested by Vaughan et al. (2016); this is likely due to the use of a DRW+sine model, as opposed to a pure sinusoid.

In the above we examined the true-positive and false-positive rates as a function of the input parameters of the noise and the signal, considering a quasar to be periodic if $\Delta\mathrm{BIC} \leq -2$. However, as is obvious from Figure 3.5, these rates would be different had we chosen a different detection threshold. This is typically quantified by a receiver operating characteristic (ROC) curve, which we construct in Figure 3.6. More specifically, we show the true positive rate against the false positive

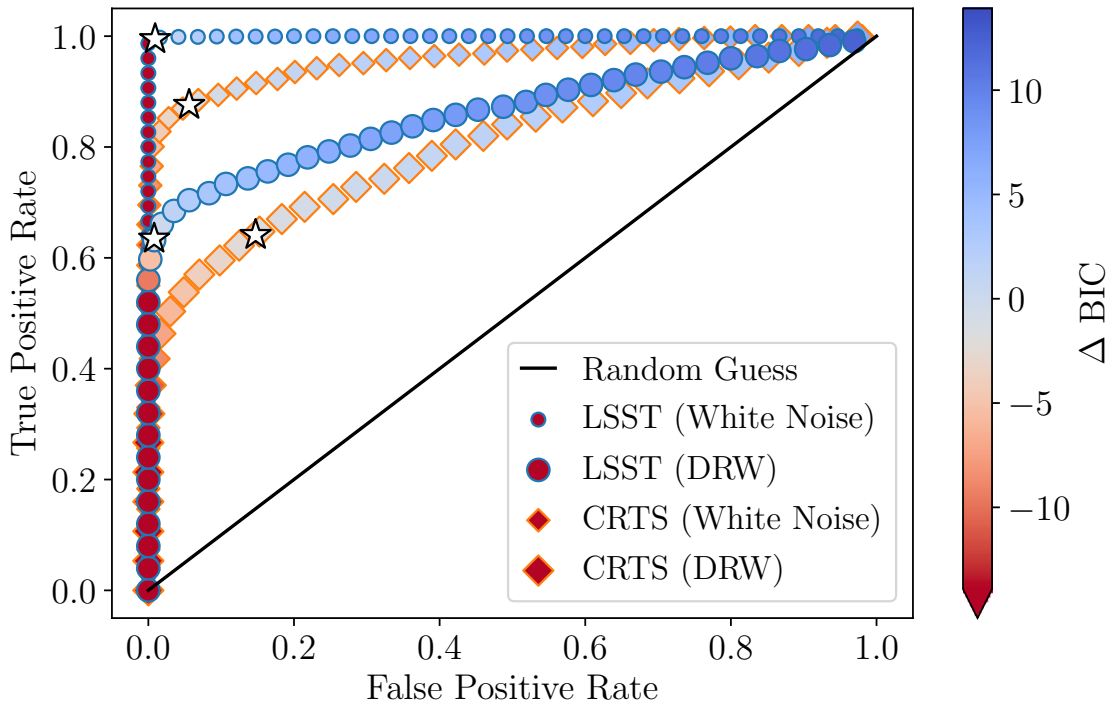

Figure 3.6: ROC curves for CRTS-like (orange-lined diamonds) and LSST-like (blue-lined circles) light curves. Overall, LSST can be seen to perform better than CRTS at selecting the correct model. For comparison, with our standard $\Delta\text{BIC} = -2$, the true positive rate is 64%, and the false positive rate is 0.8%. Stars in the curves represent the point where $\Delta\text{BIC} = -2$. Additionally, the model selection is significantly improved when white noise (small points) is present instead of a DRW process (large points), indicating that red processes are indeed a significant hindrance.

rate color-coded with the threshold value for periodicity detection (i.e. the maximum ΔBIC required for detection of a sinusoid within the light curve). We remind the reader that a smaller ΔBIC means stronger support for the binary model.

In Figure 3.6, we indicate the current threshold of $\Delta\text{BIC} \leq -2$ with a star. The corresponding true positive rate is $\sim$64% for both surveys, whereas the false positive rate is $\sim$0.8% for LSST and $\sim$14% for CRTS. We note that even though we chose this particular threshold following standard practices for model selection based on BIC, it turns out to be a reasonable threshold for both CRTS and LSST.

In fact, for a survey such as CRTS, it is sensible to set the threshold at a level that maximizes true positives, even if this allows some false positives. High-quality light curves are available for $\sim 10^5$ quasars, and given that SMBHBs are relatively rare, it is manageable to pursue follow-up observations to distinguish genuine binaries from interlopers for all candidates. For LSST, on the other hand, it is critical to minimize false positives. LSST will observe millions of quasars, and follow-up of candidates needs to be significantly more selective. The colorization of Figure 3.6 also illustrates the much larger range of ΔBIC values in an LSST-like survey, as compared to CRTS. This results in a much larger number of strongly preferred signals, which will allow for a dramatically more effective ranking system for the follow-up of binary candidates.

We also show the respective ROC curve, for both CRTS and LSST, for the case of periodicity on top of white noise. This allowed us to test the hypothesis that the classifier performs sub-optimally due to the covariance between the sinusoid and DRW. We repeated our simulations with 1,500 simulations containing white noise and a sinusoid and another 1,500 with only white noise, and performed an identical model selection procedure. Nearly all of the sinusoids were identified with accurately estimated parameters, across the entire parameter space. The ROC curve for LSST is excellent, with close to 100% recovery for true periodic signals and almost 0% false detections. The ROC curve is slightly worse for CRTS due to the lower data quality. This indicates that, without the red DRW noise process included, there was no confusion, allowing the sinusoids to be identified accurately. The white noise realization of the population, albeit unrealistic, demonstrates that the limiting factor

	CRTS	LSST
DRW	0.809	0.864
White Noise	0.960	0.999

Table 3.3: Area under curve (AUC) values for each of the ROC curves shown in Figure 3.8, including those with or without a DRW process. LSST is expected to be a much more sensitive and reliable survey for the identification of periodicities induced by SMBHBs.

in detecting quasar periodicity is primarily the stochastic DRW variability.

Finally, we quantitatively evaluated the performance of our method in each survey by computing the area under the ROC curve, also known as the AUC value. In general, a larger AUC value indicates a better performing classifier, as this metric equals the probability that the classifier will rank a positive simulation better than a negative one (Fawcett, 2006, i.e. the probability that we will calculate a lower ΔBIC if a sinusoid is present). In Table 3.3, we summarize the AUC values for CRTS and LSST both for idealized white noise simulations and for the more realistic case that includes DRW variability. The white-noise-only ROC curve for LSST has near-perfect AUC value of 0.99, indicating that the DRW process can mask a sinusoid from the model selection process, while white noise cannot.

So far we have presented our results with respect to the input signals. However, in real observations, we will not know the true parameters of the signals, and thus will be required to base our model selection conclusions on the output parameters of the MCMC method. In Figure 3.7, we present the recovered parameters A/σ versus P/τ in order to map the parts of parameter space where simulations with and without a sinusoid in addition to DRW noise are more likely to lie. For instance, if the DRW+sine model returns $A/\sigma > 1$ in LSST, it is highly likely to be a true

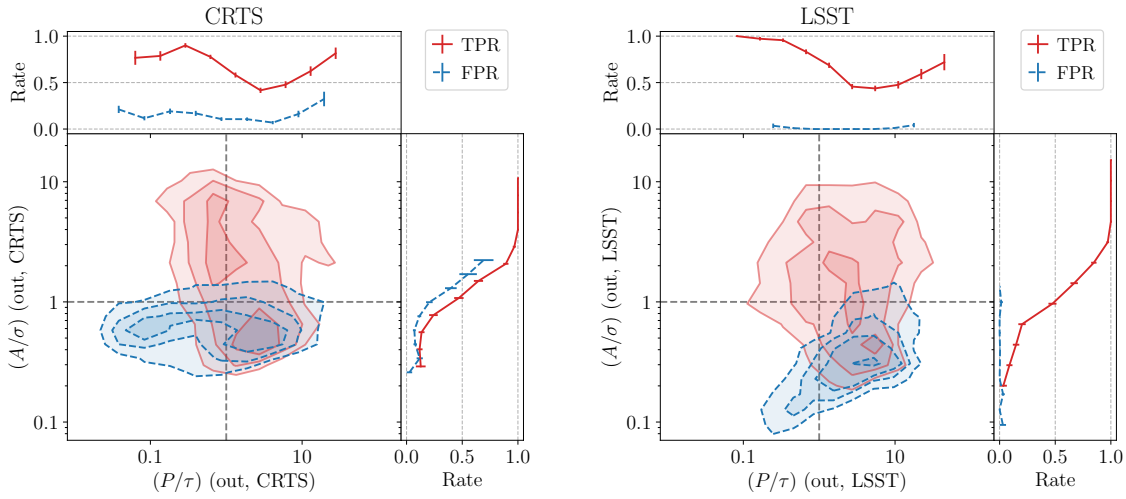

Figure 3.7: Simulations of DRW light curves with (red) and without (blue) a sinusoid lie in regular regions of parameter space when parametrized by the ratio of A/σ and P/τ. This makes apparent the cause of the location of false negatives in Figure 3.5. It is also clear that more signals will become detectable in future surveys as cadences and baselines improve.

detection regardless of P/τ. In the weak signal regime $A/\sigma \leq 1$, the two populations overlap, although given the low number of false positives, a detected signal is more likely to be genuine periodicity. In CRTS-like survey, it is more challenging to derive a conclusion about the validity of the detection based on the inferred parameters of the light curve, due to the higher rate of false positives. Overall, identifying periodicity in the strong signal regime (e.g., $A/\sigma > 2$) can boost our confidence that the detection is real, since no true negatives lie in this area.

One way to quantify the distinction between the populations with and without a simulated sinusoid is with the Mahalanobis distance (Mahalanobis, 1936). This metric measures the distance between a point and a distribution, measured in standard deviations of the distribution, while accounting for correlations between the data points. For the CRTS-like observations, the median Mahalanobis distance

between the two populations is 0.73, while for the LSST-like survey, this median distance increases to 0.97. This indicates that, in next-generation surveys, the populations of AGN with and without sinusoidal variations will become even more clearly resolved.

3.5 Discussion

3.5.1 Previous Work

In this paper, we simulated CRTS-like and LSST-like light curves and used a Bayesian model selection to assess our capability to detect SMBHBs in time-domain surveys. This is the first study that explores the parameter space of sinusoidal binary signals in the presence of a DRW process, employing an array of idealized simulated data. This allowed us to examine both the detectability/completeness of binary signals and the contamination of a sample of candidates with false detections.

We found that the sample of periodic quasars is expected to be fairly incomplete for longer period binaries and for binaries that cause weak periodic modulations in the brightness of the AGN compared to the DRW variability. This limitation is caused by the stochastic variability of quasars, since in the presence of only white noise, almost all the periodic signals would be detectable with nearly zero contamination. These results are independent of the time-domain survey setup. On the other hand, the false-positive rate is higher in the CRTS-like light curves compared to LSST. This suggests that the contamination of the samples of SMBHB candidates depends on the quality of the data. The reduced false-positive rate in LSST

is extremely encouraging for future searches for candidate signatures of SMBHBs. This is particularly important, since LSST will observe at least 20 million quasars, and a high false-positive rate would render follow-up studies of SMBHB candidates nearly impossible.

We emphasize that even though our results provide an excellent qualitative picture of limitations and detectability trends as a function of the signal and noise parameters, they cannot be directly applied to determine the number of false positives in existing samples of SMBHB candidates (Graham et al., 2015b; Charisi et al., 2016; Liu et al., 2019). These candidates were chosen with a different methodology, and likely suffer from distinct biases that cannot be captured by our analysis. We have already observed that with our algorithm, changing the detection threshold would change the true- and false-positive rate. This demonstrates that it will be enlightening for future systematic searches for quasar peridocity to use simulated light curves to carefully construct an ROC curve, as in our study, to highlight the effectiveness of the selection criteria of the search, given the specific survey properties.

We also note that to date, a Bayesian model selection method has not applied in an extensive search for binaries. This is unsurprising, as this method is computationally demanding, and thus for a large sample of quasars (of order 10^5 for CRTS and 10^7 for LSST) it is practically impossible. Our idealized data sets require a few hours of CPU time per light curve to complete the model selection analysis, and realistic data, with a larger number of associated parameters, will expand this requirement. Therefore, this method may be applied in combination with some

other classifier which will make an initial pre-selection, and therefore significantly reduce the size of the sample. Such a complementary method will filter out most non-periodic quasars, and thus the main requirements for it are speed and a high true-positive rate, rather than a perfect false-positive rate.

However, several Bayesian-model-selection algorithms have been used in multiple studies to validate (or invalidate) the periodicity for one of the most prominent candidates, quasar PG1302-102 (Graham et al., 2015a; D'Orazio et al., 2015; Vaughan et al., 2016; Liu et al., 2018; Zhu & Thrane, 2020). It is intriguing that the results of these studies are not in complete agreement, neither for the best-fit parameters nor for the preferred model. This is potentially due to choices made in these analyses; for example, Vaughan et al. (2016) introduced an extra parameter to account for poorly estimated photometric errors, Liu et al. (2018) binned the light curves in wide bins of 150days, and D'Orazio et al. (2015) fixed the parameters of the DRW model. This clearly illustrates the complexities of observed data sets that may not be reflected in idealized simulations, such as the ones we presented in this study.

3.5.2 LSST Observing Strategy and Future Improvements

To assess the prospects of detecting SMBHBs in LSST, we simulated light curves with semi-regular sampling (evenly sampled, but also adding a Gaussian error to the timestamps). As a conservative scenario for the wide-fast-deep survey, we chose a cadence of seven days, but in reality, observations of the same source

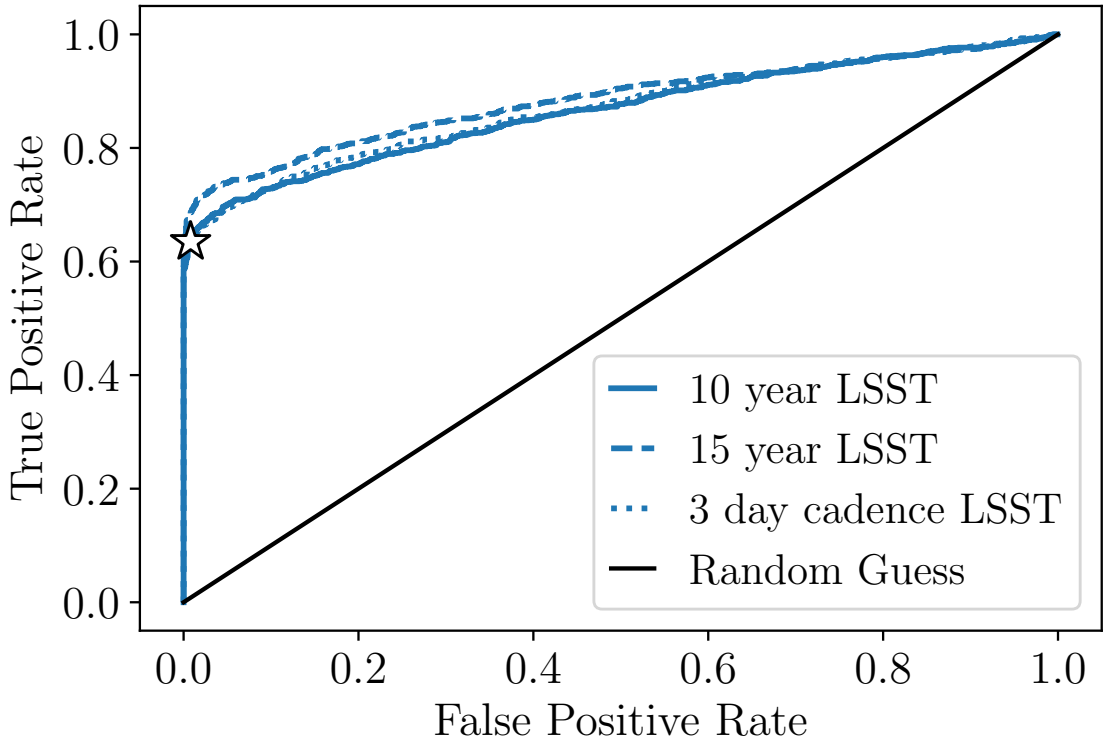

Figure 3.8: ROC curves for variations on our nominal LSST simulations (solid curve), including increased observation baseline (dashed curve) and increased obseration cadence (dotted curve).

may repeat more often. We explore two additional optimistic scenarios. First, we increase the cadence to three days, and second, we extend the observation baseline to 15 years while keeping the cadence at seven days. We simulate both DRW and DRW+sine light curves and repeat the Bayesian model selection.

In Figure 3.8, we present the ROC curve for these two scenarios, compared with the initial LSST setup as reported in Table 3.1. The AUC for these two scenarios are 0.834 (3 day cadence, 10 year baseline) and 0.883 (seven day cadence, 15 year baseline) compared to 0.864 for our typical LSST simulations with a 7 day cadence and 10 year baseline. We see both from the figure and the AUC values that increasing the baseline has a positive impact in our periodicity search, allowing us

to better constrain longer τ and detect longer periods at higher rates. This confirms that in preparation for LSST, it is advantageous to prepare strategies that will allow us to extend the LSST light curves with already existing data by combining data from multiple surveys, such as in Liu et al. (2018). On the other hand, increased cadence does not improve our results. This is expected given that the minimum searched period is set to 30 days. We note that the higher cadence will likely significantly affect the search for short period binaries, which are potential LISA sources (Xin & Haiman, 2021). In this work, we did not examine the shortest end of the period distribution, because for very short periodicities, a different method may be necessary; for example, the deep coverage and more frequent sampling planned for LSST Deep Drilling Fields will certainly be beneficial for this type of source.

Moreover, even though the finalized survey strategy has not yet been decided, we recognize that our simulations are idealized for several reasons. First, our light curves do not include gaps of no observations, which are inevitable in a real survey, since the sources will be obscured by the sun or low on the horizon. We also used an average photometric uncertainty for all of the simulated light curves, even though the photometric errors are in reality magnitude-dependent. Lastly, even though the observations in the deep-wide-fast survey will repeat semi-regularly, they will rotate among six narrow-band photometric filters from visit to visit. Therefore, if we consider light curves in only one photometric band, they will be significantly more sparse with $\sim$10 observations per year. The preferred route would be to combine the data in a single multi-band light curve. In fact, a multi-band periodogram has been developed for this purpose (VanderPlas & Ivezić, 2015). However, for quasars, this

process is more complicated, due to their color-dependent variability, which must be taken into account.

In future work, we intend to address several of the limitations of this current idealized study. In the near future, LSST's observing strategy (such as cadence, epoch/gap length, order of filter alternation and frequency of observations in each photometric band, etc.) will be finalized. Projections of the final survey have begun to be released (LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021), and as more accurate data previews, such as LSST Data Preview 0,[2] become available over the next year, we will modify our simulations to include all the above decisions to more accurately reflect the full capabilities of the survey. To do so, we must also incorporate the magnitude dependence of the photometric errors. For this, we will simulate a more realistic quasar population, based on the quasar luminosity function and incorporate correlations of the DRW parameters with the AGN properties (MacLeod et al., 2010).

In addition to improvements to the simulated observation strategies, in future work it will be critical to include a more realistic binary population and an advanced model for quasar variability. More specifically, our analysis (and most searches for periodicity) assume that quasar variability is described by a DRW model. This model, albeit successful, comes with its own limitations. A future study will include advanced noise modeling and employ a continuous-time autoregressive moving average model, which also includes quasi-periodic oscillations. For the population of SMBHBs, we randomly drew the periods and amplitudes from (log)-uniform dis-

[2] https://rtn-001.lsst.io/

tributions. However, binary evolution models suggest that binaries spend more time at larger separations (and longer periods) and should be more common than short-period binaries, while the amplitudes can be linked to the orbital properties of the binary (e.g., mass-ratio, and inclination for relativistic Doppler boost). We also modeled binary signals with pure sinusoids, which, while a decent approximation for a circular binary dominated by Doppler boost variability, real binaries can produce more complicated signatures. For instance, if the periodicity arises from periodic accretion or if the binary has an eccentric orbit, the light curves will significantly deviate from sinusoidal.

3.5.3 Covariance of Timescales

We found that the stochastic DRW noise hinders the detection of the deterministic signal of a SMBHB. One potential reason is the covariance between the parameters of the signal and the noise. For instance, both the amplitude of the sinusoid and the DRW σ determine the overall SNR of the light curves. Unsurprisingly, we saw that our ability to detect sinusoidal variability increases when σ is small and A is large, and vice versa. The covariance of the characteristic timescales P and τ is less obvious, so we explore this issue in more detail below.

First, we examined our results for potential correlations when we fit for the incorrect model (i.e. injected DRW+sine using the DRW likelihood from Equation 3.4). Searching a light curve that has a sinusoid injected with a DRW-only model will result in a biased recovery of τ, as can be seen in Figure 3.9, where the

recovered τ value is related to the injected period. For LSST-like simulations, this was best fit with a quadratic function where

$$\log_{10}\tau = -0.72\,(\log_{10} P)^2 + 5.46\,(\log_{10} P) - 6.21. \tag{3.11}$$

However, for CRTS-like simulations, this covariance is best fit with a linear function

$$\log_{10}\tau = 1.14\,(\log_{10} P) - 0.76. \tag{3.12}$$

Kozłowski et al. (2010) found a similar effect when they applied the DRW formalism to periodic stellar light curves (e.g., their Fig. 12 and the related discussion). We also confirm their finding that these correlations are sensitive to the light curve properties, since we find a different correlation in our CRTS and LSST light curves. When τ is fit in conjuction with the periodicity (i.e in the DRW+sine model), this bias is resolved.

3.5.4 Prospects for Multi-Messenger Observations

Sub-parsec SMBHBs have remained a missing piece in the puzzle of hierarchical structure formation despite decades-long observational efforts seeking their detection. The upcoming decade is expected to bring tremendous improvements both in electromagnetic observations and in GW searches. PTAs may be on the verge of detecting the GW background from a population of unresolvable SMBHBs (Arzoumanian et al., 2020a). The detection of individually resolvable SMBHBs is

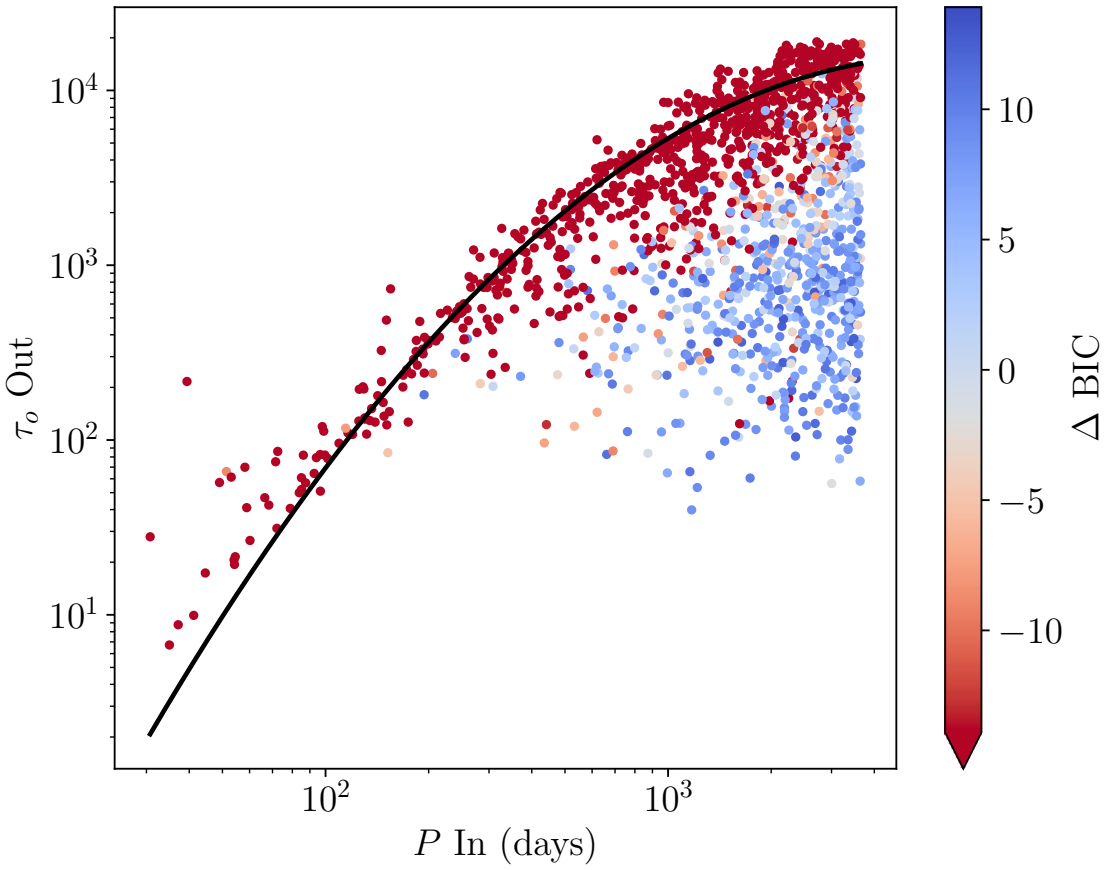

Figure 3.9: If a light curve is simulated to contain both a sinusoid and a DRW process, but is modeled with only noise, significant confusion can occur. The periods of recoverable sinusoids (red) can be confused for the DRW timescale τ if only noise is modeled.

expected to follow soon after. Addtionally, LSST will provide a revolutionary data set for searches targeting SMBHBs. In this analysis, we have demonstrated that the unprecedented quality of the light curves will minimize the false detection rate, while the large sample of quasars will likely lead to at least a few confident detections. Some of these binary candidates will have periods of a few months to a year, allowing us to see the periodicity repeat for many cycles.

This opens the possibility of combined GW and electromagnetic observations of SMBHBs. Liu & Vigeland (2021) found that incorporating priors from electromagnetic observations in the GW analysis boosts the detectability of binaries and improves parameter estimation of continuous GW searches. Arzoumanian et al. (2020b) showed that having a candidate to target significantly improves GW-derived upper limits on the binary chirp mass. Therefore, it is logical for GW searches to specifically target SMBHB candidates identified in time-domain surveys. Producing a large number of high-quality electromagnetic SMBHB candidates from LSST will provide a wealth of candidates to search for in PTA data. This population of candidates will also provide critical information about the population of SMBHBs that create the stochastic gravitational wave background, which, as stated above, may be detectable by PTAs extremely soon (Pol et al., 2021).

3.6 Conclusions

Using extensive simulations of time-domain observations of AGN, coupled with a Bayesian model selection and parameter estimation framework, we have explored

the capabilities of current and future surveys for SMBHB identification. In particular, we simulated quasar light curves with DRW variability with a realistic distribution of σ and τ, as well as binary light curves with sinusoidal variability on top of a DRW process including a wide range of periods and amplitudes. We explored the likelihoods of the respective models with an MCMC analysis, and determined the preferred model using the BIC. Our findings are summarized as follows:

- Our ability to detect periodicity on top of DRW variability depends on the parameters both of the sinusoid and of the noise. Short periods and high amplitudes are found at higher rates, whereas light curves with significant noise contribution (high σ) are recovered at lower rates. The input phase and τ do not appear to affect the detection rate.

- While our ability to discover long-period signals is decreased, about 50% are recoverable. This is significant, because longer-period SMBHBs are expected to be more common.

- The true positive rate is similar in both surveys.

- The incompleteness of the detectable binary signals is intrinsic due to the stochastic variability of quasars. In the presence of white noise, all periodic signals would be detectable almost independently of the data quality.

- The false positive rate is higher for CRTS and almost minimal for LSST. This indicates that the high quality of LSST light curves will allow for the detection of very reliable SMBHB candidates.

- The false positive rate does not depend on the input parameter of a simulated DRW signal, i.e. all DRW light curves are equally likely to produce false detections.

- There are parts of the parameters space where there is no significant overlap between true signals and false detections. If the recovered parameters of a light curve fall in that regions (e.g., $A/\sigma > 1$ for LSST) it can significantly increase our confidence in the periodicity detection.

- If periodicity is present in a light curve, and only a DRW model is fit, the recovery of the parameters is biased.

- Future work will include more realistic LSST light curves, a wider range of binary signal models, and a physically motivated binary population.

Chapter 4

The NANOGrav 12.5-year Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries

4.1 Abstract

Pulsar timing arrays (PTAs), such as the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), are the only observatories that can detect low-frequency ($10^{-9} - 10^{-7}$ Hz) gravitational waves (GWs). We have searched for continuous waves (CWs) from individual supermassive black hole binaries (SMBHBs) using NANOGrav's recent 12.5-year data set. We created new methods to accurately model the uncertainties on pulsar distances in our analysis, and implemented new techniques to account for an emerging common red noise pro-

To be submitted to ApJ.

Contributing authors: Zaven Arzoumanian, Paul T. Baker, Laura Blecha, Harsha Blumer, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, Bence Bécsy, Maria Charisi, Shami Chatterjee, James M. Cordes, Neil J. Cornish, Fronefield Crawford, H. Thankful Cromartie, Megan E. DeCesar, Paul B. Demorest, Timothy Dolch, Brendan Drachler, Justin A. Ellis, Elizabeth C. Ferrara, William Fiore, Emmanuel Fonseca, Nathan Garver-Daniels, Peter A. Gentile, Deborah C. Good, Kayhan Gültekin, Jeffrey S. Hazboun, Ross J. Jennings, Megan L. Jones, Andrew R. Kaiser, David L. Kaplan, Luke Zoltan Kelley, Joey Shapiro Key, Nima Laal, Michael T. Lam, T. Joseph W. Lazio, Tingting Liu, Duncan R. Lorimer, Jing Luo, Ryan S. Lynch, Dustin R. Madison, Alexander McEwen, Maura A. McLaughlin, Chiara M. F. Mingarelli, Cherry Ng, David J. Nice, Stella Koch Ocker, Timothy T. Pennucci, Nihan S. Pol, Scott M. Ransom, Paul S. Ray, Joseph D. Romano, Brent J. Shapiro-Albert, Xavier Siemens, Joseph Simon, Magdalena Siwek, Renée Spiewak, Ingrid H. Stairs, Daniel R. Stinebring, Kevin Stovall, Joseph K. Swiggum, Jessica Sydnor, Stephen R. Taylor, Jacob E. Turner, Michele Vallisneri, Sarah J. Vigeland, Haley M. Wahl, Gregory Walsh

cess in PTA data sets while searching for deterministic GW signals. As we found no strong evidence for CWs in our data, we placed 95% upper limits on the strain amplitude of CWs emitted by these sources. At our most sensitive frequency of 7.65 nanohertz, we placed a sky-averaged limit of $h_0 < (6.82 \pm 0.35) \times 10^{-15}$, and $h_0 < (2.66 \pm 0.15) \times 10^{-15}$ in our most sensitive sky location. Finally, we placed a multi-messenger limit of $\mathcal{M} < (1.41 \pm 0.02) \times 10^9 M_\odot$ on the chirp mass of the SMBHB candidate 3C 66B.

4.2 Introduction

Supermassive black hole binaries (SMBHBs) are expected to form in the aftermath of galaxy mergers, when the two constituent supermassive black holes eventually become gravitationally bound (Begelman et al., 1980). Once they have reached a sufficient stage of evolution, and therefore orbital separation, these binaries are predicted to be among the brightest sources of low-frequency gravitational waves (GWs) in the universe, emitting at frequencies of $\sim 1 - 100$ nHz. The GWs emitted by discrete SMBHBs are known as continuous waves (CWs), while the dominant source of nanohertz GWs is expected to be the stochastic background of GWs (GWB) that has contributions from the entire cosmic population of SMBHBs and potentially other sources (Sesana et al., 2004; Burke-Spolaor et al., 2019).

By carefully monitoring the radio pulses from stable millisecond pulsars (MSPs) over many years, pulsar timing arrays (PTAs) seek to detect correlated fluctuations in the pulse times of arrival due to the influence of low-frequency GWs (Detweiler,

1979; Foster & Backer, 1990). There are multiple PTA collaborations currently operating; among them, the North American Nanohertz Observatory for Gravitational Waves (NANOGrav; McLaughlin, 2013), the Parkes Pulsar Timing Array (PPTA; Manchester et al., 2013; Hobbs, 2013), and the European Pulsar Timing Array (EPTA; Desvignes et al., 2016), have each produced multiple pulsar timing data sets with which to search for GWs. These three groups, along with the recently formed Indian Pulsar Timing Array (InPTA), also combine efforts as a consortium known as the International Pulsar Timing Array (IPTA; Verbiest et al., 2016a).

These PTA data sets have enabled numerous searches for GWs from SMBHBs, as well as primordial GWs (e.g. Benetti et al., 2022), cosmic strings (e.g. Arzoumanian et al., 2018a), and cosmological phase transitions (Arzoumanian et al., 2021; Xue et al., 2021). Modeling has suggested that the GWB signal will be detected first (Rosado et al., 2015). While PTAs have not yet detected a GWB, they have continually placed steadily improving limits on such a signal (van Haasteren et al., 2011; Demorest et al., 2013; Shannon et al., 2013; Lentati et al., 2015; Shannon et al., 2015; Verbiest et al., 2016b; Arzoumanian et al., 2016, 2018a) until around 2015, when published limits began to stabilize at a characteristic strain value of a few times 10^{15}. In the NANOGrav 12.5-year data set (Alam et al., 2021a), PPTA second data release (Kerr et al., 2020), and IPTA data release 2 (Perera et al., 2019), not only does the upper limit fail to decrease, but a common red noise (CRN) process with some characteristics similar to those predicted for a SMBHB-origin GWB was detected to high significance (Arzoumanian et al., 2020a; Goncharov et al., 2021; Antoniadis et al., 2022).

While this common red-noise process is heartening for future GWB searches, it has sparked new challenges for CW searches, as the background takes the form of a noise process, which (like any noise process underlying a signal) may work to disrupt the sensitivity of CW searches. Over the past decades, all-sky and all-frequency CW searches have improved their sensitivity by several orders of magnitude in GW strain (e.g. Yardley et al., 2010; Arzoumanian et al., 2014; Zhu et al., 2014; Babak et al., 2016; Aggarwal et al., 2019), allowing the sensitivity horizon of PTAs to expand by several orders of magnitude. This has allowed the PTA horizon to include increasing numbers of specific systems of interest (e.g. Lommen & Backer, 2001; Jenet et al., 2004; Aggarwal et al., 2019; Charisi et al., 2022). PTAs are likely to reach the sensitivities required to detect a CW soon after the GWB is detected (Mingarelli et al., 2017; Kelley et al., 2018), and we are working to revise and improve CW search methodologies as CW upper limits decrease.

In this paper, we present the results of an all-sky search for CWs from individual SMBHBs in the NANOGrav 12.5-year data set. This work is an extension of the searches performed in previous NANOGrav datasets (presented in Arzoumanian et al. 2014 and Aggarwal et al. 2019 for the 5- and 11-year data sets, respectively). Our new search benefits from the use of the more sensitive 12.5-year data set. Most critically, however, in this work we must account for the existence of an emerging common noise signal in this data set, and understand the impact that this signal may have on CW sensitivity.

This paper is organized as follows. In section 4.3, we present an overview of the data used for our analysis, details of new pulsar distance modeling methods created

for CW searches, and a description of the GW signals and analysis methods used throughout this paper. In section 4.4, we present the results of our GW searches. For the busy reader, our main results can be summarized as follows:

- For accurate low-frequency CW searches, the common red noise (CRN) process that has been seen in GWB searches must be accounted for in our signal modeling; otherwise, our detection metrics may report a false positive result.
- Once this was taken into account, we found that no CWs were detected in the 12.5-year data set.
- With this knowledge, we place stringent limits on the CW amplitude as a function of GW frequency. For the most sensitive frequency of 7.65 nHz, this limit reaches strains of $(6.82 \pm 0.35) \times 10^{-15}$, and we also place limits on the CW amplitude at this frequency as a function of sky location.
- While our all-sky sensitivity has improved with each subsequent NANOGrav data set, we find herein that for a portion of the sky, the upper limit at the most sensitive frequency of 7.65 nHz is stable compared to or worse than in previous data sets. Through extensive simulations, we link this effect to the newly-detectable CRN process in the 12.5-year data set.
- We use these limits to make inferences about the local population of SMBHBs, and limit the distance to an SMBHB emitting at 7.65 nHz to be greater than 86.65 Mpc for a $10^9 M_\odot$ binary in the most sensitive sky location.
- We use multi-messenger techniques to update limitations on the chirp mass of

the SMBHB candidate 3C 66B to be less than $(1.41 \pm 0.02) \times 10^9 M_\odot$

In section 4.6, we discuss the implications of these results. In section 4.7, we summarize our conclusions.

4.3 Methods

4.3.1 The 12.5-year Data Set

We analyzed the NANOGrav 12.5-year data set, originally published as Alam et al. (2021a,b), which consists of times-of-arrival (TOAs) and timing models from 47 pulsars. Two versions of the data set were created from the original observations, taken between 2004 and 2017, using independent analyses. Here, we make use of the narrowband version of the data set (Alam et al., 2021a). This adds 2 pulsars and 1.5 years of observations over previous GW searches. For GW analyses, we require the pulsars to have a timing baseline of at least 3 years; therefore, we use only 45 of the 47 pulsars included in the full data set. However, as the 11-year data set included only 34 pulsars that could be used in GW analyses, this represents a significant addition of data, which will add to the sensitivity of the PTA. It is important to note that the 12.5-year data set is not merely an addition of TOAs to previous releases, but a full re-analysis with an updated pipeline, described in detail in (Alam et al., 2021a). Thus, our search will also benefit from improved timing precision for pulsars shared with previous data sets.

4.3.2 Signal Model

As in previous NANOGrav searches for continuous gravitational waves, we will describe the effect of an individual SMBHB on a pulsar's residuals δt. While the methods remain nearly identical to previous iterations, slight alterations have been made to improve consistency with other work in the field, as well as updates to reflect more recent data. As such, we will lay out the methods with particular focus on any instances that have changed since Aggarwal et al. (2019).

The pulsar residuals can be separated into multiple components as

$$\delta t = M\epsilon + n_{\text{white}} + n_{\text{red}} + s, \tag{4.1}$$

where M is the design matrix, and describes the linearized timing model, and ϵ is a vector of the timing model parameter offsets. This term allows the timing model parameters of each pulsar to be adjusted in accordance with the presence of any additional signals. The variables n_{white} and n_{red} refer to vectors describing the pulsar white and red noise, respectively, and s is a vector of GW-induced signal present in the residuals.

4.3.2.1 CW Signal

For a GW source located at right ascension α and declination δ, it is preferable to rewrite the coordinates in terms of the polar angle $\theta = \pi/2 - \delta$ and azimuthal angle $\phi = \alpha$. The strain of GWs emitted from such a source can be written in terms of two polarizations as

$$h_{ab}(t,\hat{\Omega}) = e^{+}_{ab}(\hat{\Omega})h_{+}(t,\hat{\Omega}) + e^{\times}_{ab}(\hat{\Omega})h_{\times}(t,\hat{\Omega}), \tag{4.2}$$

where $\hat{\Omega}$ is a unit vector pointing from the the GW source to the earth (along the direction of propagation), $h_{+,\times}$ are the polarization amplitudes, and $e^{+,\times}_{ab}$ are the polarization tensors. These can be written in the solar system barycenter frame as

$$\begin{aligned} e^{+}_{ab} &= \hat{p}_a\hat{p}_b - \hat{q}_a\hat{q}_b \\ e^{\times}_{ab} &= \hat{p}_a\hat{q}_b + \hat{q}_a\hat{p}_b, \end{aligned} \tag{4.3}$$

and are constructed from basis vectors[1]

$$\begin{aligned} \hat{n} =&(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) = -\hat{\Omega}, \\ \hat{p} =&(\cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi, \\ &\cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi, -\cos\psi\sin\theta), \\ \hat{q} =&(\sin\psi\cos\theta\cos\phi + \cos\psi\sin\phi, \\ &\sin\psi\cos\theta\sin\phi - \cos\psi\cos\phi, -\sin\psi\sin\theta). \end{aligned} \tag{4.4}$$

These polarization tensors are used to construct the antenna pattern function $F^{+,\times}(\hat{\Omega})$, which describes the pulsar's response to the GW source, as

$$F^{A}(\hat{\Omega}) \equiv \frac{1}{2}\frac{\hat{u}^a\hat{u}^b}{1+\hat{\Omega}\cdot\hat{u}}e^{A}_{ab}(\hat{\Omega}) \tag{4.5}$$

[1]Note: This basis is different than that used in Aggarwal et al. (2019) to maintain better consistency with previous references and the standards used by other GW detectors. Differences can be reduced to a rotation of the frame by an angle equivalent to the GW polarization angle ψ.

(Taylor et al., 2016).

Now, we can write the signal s induced by the GW as seen in the pulsar's residuals as

$$s(t, \hat{\Omega}) = F^{+}(\hat{\Omega})\Delta s_{+}(t) + F^{\times}(\hat{\Omega})\Delta s_{x}(t), \tag{4.6}$$

where $\Delta s_{+,x}$ is the difference between the signal induced at the Earth (the "Earth Term") and at the pulsar (the "pulsar term"). This can be written as

$$\Delta s_{+,x}(t) = s_{+,x}\left(t_p\right) - s_{+,\times}(t), \tag{4.7}$$

where t and t_p represent the time when the GW passes the Earth and the pulsar, respectively. These times can be related geometrically by

$$t_p = t - L(1 + \hat{\Omega} \cdot \hat{d}), \tag{4.8}$$

where $\hat{d}$ is the line of sight vector to the pulsar and L is the distance to the pulsar (see subsubsection 4.3.6.1 for further discussion of this value).

For a circular binary at zeroth post-Newtonian (0-PN) order, $s_{+,\times}$ can be written as

$$\begin{aligned} s_{+}(t) =& \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \left[-\sin 2\Phi(t) \left(1 + \cos^2 i\right)\right] \\ s_{\times}(t) =& \frac{\mathcal{M}^{5/3}}{d_L \omega(t)^{1/3}} \left[2 \cos 2\Phi(t) \cos i\right] \end{aligned} \tag{4.9}$$

where i is the inclination angle of the SMBHB, d_L is the luminosity distance to the source, and $\mathcal{M} \equiv \left(m_1 m_2\right)^{3/5} / \left(m_1 + m_2\right)^{1/5}$ is a combination of the two black hole masses known as the chirp mass. Again, note that the forms of these signals have

changed slightly from those used in Aggarwal et al. (2019); due to the rotated frame of the antenna pattern functions now in use, they are equivalent. The variables $\mathcal{M}$ and ω refer to the redshifted values of these quantities, which relate to the rest-frame versions $\mathcal{M}_r$ and ω_r as

$$\begin{aligned} \mathcal{M}_r &= \frac{\mathcal{M}}{1+z} \\ \omega_r &= \omega(1+z). \end{aligned} \tag{4.10}$$

However, PTAs are currently only sensitive to the local universe, where $(1+z) \approx 1$.

For a circular SMBHB, the initial orbital angular frequency is related to the GW frequency by $\omega_0 = \pi f_{\rm GW}$, where $\omega_0 = \omega(t_0)$. For this search, we define the reference time t_0 as MJD 57933 (June 29, 2017), the last observation date for the 12.5-year data set. The time-dependent orbital phase and frequency of the binary are given by

$$\begin{aligned} \Phi(t) &= \Phi_0 + \tfrac{1}{32}\mathcal{M}^{-5/3}\left[\omega_0^{-5/3} - \omega(t)^{-5/3}\right] \\ \omega(t) &= \omega_0\left(1 - \tfrac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}t\right)^{-3/8} \end{aligned} \tag{4.11}$$

where Φ_0 refers to the initial orbital phase (Arzoumanian et al., 2014). To account for the evolution of high chirp mass binaries over our observations, we use the full expression for $\omega(t)$ as in Aggarwal et al. (2019).

4.3.2.2 Noise Model

For individual each individual pulsar, we model both white and red noise. We use an identical white noise model as previous NANOGrav analyses, using three parameters: EFAC, EQUAD, and ECORR. EFAC scales the TOA uncertainties

by a factor, EQUAD adds white noise in quadrature, and ECORR describes white noise that is correlated only within the same observing epoch.

To model individual pulsar red noise, the noise spectrum is divided into 30 linearly spaced bins, ranging from $1/T_{obs}$ to $30/T_{obs}$, where T_{obs} is the total observation baseline for each pulsar. Then, the power spectral density of the red noise is fit to a power-law model

$$P(f) = \frac{A_{\rm red}^2}{12\pi^2} \left(\frac{f}{f_{\rm yr}}\right)^{-\gamma}. \tag{4.12}$$

Here, $f_{\rm yr} \equiv 1/(1{\rm year})$, $A_{\rm red}$ is the red noise amplitude, and γ is the power law spectral index.

As mentioned above, for the first time, a CRN signal is now detectable in the 12.5-year data set (Arzoumanian et al., 2020a). Because of this, we included a CRN term in our signal model for a portion of our analyses. The results of searches that only model a CW necessitated this addition, and are described in detail in section 4.4. The power spectral density of the CRN

$$P(f) = \frac{A_{\rm CRN}^2}{12\pi^2} \left(\frac{f}{f_{\rm yr}}\right)^{-\gamma}. \tag{4.13}$$

takes the same form as that of the pulsar red noise in Equation 4.12, but with an amplitude $A_{\rm CRN}$ and spectral index γ that are common to all of the pulsars in the array.

4.3.3 Bayesian Methods

We utilized Bayesian inference techniques to determine the posterior distributions of GW parameters. In each analysis, we include the BayesEphem model (Vallisneri et al., 2020) to account for the uncertainties in the Solar System ephemeris, which, as first described in Arzoumanian et al. (2018a), can have large impacts on the computation of GW upper limits with PTAs. We used DE438 (Folkner & Park, 2018) plus BayesEphem to transform from individual observatory reference frames to an inertial frame centered at the Solar System Barycenter.

As in previous NANOGrav CW searches, we use the `enterprise` (Ellis et al., 2019) package to construct the priors and evaluate the likelihood, which takes the same form as in Aggarwal et al. (2019) and Arzoumanian et al. (2014). The Markov Chain Monte Carlo (MCMC) sampler package `PTMCMCSampler` (Ellis & van Haasteren, 2017) was used to explore the parameter space.

The CW signal model can be described by nine global parameters:

$$\{\theta, \phi, f_{\rm GW}, \Phi_0, \psi, i, \mathcal{M}, d_{\rm L}, h_0\}, \tag{4.14}$$

which describe the circular SMBHB's:

- position on the sky (θ, ϕ);
- GW frequency, related to the orbital frequency at some reference time $(f_{\rm GW})$;
- orbital phase at some reference time (Φ_0);
- GW polarization angle (ψ);
- orbital inclination (i);

- chirp mass ($\mathcal{M}$);
- luminosity distance ($d_{\rm L}$);
- strain amplitude (h_0), which is related to the chirp mass, GW frequency, and luminosity distance.

However, due to a degeneracy between h_0, $\mathcal{M}$, $f_{\rm GW}$, and D_L, since h_0 can be defined as

$$h_0 = \frac{2\mathcal{M}^{5/3}(\pi f_{GW})^{2/3}}{d_L}, \tag{4.15}$$

only eight of these parameters are required to fully describe the global CW signal. The following types of searches use a variety of prior setups to sample the necessary eight global parameters, and are described below and summarized in Table 4.1.

4.3.4 All-Sky Searches

To search for GWs from SMBHBs located in any direction, we use uniform priors on the source sky position $(cos(\theta), \phi)$, as well as the cosine of the source inclination $cos(i)$, polarization angle ψ, and GW phase Φ_0. We used log-uniform priors on h_0 for detection analyses, and uniform priors on h_0 for upper limit analyses, so as to set the most conservative upper limit. For both analysis types, priors on $\log_{10}(h_0)$ span the range $[-18, -11]$, which accounts for an over-conservative range around the sensitivity of the most recent data sets (order -15).

We performed many searches at fixed values of $f_{\rm GW}$, to evaluate the detection statistics and sensitivity of the PTA across the entire nanohertz GW band.

The lowest frequency value was set by the total observation time of the PTA, $f_{\rm GW} = 1/(12.9 \text{ years}) = 2.45$ nHz. The highest frequency value is limited by the observation cadence of the PTA (approximately one observation per 2-4 weeks). However SMBHBs at that frequency, at the mass range where their strains would be large enough to be detectable by PTAs, have exceedingly short inspiral timescales (a few weeks up to $\sim$ 3 months). Thus, they are unlikely to be detectable in our data set. Therefore, we set our maximum frequency to 317.8 nHz (equivalent to around one GW cycle every 36 days). This is the same high-frequency cutoff value used in Arzoumanian et al. (2014); Aggarwal et al. (2019).

For most of the frequency band, we searched over $\log_{10}\frac{\mathcal{M}}{M_\odot}$ with a log-uniform prior with a range of $[7, 10]$. However, for very high-frequency sources, we limit the maximum value of the prior to account for high-chirp-mass binaries never emitting GWs at the highest frequencies in our band, as they will have merged prior to emitting GWs at $f_{\rm GW} \geq 191.3$ nHz. Assuming binaries merge when the orbital frequency is equal to the innermost stable circular orbit (ISCO) frequency, $\mathcal{M}$ must satisfy

$$\mathcal{M} \leq \frac{1}{6^{3/2}\pi f_{\rm gw}} \left[\frac{q}{(1+q)^2}\right]^{3/5}, \tag{4.16}$$

where q is the SMBHB mass ratio. Here, we calculated the chirp mass cutoff for $q = 1$.

4.3.5 Sky Map

Due to the non-uniform distribution of pulsars on the sky, PTAs are typically not equally sensitive in all directions. To analyze the differences in sensitivity, once detection analyses have been completed, we also place upper limits on 768 pixels distributed isotropically across the sky. The sampler is able to search a uniform prior across the small pixel, so as to still sample the entire sky across the entire analysis.

Due to the large computational cost required to conduct 768 independent runs, the sky map is created at only a single frequency. We selected 7.65 nHz, as it was the most sensitive in the sky-averaged analysis. As this is in the low-frequency regime where we expect the inclusion of the CRN to be significant, it is included in our signal model. All other modeling is done identically to subsection 4.3.4, and is summarized in Table 4.1.

4.3.6 Targeted Search

In addition to the two variations searches described above, we also perform a targeted search for a known SMBHB candidate, 3C 66B. This source was the subject of study by Arzoumanian et al. (2020b), and here, we are able to provide an updated analysis with the addition of the new data included in the 12.5-year data set.

For the targeted search, we perform detection and upper limit analyses in the same way as in subsection 4.3.4, with a few differences in the model priors. Because

we know the sky location and luminosity distance to 3C 66B, as well as a frequency estimate, these parameters are set to constants in this search. This allows us to place constraints directly on the chirp mass of the binary, rather than its GW strain amplitude. For a detection analysis, the prior on $\log_{10}\left(\frac{\mathcal{M}}{M_\odot}\right)$ is log-uniform in the range $[7, 10]$, while for upper limit analyses, the prior is uniform over this range. The remaining priors are identical to the above analyses, and are summarized in Table 4.1.

4.3.6.1 Pulsar Distance Priors

In this work, we adopt a data-driven approach to handle the large uncertainties on pulsar distance measurements. As in previous searches, the pulsar distance is used as a free parameter in the search. This allows us to marginalize over the pulsar distance, and avoid incorrect modeling of the signal at the the location of the pulsar.

In previous versions of this search (e.g. Aggarwal et al., 2019; Arzoumanian et al., 2020b), the pulsar distance prior was constructed from a Gaussian scaled to the parallax distance and associated uncertainty listed in Verbiest et al. (2012). If no distance was listed, a value of 1.0 ± 0.2 kpc was assumed. While this assumption is reasonable while placing upper limits (see discussion within Arzoumanian et al. 2020b), as the PTA reaches sensitivities where a detection is nearly possible, an improvement is needed.

In this work, every pulsar distance prior is constructed from a measurement or estimate. If a pulsar has a significant independent parallax measurement[2], such

[2] http://hosting.astro.cornell.edu/research/parallax/, with values compiled from

	All-Sky		Sky Map	Targeted	
Analysis Type	Detection	Upper Limit	Upper Limit	Detection	Upper Limit
CRN	Y/N	Y/N	Y	Y/N	Y/N
$\log_{10}h$	Uniform(–18,–11)	LinExp(–18,–11)	LinExp(–18,–11)	–	–
$\log_{10}\mathcal{M}$	Uniform(7,$\mathcal{M}_{max}$)	Uniform(7,$\mathcal{M}_{max}$)	Uniform(7,$\mathcal{M}_{max}$)	Uniform(7,$\mathcal{M}_{max}$)	LinExp(7,$\mathcal{M}_{max}$)
$\log_{10}d_L$	–	–	–	Constant	Constant
$\log_{10}f_{\rm GW}$	Constant (many)	Constant (many)	Constant (single)	Constant	Constant
ϕ	Uniform(0,2π)	Uniform(0,2π)	Uniform(pixel)	Constant	Constant
$\cos\theta$	Uniform(–1,1)	Uniform(–1,1)	Uniform(pixel)	Constant	Constant
ψ	Uniform(0,π)	Uniform(0,π)	Uniform(0,π)	Uniform(0,π)	Uniform(0,π)
Φ_0	Uniform(0,2π)	Uniform(0,2π)	Uniform(0,2π)	Uniform(0,2π)	Uniform(0,2π)
$\cos\iota$	Uniform(–1,1)	Uniform(–1,1)	Uniform(–1,1)	Uniform(–1,1)	Uniform(–1,1)

Table 4.1: CW parameter priors for each analysis.

as from Very Long Baseline Interferometry (VLBI), or timing parallax measured in the 12.5-year data set, this value is used to construct a prior on pulsar distance (d_p)

$$p(d_p) = \frac{1}{\sqrt{2\pi}\sigma_{PX} d_p^2} \exp \frac{-(PX - d_p^{-1})^2}{2\sigma_{PX}^2}, \tag{4.17}$$

which inverts the approximately Gaussian shape of a parallax prior to describe the prior for distance (Vigeland & Vallisneri, 2014). Here, significance is defined by the parallax measurement (PX) having an associated uncertainty (σ_{PX}) of less than 30%, so as to avoid the introduction of any errors due to the Lutz-Kelker bias (Lutz & Kelker, 1973). If multiple measurements of sufficient quality exist, these values and uncertainties are combined with a weighted average before being used to construct the prior.

If there are no parallax measurements that can be used, the pulsar's dispersion measure (DM) is used to construct a distance estimate using NE2001 (Cordes & Lazio, 2002) and subsequently, the distance prior. Since these values are only an estimate, we construct a broad, nearly uniform prior for the DM-distance value and a 20% uncertainty (Cordes & Lazio, 2002; Jones et al., 2017; Lam et al., 2016), with the shape

$$p(d_p) = \begin{cases} \text{Gaussian} & \text{if } d_p < 0.8\ d_{DM} \\ \text{Uniform} & \text{if } 0.8\ d_{DM} \leq d_p \leq 1.2\ d_{DM} \\ \text{Gaussian} & \text{if } d_p > 1.2\ d_{DM} \end{cases} \tag{4.18}$$

(Ding et al., 2020; Jennings et al., 2018; Deller et al., 2019; Guillemot et al., 2016; Stovall et al., 2014; Abdo et al., 2013; Freire et al., 2012; Verbiest et al., 2009; Lazaridis et al., 2009; Chatterjee et al., 2009; Hotan et al., 2006; Lommen et al., 2006; Jacoby et al., 2005; Splaver et al., 2005; Löhmer et al., 2004; Toscano et al., 1999; Camilo et al., 1994)

Here, the Gaussian additions have standard deviations of 0.25 times the DM-distance uncertainty, to allow the sampler to move into the edges of this prior range, to account for any differences in distance estimates by alternative electron density models, such as Yao et al. (2017). While pulsar distance priors will still only induce minor influences on the results of an upper limit analysis (Arzoumanian et al., 2020b), by constructing new priors to accurately handle pulsar distance measurements and estimates, we have prepared our methods for the eventual detection of a CW, which will be more reliant on the pulsar term of the signal than upper limit evaluations.

4.4 Results

4.4.1 All-Sky Searches

For each GW frequency in our search, we performed a detection analysis on the 12.5-year data which marginalized over the source sky location. Figure 4.1 shows the Bayes factor for a CW at each searched GW frequency in purple. However, it is important to note the large (near-infinite) Bayes factor for $f_{\rm GW} = 2.45$ nHz, with a steady decrease in the following few frequency bins. Ordinarily, this would be a first indication for the detection of a CW. However, given the strong evidence for the existence of a CRN process in the 12.5-year data set (Arzoumanian et al., 2020a), it is clear that this signal appears to be of similar form; that is, what we have detected is bright at low frequencies and declines toward higher frequency. Once a common red-noise process is added to the model, with the A_{CRN} and γ parameters fixed to the maximum likelihood values found by the search of Arzoumanian et al. (2020a),

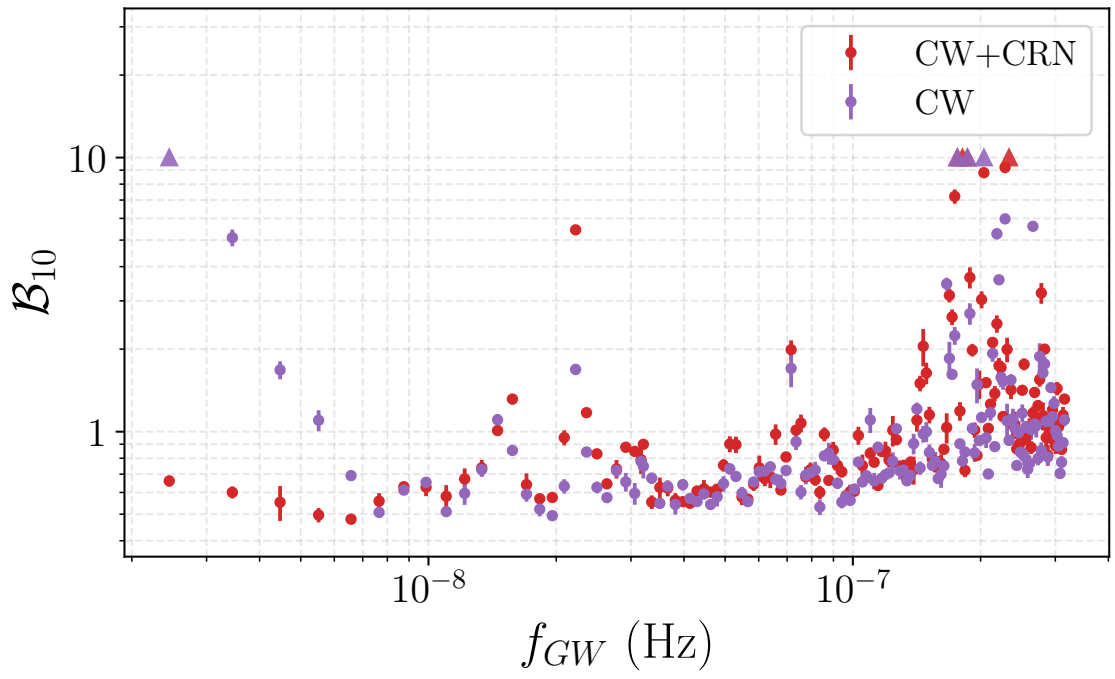

Figure 4.1: Savage-Dickey Bayes factors for a CW at each GW frequency. At low frequencies, inclusion of a CRN in the model (red) is necessary to avoid a false CW detection (purple). Triangle markers indicate an infinite Savage-Dickey Bayes Factor, which could be an effect of inadequate sampling at high frequencies.

the Bayes factors for a CW at low $f_{\rm GW}$ return to < 1 (red points in the figure). Therefore, throughout this paper, we will present the results of many analyses with a fixed CRN included in our model.

We note that a few frequencies above $f_{GW} = 100$ nHz have $\mathcal{B}_{10}$ values that are returned as very large. However, upon inspection, this is due to inadequate sampling in a few frequency bins, rather than a detection of a CW. With further sampling, it is expected that these values will settle near $\mathcal{B}_{10} \sim 10$, and therefore, these results should be treated as preliminary for frequencies greater than 1×10^{-7}nHz.

As we found no strong evidence for a GW from an individual SMBHB in the 12.5-year data set, we proceeded to place all-sky upper limits on the GW strain,

with results shown in Figure 4.2. We again conduct this analysis using two different models, one which includes only a CW (purple) and one which includes both a CW and a CRN process (red). While in both cases, the most sensitive frequency (that with the lowest strain upper limit) is 7.65 nHz, the strain upper limits are lower when the CRN is included in the model. In this case, we can limit the strain to $h_0 < (6.82 \pm 0.35) \times 10^{-15}$, while when the CRN is neglected, the best limit we can place on CW strain is $h_0 < (9.11 \pm 0.10) \times 10^{-15}$. This trend of the CW+CRN model resulting in lower upper limits than a CW-only model continues until frequencies of approximately 10 nHz, above which the upper limit values are nearly equal. Therefore, throughout the remainder of this work, we will opt to include the CRN in computationally expensive analyses which can not be completed with both models, such as the sky map analyses those described in subsection 4.4.2.

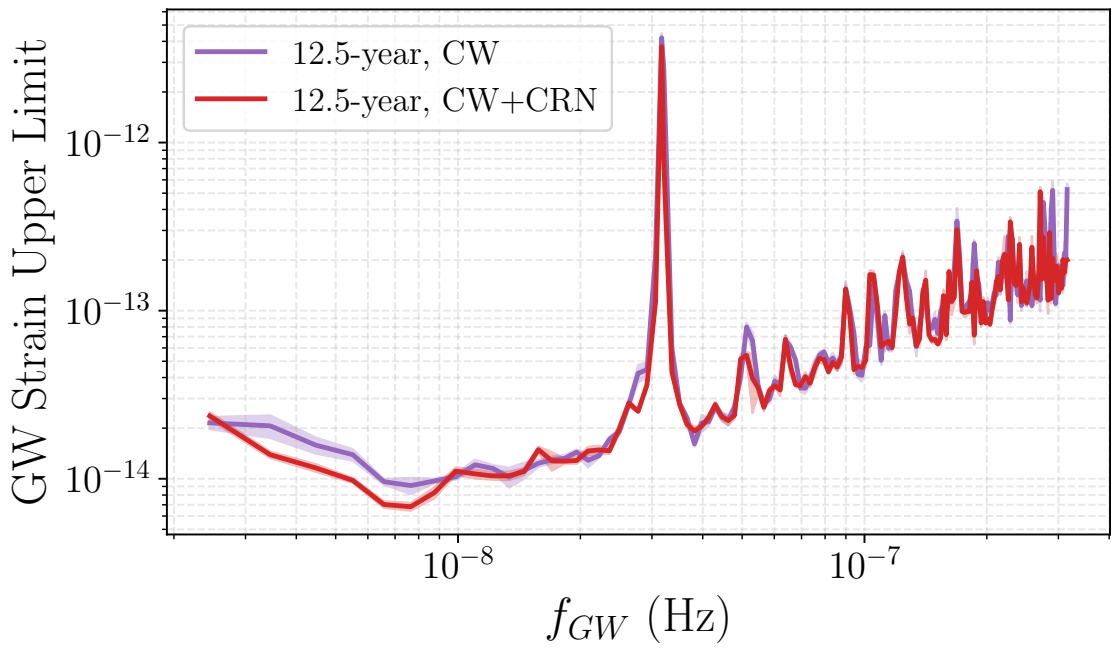

Figure 4.2: CW strain 95% upper limits and associated error regions, with (red) and without (purple) a CRN included in the model.

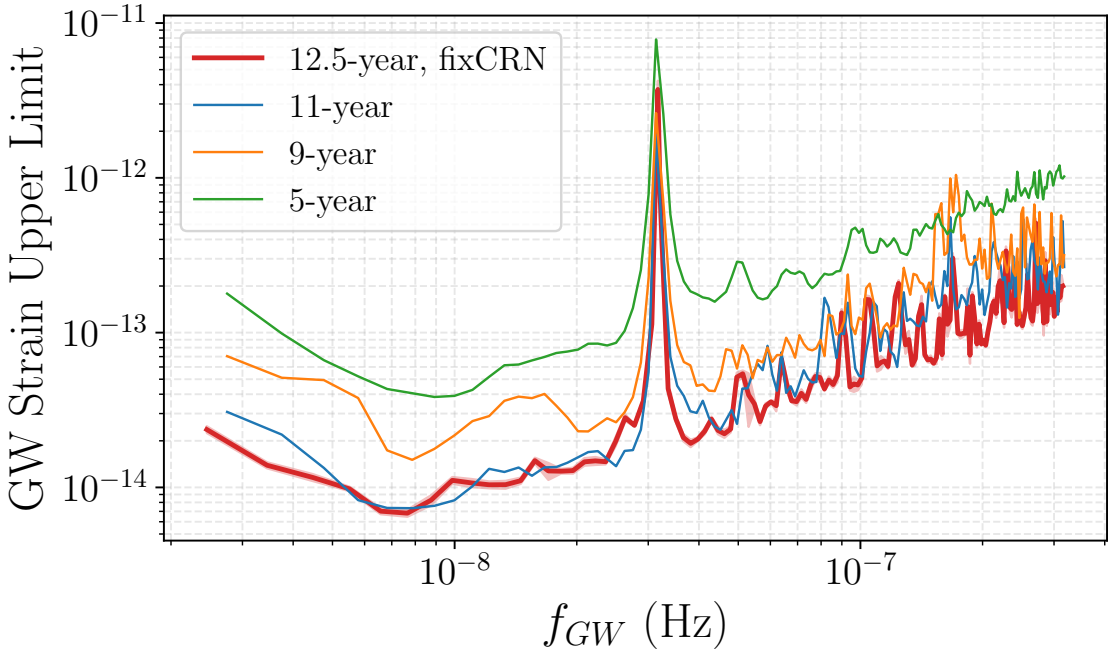

Figure 4.3: The upper limits on CW strain are continuing to decrease. The 12.5-year data set (red curve and error region is more sensitive than the 11-year, 9-year, and 5-year (blue, orange, and blue curves, respectively).

In Figure 4.3, we compare this result to those of previous NANOGrav searches for CWs (Aggarwal et al., 2019). While analyses have shown a factor of ~ 2 improvement between the previous three data sets, we see only a modest sensitivity improvement between the 11-year and 12.5-year data, with only a factor of 1.07 between the two lowest strain limits. In addition to the smaller fractional increase in observing baseline between the 11- and 12.5-year data sets as compared to previous data sets, this is likely due to the presence of the CRN, which, while it is no longer causing a false positive in the CW search if included in the model, does represent a significant noise process that will limit our sensitivity to low-frequency CWs over the years to come (Hazboun et al., 2019b).

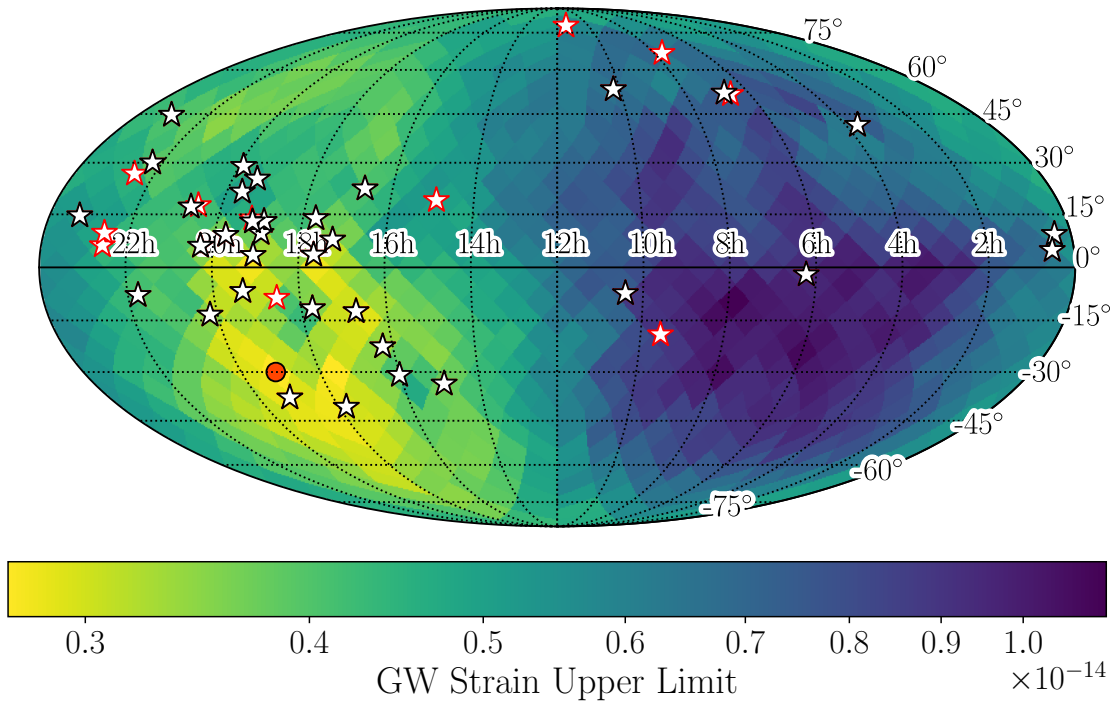

Figure 4.4: Map of CW strain 95% upper limits at $f_{\rm GW} = 7.65$ nHz, the most sensitive frequency searched, for the 12.5-year data set. Pulsar locations are shown as white stars, with new pulsars added in the 12.5-year data set outlined in red. The most sensitive pixel is marked with a red dot.

4.4.2 Sky Map

The GW strain upper limits, for a model including a CRN, at the most sensitive CW frequency $f_{\rm GW}$ =7.65 nHz as a function of sky location are shown as a map in Figure 4.4. As expected, the portion of the sky that is the least sensitive to CWs is that which contains the fewest pulsars in the PTA. At the most sensitive pixel, the strain upper limit is $h_0 < (2.66 \pm 0.15) \times 10^{-15}$, while at the least sensitive pixel, $h_0 < (1.12 \pm 0.05) \times 10^{-14}$, a range of sensitivities that varies by a factor of ~ 4.

In Figure 4.5, we compare the 12.5-year CW strain map to that constructed in Aggarwal et al. (2019) for the 11-year data set by plotting $\Delta h_{95} = h_{95,12.5} - h_{95,11}$. While a portion of the sky shows a significant reduction in strain upper limits, many

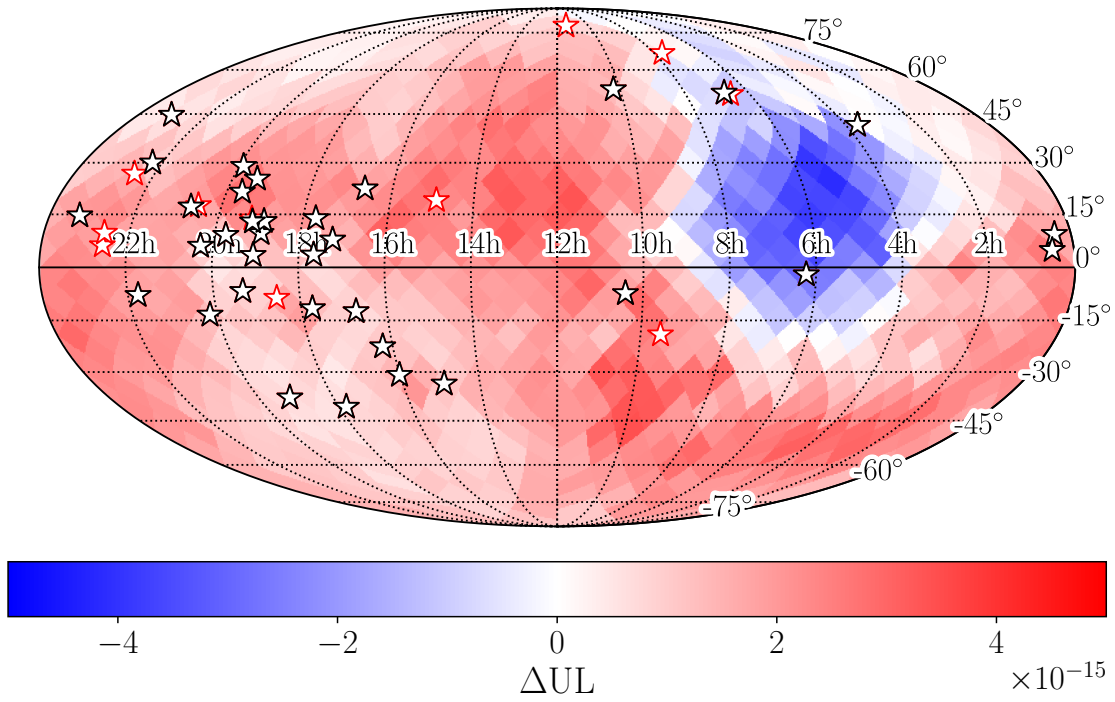

Figure 4.5: Difference in strain 95% upper limits for the 12.5-year data set versus the 11-year data set. Blue pixels indicate a decrease in upper limit, while red pixels indicate an increase.

pixels show an increase in strain upper limit, indicating a loss of sensitivity in the newest data set for much of the sky at the PTA's most sensitive frequency, including in the most sensitive area of the sky.

To investigate the cause of this apparent sensitivity loss, we conducted an analysis of the simulated data utilized in Pol et al. (2021). We selected portions of the data set with included pulsars and observation baselines corresponding to the 11- and 12.5-year data sets that also included a CRN corresponding to that found in Arzoumanian et al. (2018a). Then, we conducted identical upper limit analyses for an equatorial slice of sky pixels (i.e. for the pixels with $\theta_{GW} \sim \pi/2$). When plotted against ϕ_{GW} in Figure 4.6, the patterns in Δh_{95} in the real data are well within the range represented by the same analysis in the 10 simulated data sets, each

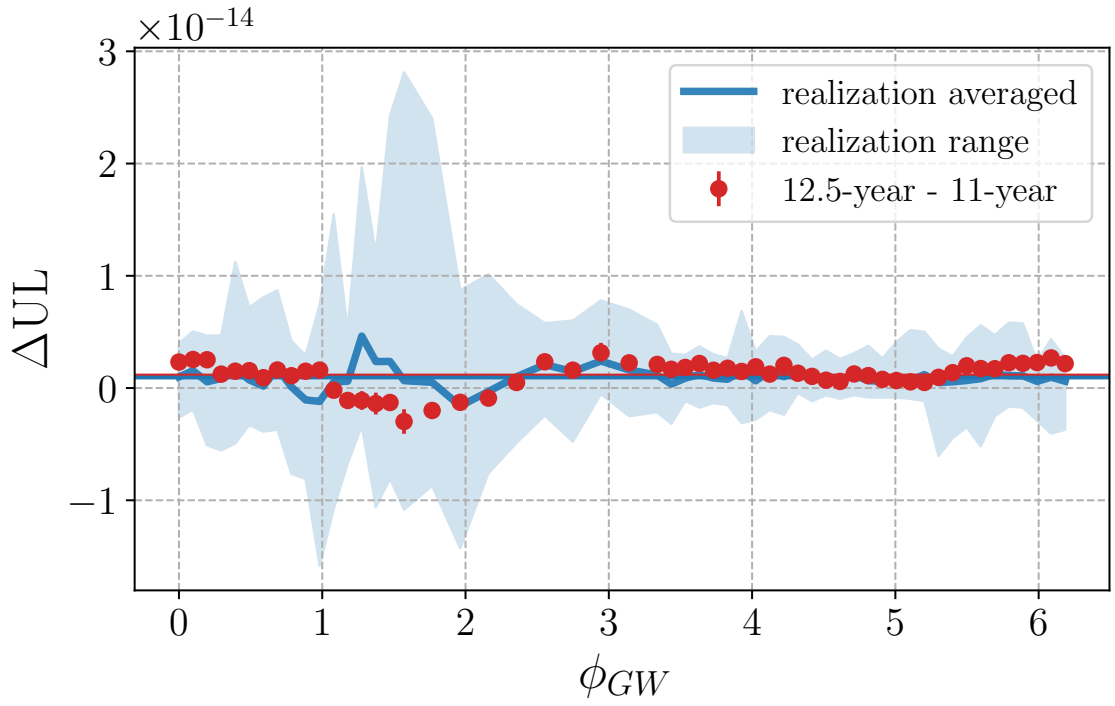

Figure 4.6: The difference in strain upper limits for an equatorial slice of the sky map shown in Figure 4.5 plotted against ϕ_{GW}. The results for the real data (red points) are well within the range of values encompassed by the ten realizations simulated (blue), with near-identical mean values of Δh_{95} (horizontal red and blue lines).

containing a different realization of the CRN. The mean value of Δh_{95} across each included pixel is nearly identical for the real data and the simulations. Together, this allows us to confidently state that this apparent pattern in our evolving sensitivity across the sky is due to the emerging CRN.

4.5 Astrophysical Limitations of Nearby SMBHBs

In recent years, numerous studies have modeled the SMBHB population in the nearby universe (Mingarelli et al., 2017) and multiple SMBHB candidates have been discovered with electromagnetic techniques (Sudou et al., 2003; Graham et al., 2015b; Hu et al., 2020; Lehto & Valtonen, 1996). Even without a CW detection,

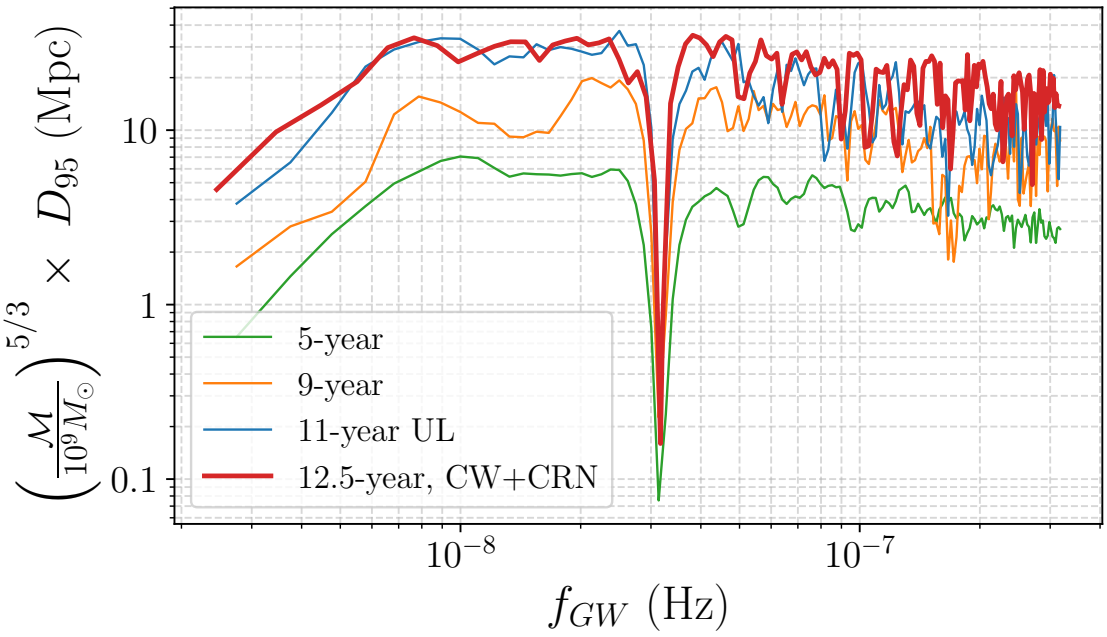

Figure 4.7: The 95% lower limits on the luminosity distance to an individual SMBHB. While we can limit SMBHBs emitting GWs at the most sensitive value of $f_{\rm GW} = 7.65$ nHz to $d_L > 33.85$ Mpc, at $f_{\rm GW} = 38.17$ nHz, they can be limited to farther away at $d_L > 34.99$ Mpc.

our limits can add crucial insights to these works, including limiting the distance to nearby SMBHBs and placing multi-messenger mass constraints on SMBHB candidates.

4.5.1 Distance Limitations

Our limits on CW strain can be transformed using Equation 4.15 to calculate the 95% lower limit on the luminosity distance to a source of a given chirp mass. The distance limits for an SMBHB with $\mathcal{M} = 10^9 M_\odot$ are shown in Figure 4.7. For the most sensitive frequency of $f_{\rm GW} = 7.65$ nHz, we can limit the distance to an SMBHB with $\mathcal{M} = 10^9 M_\odot$ to $d_L > 33.85$ Mpc. These limits may be scaled to larger

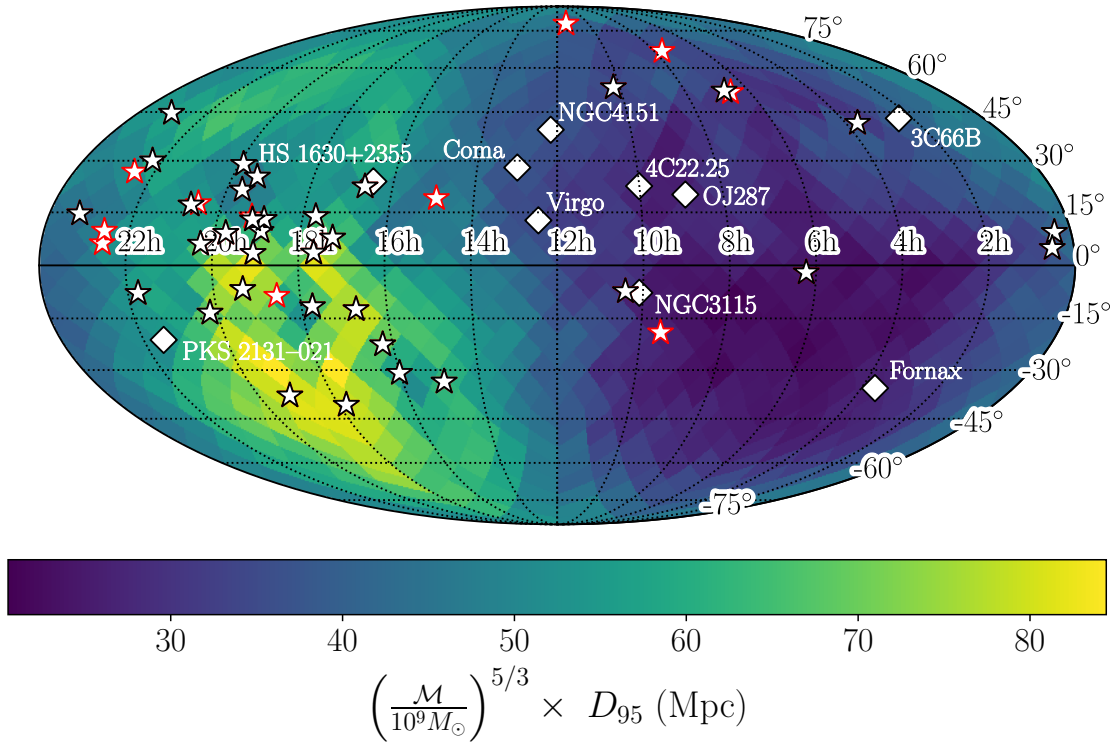

Figure 4.8: Map of the 95% lower limit on the distance to individucal SMBHBs with $\mathcal{M} = 10^9 M_\odot$ and $f_{\rm GW} = 7.5$ nHz. White diamonds indicate the positions of known SMBHB candidates and large galaxy clusters that could contain an SMBHB.

or smaller SMBHBs directly using Equation 4.15, which results in a factor of

$$D_{95,\mathcal{M}} = D_{95,10^9 M_\odot} \times \left(\frac{\mathcal{M}}{10^9 M_\odot}\right)^{5/3}. \tag{4.19}$$

However, it is important to note that while this frequency is produces the lowest strain upper limit, it does not produce the farthest luminosity distance lower limit. This value is $d_L > 34.99$ Mpc at $f_{\rm GW} = 38.17$ nHz.

This technique can be applied to the strain upper limit sky map as well, to calculate the 95% luminosity distance lower limit for an SMBHB emitting CWs at $f_{\rm GW}$ =7.65 nHz as a function of sky location. The results of this transformation

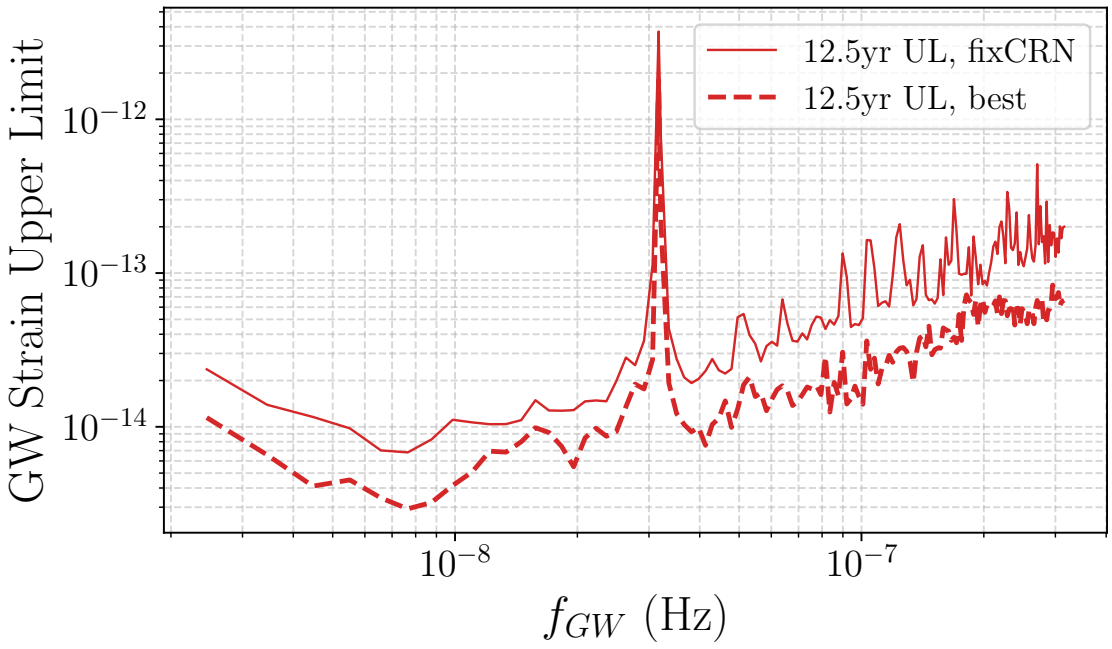

Figure 4.9: The 95% strain upper limit curve for the all-sky (solid red) CW search compared with the 95% strain upper limit curve in the most sensitive sky location (red dashed).

are shown in Figure 4.8. At the most sensitive sky location, we can limit the minimum distance to an $\mathcal{M} = 10^9 M_\odot$ SMBHB to be $d_L > 86.65$ Mpc, and that to an $\mathcal{M} = 10^{10} M_\odot$ SMBHB to $d_L > 4.02$ Gpc. In the least sensitive sky location, we can limit the minimum distance to an $\mathcal{M} = 10^9 M_\odot$ SMBHB to be $d_L > 20.50$ Mpc, and that to an $\mathcal{M} = 10^{10} M_\odot$ SMBHB to $d_L > 0.95$ Gpc. These values vary by over a factor of 4 between the most and least sensitive parts of the sky.

At the most sensitive sky pixel, we conducted a final upper limit analysis across the entire frequency band, with results plotted in Figure 4.9. Across the entire nanohertz frequency band, the PTA is dramatically more sensitive to CWs from sources at this sky location than across the entire sky on average.

4.5.2 Multi-Messenger Analyses

Using the methodology described in subsection 4.3.6, we conduct a multi-messenger search for GWs from the SMBHB candidate 3C 66B to provide an update to results of (Arzoumanian et al., 2020b). The detection analyses result in nearly identical Savage-Dickey Bayes factors, whether the CRN was included or not. This is to be expected, as the CRN is very weak at frequencies as high as that of 3C 66B. The Bayes factors for the CW-only analysis and the CW+CRN analysis are 0.70 ± 0.02 and 0.67 ± 0.01, respectively. Both of these values are very near 1, meaning that the data does not indicate the presence of a CW corresponding to a binary within 3C 66B.

Because no GW was detected, we constrain the chirp mass of a potential binary with an upper limit analysis, again performed with and without a CRN to confirm consistency. The posteriors from these two searches are plotted in Figure 4.10, with resulting 95% upper limits of $(1.41 \pm 0.02) \times 10^9 M_\odot$when a CRN is included, and $(1.34 \pm 0.01) \times 10^9 M_\odot$when only CWs are included in the signal. For comparison, the 95% chirp mass upper limit for 3C 66B from the 11-year data set was $1.65 \times 10^9 M_\odot$. This represents an improvement of $1 \times 10^8 M_\odot$, or a factor of 1.5 smaller; by adding pulsars, extending timing baselines, and improving timing and searching methods, the PTA's sensitivity has clearly improved. These upper limits are nearer to the value of the upper bound of the Iguchi et al. (2010) chirp mass estimate. In subsequent data sets, or by using more sophisticated analyses such as advanced noise modeling (Simon & Hazboun, in prep), this error region will be within reach.

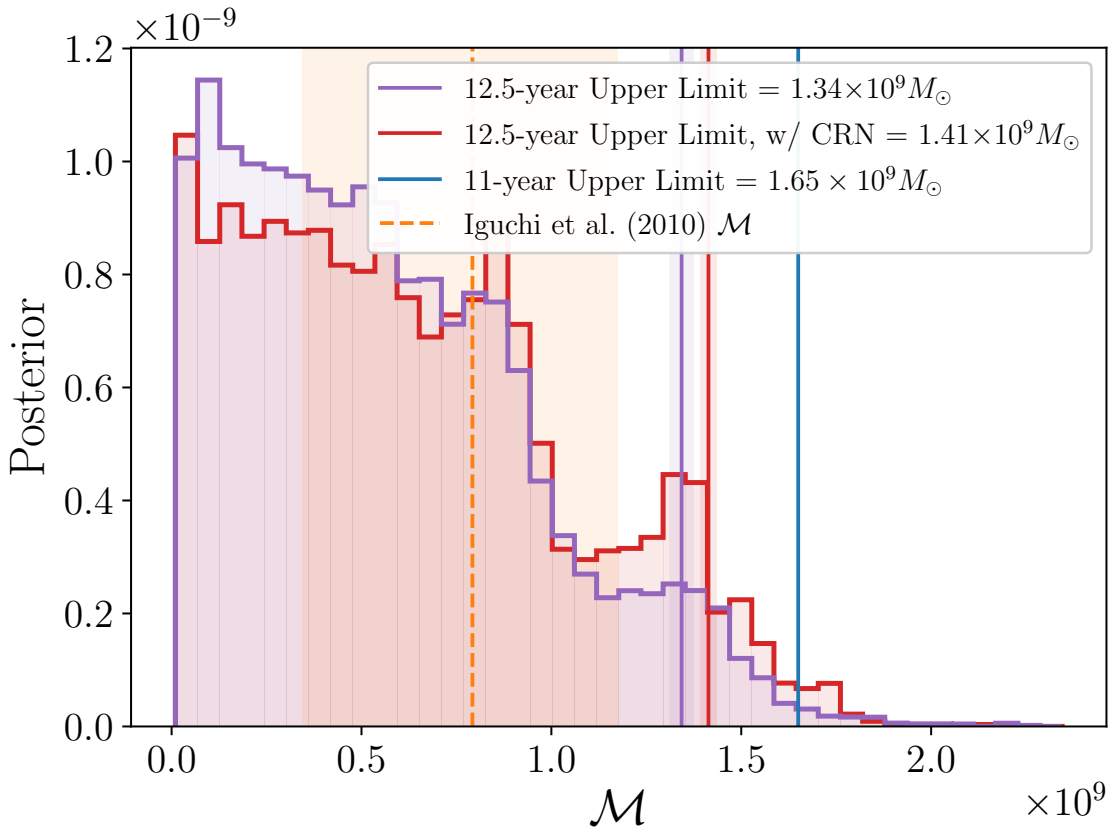

Figure 4.10: Posterior distributions for a targeted upper limit analysis of the SMBHB candidate 3C 66B. While 95% upper limits (red and purple lines) are lower than in the 11-year data set (blue line), they cannot rule out the model from (Iguchi et al., 2010) (orange region).

In Arzoumanian et al. (2020b), it was shown that a targeted search, like this analysis, results in a factor of ~ 2 reduction in upper limits than those of an all-sky search at a corresponding GW frequency. When converted to strain amplitudes rather than chirp masses, the 95% upper limits are 1.90×10^{-14} and 1.74×10^{-14} for the searches with and without a CRN, respectively. In comparison, the all-sky analysis in subsection 4.4.1 returned strain upper limits of 3.56×10^{-14} and 3.82×10^{-14} at 60.1 nHz, the nearest frequency to that of 3C 66B at 60.4 nHz. These all-sky strain upper limits are a factor of 1.88 and 2.20 larger, very similar

to the value for the 11-year data set. Therefore, the improvement in upper limits gained by using multi-messenger techniques has stayed stable accross the addition of new pulsars, more data, and the emergence of the CRN.

4.6 Discussion and Future Prospects

While the NANOGrav PTA is continuing to add data from ongoing observations, discover new pulsars, and therefore increase our sensitivity to GWs, we have entered an interesting era where surprising results will continue to be uncovered. Due to the CRN first detected in the 12.5-year data set in Arzoumanian et al. (2020a), our limitations on CW strains across the nanohertz GW frequency band (Figure 4.3) and the sky (Figure 4.5) have not improved as steadily as in previous data sets. While adding a CRN to the search model that is fixed to the maximum-likelihood values from a dedicated search avoids confusion in detection analyses, this adds a significant source of noise to the PTA, and therefore limits our sensitivity to CWs at frequencies below 10 nHz.

In future data sets, the CRN will likely be even more apparent in the data, and may eventually resolve to be due to a stochastic GWB from SMBHBs (Pol et al., 2021). In any case, this will continue to impact CW searches, and significant efforts will be needed to continue development on methods that will allow for efficient detection of both types of nanohertz GW signal such as in Bécsy & Cornish (2020), as well as extensive simulations that evaluate detection possibilities, as in Pol et al. (2021), that include multiple types of GW signal in the simulated data

sets. Additionally, significant effort will be needed to improve sampling methods that can efficiently explore the complex CW parameter space, particularly at high GW frequencies, complexities which will only be exacerbated as data sets expand.

4.7 Conclusions

With extensive Bayesian analyses, we have searched the NANOGrav 12.5-year data set for CWs from individual SMBHBs. In our detection analyses, we showed that no CWs were detected to a high degree of confidence. We then placed all-sky upper limits on the strain amplitude for all CWs emitting between 2.5 and 319 nHz, as well as upper limits as a function of sky location for the 12.5-year data set's most sensitive frequency of 7.65 nHz.

This analysis also included the development of new methods to accurately reflect the realistic distribution of possible values of pulsar distances from updated measurements. The way we treat these values in search pipelines has a significant impact on our ability to detect the pulsar term of a CW signal, and these methods will be critical as we proceed towards PTA sensitivities which enable a CW detection.

Unlike previous data sets, the 12.5-year data set contains a significant CRN. Therefore, for the first time, we included the CRN in our Bayesian searches by fixing the model parameters to those recovered in Arzoumanian et al. (2020a). This had a significant effect on the results of many of our analyses, and proved critical to avoid a false detection of a CW at 2.45 nHz. This process also significantly impeded the improvement of our upper limits between the 11-year and 12.5-year NANOGrav

searches at the most sensitive frequency of 7.65 nHz in most areas of the sky.

Despite these new necessities, we are able to place significant astrophysical constraints on the local SMBHB population. In our most sensitive sky location, we can rule out the existance of any SMBHB with a mass of at least $10^9 M_\odot$ emitting at 7.65 nHz within 86.65 Mpc. Furthermore, we demonstrate significant improvements to chirp mass upper limits of SMBHB candidates can be made through multi-messenger analysis techniques, and limit the chirp mass of 3C 66B to $(1.34 \pm 0.01) \times 10^9 M_\odot$. With the inclusion of more data, we will soon be able to rule out or confirm this source and other binary candidates.

Chapter 5

Conclusion

We have created Bayesian methods critical to the analysis of supermassive black hole binaries (SMBHBs), from model selection techniques to identify electromagnetic SMBHB markers from within complex noise processes to advanced searches for continuous gravitational waves (CWs) in pulsar timing array (PTA) data across the entire sky. These were applied to detailed simulations and real PTA data, to create realistic expectations on our ability to discover SMBHBs in time domain surveys, and place sensitive limits on CWs in the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) 12.5-year data set. We also combined these two identification methods to create multi-messenger techniques, which we applied to a specific SMBHB candidate in NANOGrav's first multi-messenger search on the 11-year data set. Through this broad effort, we have expanded our ability to investigate the wide aspects linked to SMBHB science. In the near future, as the window to the low-frequency gravitational wave (GW) universe steadily opens, we will be well prepared to use each messenger at our disposal to characterize these dynamic pairs.

5.1 NANOGrav Multi-Messenger Searches: 3C66B

All-sky and frequency-agnostic searches for CWs, while necessary, are becoming increasingly computationally expensive for ever-expanding PTA data sets. By specifically targeting SMBHB candidates in CW searches, the computational footprint is reduced by multiple orders of magnitude, and as we have shown in this thesis, limitations on the SMBHB mass can be even more stringent through multi-messenger techniques, even without a GW detection. In chapter 2, we constructed foundational multi-messenger methods to search for CWs from specific SMBHB candidates in PTA data sets, and applied them to a well-known candidate, 3C 66B. After searching for CWs from 3C 66B and finding no evidence for detectable GW emission, we were able to limit 3C 66B's chirp mass, at 95% confidence, to $(1.65 \pm 0.02) \times 10^9\ M_\odot$, a factor of 4.3 smaller than the limit set in J04, and a factor of 2 lower than the mass limit calculated for the same GW frequency from an all-sky search using the same data (Aggarwal et al., 2019). However, we were unable to rule out the existence of a binary corresponding to the revised model proposed in I10.

While 3C 66B's measured orbital period of 1.05 years is reasonably well constrained (Sudou et al., 2003), there are still associated uncertainties on this measurement. For other candidates, this quantity may be even less constrained. We investigated this possibility by varying the GW frequency within the CW search pipeline for the first time on real data, and discovered that knowledge of the orbital period of a binary target improves upper limits by up to an order of magnitude.

However, perfect measurements are not necessary, and upper limits are consistent with prior widths of up to an order of magnitude in GW frequency. This vastly expands the possibilities for targeted GW searches for SMBHB candidates in the near future, as it will take many years to improve orbital models of long-period binaries emerging in time domain surveys. We also showed that testing a specific binary model by constructing priors corresponding to measured quantities is possible, and new information can be gained without a detection of GWs. These methods will be critical in the near future as PTAs approach the sensitivities required to detect CWs and new SMBHB candidates emerge.

5.2 Searching for Periodic Variability in Quasars

Time-domain surveys of quasars provide an excellent place to search for SMBHB candidates. However, these same quasars exhibit significant intrinsic variability that can be modeled as a damped random walk (DRW), which may mimic or mask periodicities used to identify the presence of a binary (Vaughan et al., 2016). In chapter 3, we conducted large-scale simulations to test a Bayesian model selection pipeline to identify periodic light curves that contain DRW noise processes. By simulating data corresponding to two different optical surveys, we evaluated the effectiveness of current and future SMBHB identification efforts.

For a wide range of DRW and sinusoid parameters, we determined that a strong DRW process can effectively mask low-amplitude and long-period sinusoids, regardless of the observation strategy used. However, denser observations signifi-

cantly lowered the rate of false positive SMBHB detections, meaning that candidates observed in future, high-cadence surveys such as the Legacy Survey of Space and Time (LSST) can be considered more reliable. These signals will also become more distinct from DRW processes in future data from LSST. We also confirmed that if the simulated quasars did not contain red noise, nearly all sinusoids were identified, and there were no false detections; therefore, these effects are purely due to the similarities between a DRW process and a periodicity in unevenly-sampled observations. Finally, we emphasized the need to fit all DRW and sinusoid parameters simultaneously. If this is not done, significant biases may be introduced into either measurement.

These results are extremely promising for future SMBHB searches in time-domain surveys, and provide necessary recommendations for how they may be discovered reliably. These sources will provide ideal candidates for multi-messenger GW searches with PTAs, due to the two techniques' common frequency sensitivities.

5.3 NANOGrav 12.5-year Continuous Wave Limits

CWs provide one of the best opportunities to study individual SMBHBs, and a detection of these bright GWs will provide a unique insight into these complex systems through a new lens. However, the recent strong detection of a common red noise process (CRN) in PTA data sets has complicated searches for CWs, and if future searches for the GWB proceed as anticipated through simulations, this

impediment will only grow more complex.

In chapter 4, we performed Bayesian searches for CWs from sources located anywhere on the sky, across two orders of magnitude in GW frequency in NANOGrav's 12.5-year data set. No CWs were detected in our detection analyses; however, we discovered that to avoid a false detection at the lowest frequency of 2.45 nHz, it was critical to include the CRN in our signal model. Therefore, for the first time, we search for CWs alongside the presence of a CRN. In the near future, we will expand these searches to also include a Frequentist search for CWs across the nanohertz GW frequency band.

Next, we placed all-sky upper limits on the CW strain at each frequency, as well as upper limits as a function of sky location at the PTA's most sensitive frequency of 7.65 nHz. These analyses were also heavily effected by the presence of the CRN. We confirmed the source of these effects through searches on simulated PTA data sets containing a CRN, and found that the effects we observed were well within the range of possibilities observed in the simulated searches.

Additionally, even without a CW detection, we placed stringent astrophysical constraints on the existence of nearby SMBHBs. In our most sensitive sky location, we can rule out the existance of any SMBHB with a mass of at least $10^9 M_\odot$ emitting at 7.65 nHz within 86.65 Mpc. We also maintained the improvements made by conducting multi-messenger CW searches for specific SMBHB candidates, and limited the chirp mass of 3C 66B to $(1.41 \pm 0.02) \times 10^9 M_\odot$.

5.4 Looking Forward

While no multi-messenger detections of SMBHBs have been made thus far, the future is blindingly bright. PTAs are rapidly approaching the sensitivities required to detect a GWB (Pol et al., 2021), and CWs are expected to follow soon after (Kelley et al., 2018). Concurrently, next-generation time-domain surveys, including LSST, are preparing to commence within the next few years (LSST Science Collaborations et al., 2017) and will dramatically increase our capabilities of detecting SMBHB candidates. Through strong cooperation by both efforts, the field will enter a unique era of opportunity for multi-messenger studies of SMBHBs.

However, significant efforts are still needed to ensure that a multi-messenger detection will be possible. As the GWB becomes detectable in PTA data sets, new methods, in addition to those developed in chapter 4 will be critical to differentiate between a detection of a CW or an effect of the GWB. To test these methods, extensive simulations of PTA data sets will be necessary to identify clear indicators for the presence of a CW in addition to the GWB.

While LSST will provide an incredible data set with which to search for periodicities induced by SMBHBs, the data volume of over 20 million quasar light curves will prohibit detailed Bayesian analyses, such as those conducted in chapter 3, on every quasar in the database. Instead, a fast evaluation technique will be needed to identify promising candidates for follow-up with the full pipeline. Additionally, further modeling of the intrinsic quasar population and more accurate SMBHB orbital models will allow us to improve predictions about the overall expectations for

finding these sources in time-domain surveys.

The results of chapter 2 have shown that the possibilities created by multi-messenger analyses of SMBHBs are broader than expected, and the methods developed are already being applied to numerous other candidates, and new methods are being developed to jointly analyze electromagnetic and GW data sets. In the near future, as PTA sensitivities grow and more periodic candidates are identified, the possibilities for multi-messenger analyses will increase dramatically. However, predictions are often incomplete; as the first detections are made, surprises are sure to abound. Fortunately, surprising discoveries are often the most thrilling, and will provide ample opportunities for the development of new physics. Therefore, it is critical to use every tool and messenger in our arsenal as we explore the unknown.

Bibliography

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102, doi: 10.1103/PhysRevLett.116.061102

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, ApJL, 848, L12, doi: 10.3847/2041-8213/aa91c9

Abbott, R., Abbott, T. D., Abraham, S., et al. 2021, ApJL, 915, L5, doi: 10.3847/2041-8213/ac082e

Abdo, A. A., Ajello, M., Allafort, A., et al. 2013, ApJS, 208, 17, doi: 10.1088/0067-0049/208/2/17

Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2019, ApJ, 880, 116, doi: 10.3847/1538-4357/ab2236

Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021a, ApJS, 252, 4, doi: 10.3847/1538-4365/abc6a0

—. 2021b, ApJS, 252, 5, doi: 10.3847/1538-4365/abc6a1

Amaro-Seoane, P., Audley, H., Babak, S., et al. 2017, arXiv e-prints, arXiv:1702.00786. https://arxiv.org/abs/1702.00786

Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, MNRAS, 510, 4873, doi: 10.1093/mnras/stab3418

Aretxaga, I., Cid Fernandes, R., & Terlevich, R. J. 1997, MNRAS, 286, 271, doi: 10.1093/mnras/286.2.271

Armitage, P. J., & Natarajan, P. 2002, ApJL, 567, L9, doi: 10.1086/339770

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2014, ApJ, 794, 141, doi: 10.1088/0004-637X/794/2/141

—. 2016, ApJ, 821, 13, doi: 10.3847/0004-637X/821/1/13

Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018a, ApJ, 859, 47, doi: 10.3847/1538-4357/aabd3b

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2018b, ApJS, 235, 37, doi: 10.3847/1538-4365/aab5b0

Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020a, ApJL, 905, L34, doi: 10.3847/2041-8213/abd401

Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2020b, ApJ, 900, 102, doi: 10.3847/1538-4357/ababa1

Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2021, Phys. Rev. Lett., 127, 251302, doi: 10.1103/PhysRevLett.127.251302

Babak, S., Petiteau, A., Sesana, A., et al. 2016, MNRAS, 455, 1665, doi: 10.1093/mnras/stv2092

Baker, P. T., Brook, P. R., Fiore, W. C., et al. 2019, arXiv e-prints, arXiv:1912.12939. https://arxiv.org/abs/1912.12939

Barnes, J. E. 2002, MNRAS, 333, 481, doi: 10.1046/j.1365-8711.2002.05335.x

Bécsy, B., & Cornish, N. J. 2020, Classical and Quantum Gravity, 37, 135011, doi: 10.1088/1361-6382/ab8bbd

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, Nature, 287, 307, doi: 10.1038/287307a0

Benetti, M., Graef, L. L., & Vagnozzi, S. 2022, Phys. Rev. D, 105, 043520, doi: 10.1103/PhysRevD.105.043520

Bogdanovic, T., Miller, M. C., & Blecha, L. 2021, arXiv e-prints, arXiv:2109.03262. https://arxiv.org/abs/2109.03262

Bromm, V., & Loeb, A. 2003, ApJ, 596, 34, doi: 10.1086/377529

Burke-Spolaor, S., Taylor, S. R., Charisi, M., et al. 2019, A&A Rev., 27, 5, doi: 10.1007/s00159-019-0115-7

Caballero, R. N., Guo, Y. J., Lee, K. J., et al. 2018, MNRAS, 481, 5501, doi: 10.1093/mnras/sty2632

Camilo, F., Foster, R. S., & Wolszczan, A. 1994, ApJL, 437, L39, doi: 10.1086/187677

Charisi, M., Bartos, I., Haiman, Z., et al. 2016, MNRAS, 463, 2145, doi: 10.1093/mnras/stw1838

Charisi, M., Bartos, I., Haiman, Z., Price-Whelan, A. M., & Márka, S. 2015, MNRAS, 454, L21, doi: 10.1093/mnrasl/slv111

Charisi, M., Haiman, Z., Schiminovich, D., & D'Orazio, D. J. 2018, MNRAS, 476, 4617, doi: 10.1093/mnras/sty516

Charisi, M., Taylor, S. R., Runnoe, J., Bogdanovic, T., & Trump, J. R. 2022, MNRAS, 510, 5929, doi: 10.1093/mnras/stab3713

Chatterjee, S., Brisken, W. F., Vlemmings, W. H. T., et al. 2009, ApJ, 698, 250, doi: 10.1088/0004-637X/698/1/250

Chen, Y.-C., Liu, X., Liao, W.-T., et al. 2020, MNRAS, 499, 2245, doi: 10.1093/mnras/staa2957

Colpi, M. 2014, Space Sci. Rev. , 183, 189, doi: 10.1007/s11214-014-0067-1

Condon, J. J., & Ransom, S. M. 2016, Essential Radio Astronomy

Corbin, V., & Cornish, N. J. 2010, arXiv e-prints, arXiv:1008.1782. https://arxiv.org/abs/1008.1782

Cordes, J. M., & Lazio, T. J. W. 2002, arXiv e-prints, astro. https://arxiv.org/abs/astro-ph/0207156

d'Ascoli, S., Noble, S. C., Bowen, D. B., et al. 2018, ApJ, 865, 140, doi: 10.3847/1538-4357/aad8b4

De Rosa, A., Vignali, C., Bogdanović, T., et al. 2019, New Astro. Rev., 86, 101525, doi: 10.1016/j.newar.2020.101525

Deller, A. T., Goss, W. M., Brisken, W. F., et al. 2019, ApJ, 875, 100, doi: 10.3847/1538-4357/ab11c7

Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94, doi: 10.1088/0004-637X/762/2/94

Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341, doi: 10.1093/mnras/stw483

Detweiler, S. 1979, ApJ, 234, 1100, doi: 10.1086/157593

Dickey, J. M. 1971, The Annals of Mathematical Statistics, 42, 204. http://www.jstor.org/stable/2958475

Ding, H., Deller, A. T., Freire, P., et al. 2020, ApJ, 896, 85, doi: 10.3847/1538-4357/ab8f27

Donea, A.-C., & Biermann, P. L. 2002, PASA, 19, 125, doi: 10.1071/AS01078

D'Orazio, D. J., Haiman, Z., & MacFadyen, A. 2013, MNRAS, 436, 2997, doi: 10.1093/mnras/stt1787

D'Orazio, D. J., Haiman, Z., & Schiminovich, D. 2015, Nature, 525, 351, doi: 10.1038/nature15262

Einstein, A. 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, 688

Ellis, J., & van Haasteren, R. 2017, doi: 10.5281/zenodo.1037579

Ellis, J. A. 2013, Classical and Quantum Gravity, 30, 224004, doi: 10.1088/0264-9381/30/22/224004

Ellis, J. A., & Cornish, N. J. 2016, Phys. Rev. D, 93, 084048, doi: 10.1103/PhysRevD.93.084048

Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012, ApJ, 756, 175, doi: 10.1088/0004-637X/756/2/175

Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2019, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE. http://ascl.net/1912.015

Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, ApJS, 201, 23, doi: 10.1088/0067-0049/201/2/23

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019, ApJL, 875, L1, doi: 10.3847/2041-8213/ab0ec7

Farris, B. D., Duffell, P., MacFadyen, A. I., & Haiman, Z. 2014, ApJ, 783, 134, doi: 10.1088/0004-637X/783/2/134

Fawcett, T. 2006, Pattern Recognition Letters, 27, 861, doi: 10.1016/j.patrec.2005.10.010

Feng, Y., Li, D., Li, Y.-R., & Wang, J.-M. 2019, arXiv e-prints, arXiv:1907.03460. https://arxiv.org/abs/1907.03460

Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9, doi: 10.1086/312838

Fey, A. L., Ma, C., Arias, E. F., et al. 2004, AJ, 127, 3587, doi: 10.1086/420998

Folkner, W. M., & Park, R. S. 2018, Planetary ephemeris DE438 for Juno, Tech. Rep. IOM 392R-18-004, Jet Propulsion Laboratory, Pasadena, CA

Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300, doi: 10.1086/169195

Freire, P. C. C., Wex, N., Esposito-Farèse, G., et al. 2012, MNRAS, 423, 3328, doi: 10.1111/j.1365-2966.2012.21253.x

Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, ApJ, 689, 1044, doi: 10.1086/592738

Goncharov, B., Shannon, R. M., Reardon, D. J., et al. 2021, ApJL, 917, L19, doi: 10.3847/2041-8213/ac17f4

Goulding, A. D., Greene, J. E., Bezanson, R., et al. 2018, PASJ, 70, S37, doi: 10.1093/pasj/psx135

Graham, M. J., Djorgovski, S. G., Stern, D., et al. 2015a, Nature, 518, 74, doi: 10.1038/nature14143

—. 2015b, MNRAS, 453, 1562, doi: 10.1093/mnras/stv1726

Guillemot, L., Smith, D. A., Laffon, H., et al. 2016, A&A, 587, A109, doi: 10.1051/0004-6361/201527847

Haehnelt, M. G., & Kauffmann, G. 2002, MNRAS, 336, L61, doi: 10.1046/j.1365-8711.2002.06056.x

Haiman, Z., Kocsis, B., & Menou, K. 2009, ApJ, 700, 1952, doi: 10.1088/0004-637X/700/2/1952

Hawkins, M. R. S. 1993, Nature, 366, 242, doi: 10.1038/366242a0

Hazboun, J., Romano, J., & Smith, T. 2019a, The Journal of Open Source Software, 4, 1775, doi: 10.21105/joss.01775

Hazboun, J. S., Romano, J. D., & Smith, T. L. 2019b, Phys. Rev. D, 100, 104028, doi: 10.1103/PhysRevD.100.104028

Hazboun, J. S., Simon, J., Taylor, S. R., et al. 2020, ApJ, 890, 108, doi: 10.3847/1538-4357/ab68db

Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39, doi: 10.1086/183954

Hobbs, G. 2013, Classical and Quantum Gravity, 30, 224007, doi: 10.1088/0264-9381/30/22/224007

Hobbs, G., Guo, L., Caballero, R. N., et al. 2020, MNRAS, 491, 5951, doi: 10.1093/mnras/stz3071

Holz, D. E., & Hughes, S. A. 2005, ApJ, 629, 15, doi: 10.1086/431341

Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, MNRAS, 268, 305, doi: 10.1093/mnras/268.2.305

Hotan, A. W., Bailes, M., & Ord, S. M. 2006, MNRAS, 369, 1502, doi: 10.1111/j.1365-2966.2006.10394.x

Hu, B. X., D'Orazio, D. J., Haiman, Z., et al. 2020, MNRAS, 495, 4061, doi: 10.1093/mnras/staa1312

Huchra, J. P., Vogeley, M. S., & Geller, M. J. 1999, ApJS, 121, 287, doi: 10.1086/313194

Iguchi, S., Okuda, T., & Sudou, H. 2010, ApJL, 724, L166, doi: 10.1088/2041-8205/724/2/L166

Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2014, Statistics, Data Mining and Machine Learning in Astronomy (Princeton University Press)

Jacoby, B. A., Hotan, A., Bailes, M., Ord, S., & Kulkarni, S. R. 2005, ApJL, 629, L113, doi: 10.1086/449311

Jeffreys, H. 1939, The Theory of Probability

Jenet, F. A., Lommen, A., Larson, S. L., & Wen, L. 2004, ApJ, 606, 799, doi: 10.1086/383020

Jennings, R. J., Kaplan, D. L., Chatterjee, S., Cordes, J. M., & Deller, A. T. 2018, ApJ, 864, 26, doi: 10.3847/1538-4357/aad084

Jones, M. L., McLaughlin, M. A., Lam, M. T., et al. 2017, ApJ, 841, 125, doi: 10.3847/1538-4357/aa73df

Kaspi, V. M., Taylor, J. H., & Ryba, M. F. 1994, ApJ, 428, 713, doi: 10.1086/174280

Kass, R. E., & Raftery, A. E. 1995, Journal of the American Statistical Association, 90, 773, doi: 10.1080/01621459.1995.10476572

Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L. 1998, ApJ, 504, 671, doi: 10.1086/306105

Kelley, L., Charisi, M., Burke-Spolaor, S., et al. 2019a, BAAS, 51, 490. https://arxiv.org/abs/1903.07644

Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2018, MNRAS, 477, 964, doi: 10.1093/mnras/sty689

Kelley, L. Z., D'Orazio, D. J., & Di Stefano, R. 2021, arXiv e-prints, arXiv:2107.07522. https://arxiv.org/abs/2107.07522

Kelley, L. Z., Haiman, Z., Sesana, A., & Hernquist, L. 2019b, MNRAS, 485, 1579, doi: 10.1093/mnras/stz150

Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, ApJ, 698, 895, doi: 10.1088/0004-637X/698/1/895

Kerr, M., Reardon, D. J., Hobbs, G., et al. 2020, PASA, 37, e020, doi: 10.1017/pasa.2020.11

Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511, doi: 10.1146/annurev-astro-082708-101811

Kozłowski, S. 2017, A&A, 597, A128, doi: 10.1051/0004-6361/201629890

Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010, ApJ, 708, 927, doi: 10.1088/0004-637X/708/2/927

Kulier, A., Ostriker, J. P., Natarajan, P., Lackner, C. N., & Cen, R. 2015, ApJ, 799, 178, doi: 10.1088/0004-637X/799/2/178

Kullback, S., & Leibler, R. A. 1951, Ann. Math. Statist., 22, 79, doi: 10.1214/aoms/1177729694

Kun, E., Frey, S., Gabányi, K. É., et al. 2015, MNRAS, 454, 1290, doi: 10.1093/mnras/stv2049

Lam, M. T., Cordes, J. M., Chatterjee, S., et al. 2016, ApJ, 821, 66, doi: 10.3847/0004-637X/821/1/66

Lam, M. T., Ellis, J. A., Grillo, G., et al. 2018, ApJ, 861, 132, doi: 10.3847/1538-4357/aac770

Lam, M. T., McLaughlin, M. A., Arzoumanian, Z., et al. 2019, ApJ, 872, 193, doi: 10.3847/1538-4357/ab01cd

Lazaridis, K., Wex, N., Jessner, A., et al. 2009, MNRAS, 400, 805, doi: 10.1111/j.1365-2966.2009.15481.x

Lee, K. J., Wex, N., Kramer, M., et al. 2011, MNRAS, 414, 3251, doi: 10.1111/j.1365-2966.2011.18622.x

Lehto, H. J., & Valtonen, M. J. 1996, ApJ, 460, 207, doi: 10.1086/176962

Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, MNRAS, 453, 2576, doi: 10.1093/mnras/stv1538

Liddle, A. R. 2007, MNRAS, 377, L74, doi: 10.1111/j.1745-3933.2007.00306.x

Liu, F. K., Li, S., & Komossa, S. 2014, ApJ, 786, 103, doi: 10.1088/0004-637X/786/2/103

Liu, T., Gezari, S., & Miller, M. C. 2018, ApJL, 859, L12, doi: 10.3847/2041-8213/aac2ed

Liu, T., & Vigeland, S. J. 2021, ApJ, 921, 178, doi: 10.3847/1538-4357/ac1da9

Liu, T., Gezari, S., Ayers, M., et al. 2019, ApJ, 884, 36, doi: 10.3847/1538-4357/ab40cb

Löhmer, O., Kramer, M., Driebe, T., et al. 2004, A&A, 426, 631, doi: 10.1051/0004-6361:20041031

Lommen, A. N., & Backer, D. C. 2001, ApJ, 562, 297, doi: 10.1086/323491

Lommen, A. N., Kipphorn, R. A., Nice, D. J., et al. 2006, ApJ, 642, 1012, doi: 10.1086/501067

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

LSST Dark Energy Science Collaboration (LSST DESC), Abolfathi, B., Alonso, D., et al. 2021, ApJS, 253, 31, doi: 10.3847/1538-4365/abd62c

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201. https://arxiv.org/abs/0912.0201

LSST Science Collaborations, Marshall, P., Anguita, T., et al. 2017, ArXiv e-prints, doi: 10.5281/zenodo.842712

Lutz, T. E., & Kelker, D. H. 1973, PASP, 85, 573, doi: 10.1086/129506

MacFadyen, A. I., & Milosavljević, M. 2008, ApJ, 672, 83, doi: 10.1086/523869

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, ApJ, 721, 1014, doi: 10.1088/0004-637X/721/2/1014

Maggiore, M. 2007, Gravitational Waves. Vol. 1: Theory and Experiments, Oxford Master Series in Physics (Oxford University Press)

Mahalanobis, P. C. 1936, in On the generalized distance in statistics, National Institute of Science of India

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993, doi: 10.1086/428488

Manchester, R. N., Hobbs, G., Bailes, M., et al. 2013, PASA, 30, e017, doi: 10.1017/pasa.2012.017

McLaughlin, M. A. 2013, Classical and Quantum Gravity, 30, 224008, doi: 10.1088/0264-9381/30/22/224008

Mingarelli, C. M. F., Lazio, T. J. W., Sesana, A., et al. 2017, Nature Astronomy, 1, 886, doi: 10.1038/s41550-017-0299-6

Mohan, P., An, T., Frey, S., et al. 2016, MNRAS, 463, 1812, doi: 10.1093/mnras/stw2154

Moore, C. J., Cole, R. H., & Berry, C. P. L. 2015, Classical and Quantum Gravity, 32, 015014, doi: 10.1088/0264-9381/32/1/015014

National Academies, T. 2021, doi: 10.17226/26141

Newcombe, R. G. 1998, Statistics in Medicine, 17, 857, doi: https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, MNRAS, 490, 4666, doi: 10.1093/mnras/stz2857

Pesce, D. W., Braatz, J. A., Condon, J. J., & Greene, J. E. 2018, ApJ, 863, 149, doi: 10.3847/1538-4357/aad3c2

Pol, N. S., Taylor, S. R., Kelley, L. Z., et al. 2021, ApJL, 911, L34, doi: 10.3847/2041-8213/abf2c9

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C. The art of scientific computing

Qian, S. J., Britzen, S., Krichbaum, T. P., & Witzel, A. 2019, A&A, 621, A11, doi: 10.1051/0004-6361/201833508

Rajagopal, M., & Romani, R. W. 1995, ApJ, 446, 543, doi: 10.1086/175813

Rodriguez, C., Taylor, G. B., Zavala, R. T., et al. 2006, The Astrophysical Journal, 646, 49, doi: 10.1086/504825

Roedig, C., Sesana, A., Dotti, M., et al. 2012, A&A, 545, A127, doi: 10.1051/0004-6361/201219986

Romano, J. D., & Cornish, N. J. 2017, Living Reviews in Relativity, 20, 2, doi: 10.1007/s41114-017-0004-1

Rosado, P. A., Sesana, A., & Gair, J. 2015, MNRAS, 451, 2417, doi: 10.1093/mnras/stv1098

Saade, M. L., Stern, D., Brightman, M., et al. 2020a, ApJ, 900, 148, doi: 10.3847/1538-4357/abad31

—. 2020b, ApJ, 900, 148, doi: 10.3847/1538-4357/abad31

Sathyaprakash, B. S., & Schutz, B. F. 2009, Living Reviews in Relativity, 12, 2, doi: 10.12942/lrr-2009-2

Sesana, A., Haardt, F., Madau, P., & Volonteri, M. 2004, ApJ, 611, 623, doi: 10.1086/422185

Sesana, A., Haiman, Z., Kocsis, B., & Kelley, L. Z. 2018, ApJ, 856, 42, doi: 10.3847/1538-4357/aaad0f

Sesana, A., & Vecchio, A. 2010, Phys. Rev. D, 81, 104008, doi: 10.1103/PhysRevD.81.104008

Shankar, F., Bernardi, M., Sheth, R. K., et al. 2016, MNRAS, 460, 3119, doi: 10.1093/mnras/stw678

Shannon, R. M., Ravi, V., Coles, W. A., et al. 2013, Science, 342, 334, doi: 10.1126/science.1238012

Shannon, R. M., Ravi, V., Lentati, L. T., et al. 2015, Science, 349, 1522, doi: 10.1126/science.aab1910

Simon, J., & Hazboun, J. in prep

Speagle, J. S. 2019, arXiv e-prints, arXiv:1909.12313. https://arxiv.org/abs/1909.12313

Splaver, E. M., Nice, D. J., Stairs, I. H., Lommen, A. N., & Backer, D. C. 2005, ApJ, 620, 405, doi: 10.1086/426804

Stovall, K., Lynch, R. S., Ransom, S. M., et al. 2014, ApJ, 791, 67, doi: 10.1088/0004-637X/791/1/67

Sudou, H., Iguchi, S., Murata, Y., & Taniguchi, Y. 2003, Science, 300, 1263, doi: 10.1126/science.1082817

Tanaka, T., Menou, K., & Haiman, Z. 2012, MNRAS, 420, 705, doi: 10.1111/j.1365-2966.2011.20083.x

Tanaka, T. L., & Haiman, Z. 2013, Classical and Quantum Gravity, 30, 224012, doi: 10.1088/0264-9381/30/22/224012

Tang, Y., Haiman, Z., & MacFadyen, A. 2018, ArXiv e-prints. https://arxiv.org/abs/1801.02266

Taylor, S., Burke-Spolaor, S., Baker, P. T., et al. 2019, BAAS, 51, 336. https://arxiv.org/abs/1903.08183

Taylor, S. R. 2021, arXiv e-prints, arXiv:2105.13270. https://arxiv.org/abs/2105.13270

Taylor, S. R., Huerta, E. A., Gair, J. R., & McWilliams, S. T. 2016, ApJ, 817, 70, doi: 10.3847/0004-637X/817/1/70

Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, MNRAS, 455, 4339, doi: 10.1093/mnras/stv2143

Timmer, J., & Koenig, M. 1995, A&A, 300, 707

Toscano, M., Britton, M. C., Manchester, R. N., et al. 1999, ApJL, 523, L171, doi: 10.1086/312276

Trèvese, D., & Vagnetti, F. 2002, ApJ, 564, 624, doi: 10.1086/324541

Urry, C. M., & Padovani, P. 1995, PASP, 107, 803, doi: 10.1086/133630

Vallisneri, M., Taylor, S. R., Simon, J., et al. 2020, ApJ, 893, 112, doi: 10.3847/1538-4357/ab7b67

van Haasteren, R., Levin, Y., Janssen, G. H., et al. 2011, MNRAS, 414, 3117, doi: 10.1111/j.1365-2966.2011.18613.x

Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47 –54, doi: 10.1109/CIDU.2012.6382200

VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18, doi: 10.1088/0004-637X/812/1/18

Vaughan, S., Uttley, P., Markowitz, A. G., et al. 2016, MNRAS, 461, 3145, doi: 10.1093/mnras/stw1412

Verbiest, J. P. W., Weisberg, J. M., Chael, A. A., Lee, K. J., & Lorimer, D. R. 2012, ApJ, 755, 39, doi: 10.1088/0004-637X/755/1/39

Verbiest, J. P. W., Bailes, M., Coles, W. A., et al. 2009, MNRAS, 400, 951, doi: 10.1111/j.1365-2966.2009.15508.x

Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016a, MNRAS, 458, 1267, doi: 10.1093/mnras/stw347

—. 2016b, MNRAS, 458, 1267, doi: 10.1093/mnras/stw347

Vigeland, S. J., & Vallisneri, M. 2014, MNRAS, 440, 1446, doi: 10.1093/mnras/stu312

Wahlquist, H. 1987, General Relativity and Gravitation, 19, 1101, doi: 10.1007/BF00759146

Wegg, C., & Nate Bode, J. 2011, ApJL, 738, L8, doi: 10.1088/2041-8205/738/1/L8

Xin, C., Charisi, M., Haiman, Z., et al. 2020a, MNRAS, 496, 1683, doi: 10.1093/mnras/staa1643

Xin, C., & Haiman, Z. 2021, MNRAS, 506, 2408, doi: 10.1093/mnras/stab1856

Xin, C., Mingarelli, C. M. F., & Hazboun, J. S. 2020b, arXiv e-prints, arXiv:2009.11865. https://arxiv.org/abs/2009.11865

Xue, X., Bian, L., Shu, J., et al. 2021, Phys. Rev. Lett., 127, 251303, doi: 10.1103/PhysRevLett.127.251303

Yao, J. M., Manchester, R. N., & Wang, N. 2017, ApJ, 835, 29, doi: 10.3847/1538-4357/835/1/29

Yardley, D. R. B., Hobbs, G. B., Jenet, F. A., et al. 2010, MNRAS, 407, 669, doi: 10.1111/j.1365-2966.2010.16949.x

Zhu, X.-J., Cui, W., & Thrane, E. 2019, MNRAS, 482, 2588, doi: 10.1093/mnras/sty2849

Zhu, X.-J., & Thrane, E. 2020, ApJ, 900, 117, doi: 10.3847/1538-4357/abac5a

Zhu, X.-J., Hobbs, G., Wen, L., et al. 2014, MNRAS, 444, 3709, doi: 10.1093/mnras/stu1717