Draft version April 24, 2021
Typeset using LaTeX **twocolumn** style in AASTeX631

Pulsar Timing Arrays: An Overview

Zachary Stone[1]

[1] *Stony Brook University*
100 Circle Rd
Stony Brook, NY 11790, USA

ABSTRACT

The first evidence of gravitational radiation came from observing the decay of the orbital period of a pulsar in a system with another neutron star (NS). Now, we use arrays of pulsars, termed pulsar timing arrays (PTAs), to similarly determine the existence of gravitational waves (GWs), but by different means. PTAs correlate the timing residuals between pulsars in their arrays to confirm the existence of GWs passing through the path of the radio pulses of the pulsars as they traverse the Galaxy on their way to Earth. In this article, I discuss the concepts of pulsar timing and PTAs, and how they can measure GWs. In particular, PTAs are prone to measure the gravitational wave background (GWB), a stochastic background formed from the incoherent superposition of a large number of isotropic GWs, due to their sensitivity range in frequency-space. I discuss the types of noise introduced in pulsar timing and how they need to be accounted for in order to make a significant detection of the GWB. I then discuss different possible sources of the GWB, and what parameters we can infer from the GWB by using PTAs. The most likely source of the GWB are merging supermassive black holes (SMBHs), present over the whole Universe. I discuss the formation and dynamics of the SMBH binaries that lead to these mergers and what characteristics they imprint upon the GWB. Finally, I state certain constraints placed upon the GWB and expectations for the future detection of GWs with PTAs.

Keywords: astrophysics, pulsars, observations, compact objects — gravitational waves, black holes

1. INTRODUCTION

One of the most studied objects in space, outside of objects local to our solar system, are stars. Stars are one of the most ubiquitous objects in the Galaxy, let alone the entire local Universe. We observe millions of stars, piecing together how they evolve over their lifetimes from the snapshots we get of each star when we observe them. Some stars never become stars at all, and can never start the fusion process that would make them like our Sun. Some stars, like our Sun, will eventually shed their outer layers in the red giant phase, become a white dwarf, and cool off for billions of years. However, many stars much more massive that stars like our Sun on the main-sequence will undergo very violent transformations near the end of their lifespans.

As stars run out of fuel to create energy to create a radiative pressure to counter the force of gravity, they often switch to heavier fuels. For example, as our own Sun runs out of Hydrogen to power its fusion reactions in its core, it will eventually move to fusing Helium. In more massive stars, this fusing of heavier elements continues up to Iron. Iron is shown to have the optimal binding energy per nucleon. As such, all lighter elements will have a lower "energy requirement" to overcome the binding energy. However, all elements heavier than Iron have a higher energy requirement. Since Iron has the optimal binding energy, all elements of equal or greater mass than Iron will require more energy put into their fusion reaction than the energy released. This means that there can be no outward pressure in the star's core from these reactions to counter gravity, and the star collapses.

These massive stars can have two possible ends to their lives on the main-sequence, either becoming a neutron star (NS) or a black hole (BH). Neutron stars form as the outer layers of a star collapse in on the core. This collapse will significantly increase the density of the core of the star, eventually making it degenerate.

zachary.stone@stonybrook.edu

However, the pressure from this collapse is high enough to break this degeneracy and cause the core to become even more dense. Eventually, the core stops contracting as it reaches nuclear densities, not being able to become denser due to neutron degeneracy. The outer layers of the star proceed to bounce of the surface of this newly-formed neutron star. Thus, we are left will a single compact object after this core collapse supernova (CCSNe). For stars more massive than ones that produce NSs, the pressure of the collapse will be great enough to break the neutron degeneracy in the core. The core will collapse so quickly and with so much force that it will become a black hole.

Seeing as NSs contain matter in nuclear densities, they are very compact objects. As such, they have very high rotation rates due to the conservation of angular momentum from when they formed in CCSNe. Furthermore, this high energy and significant rotation causes NSs to have strong magnetic fields, some being on the order of 10^{12} G. Furthermore, the dense matter within them can help us understand the properties of condensed matter and interactions in particle physics previously unseen. However, the equation of state (EoS) governing how this dense matter behaves is still somewhat poorly understood. Therefore, we can use NSs as probes of highly-condensed matter, as well as general relativity (GR), due to the strong gravitational effect they have from their compact nature.

1.1. *Pulsars*

Many NSs also emit very powerful beams of radio waves from their magnetic poles, due to the strong magnetic fields they have at their surfaces. If the magnetic and rotational axes of the NS are misaligned, then this beam can sweep across space like a lighthouse. We call these types of NSs pulsars, as they send pulses of light along our line of sight (LOS) whenever this radio emission passes in front of us. Not only do these pulses of radio emission sweep across our LOS very frequently due to their high rotational frequencies, but they also appear regularly. Pulsars are very stable rotators due to their large moment of inertia, large rotational energies, and low energy loss rates [16]. Consequently, pulsars continually have a train of pulsed information, that will not deviate much from the frequency with which we observe it. We can then use this train of information to probe interesting physics that affects the pulsar, the environment surrounding the pulsar, or anything that interacts with the radio pulse as it propagates through space towards Earth.

There are also various types of pulsars, considering that there are various methods in which they could form and various progenitor stars they could form from. Right after pulsars are born from a CCSNe, they are known as "young" pulsars, and have periods $\sim$10 ms and magnetic fields $B \sim 10^{12}$G at the surface. After the first few years of them being a "young pulsar" they slow down appreciably into a "slow" or "canonical" pulsar, which has periods $\sim$.1s and magnetic fields $\sim 10^{10-12}$G [56]. There are also more exotic pulsars, such as Anomalous X-ray Pulsars (AXPs) and Soft Gamma Ray Pulsars, which are types of "young" pulsars with large periods (5-12s) and significant magnetic fields (10^{14-15} G) [6]. There are also magnetars, which are pulsars with extreme magnetic fields, which decay and rearrange to produce large and variable X-ray luminosity. One of the most important types of pulsars in observational astronomy are pulsars which had/have stellar companions. In theses scenarios, the pulsar will accrete matter from its stellar companion, and spin-up due to transfer of mass and angular momentum. These millisecond pulsars (MSPs) have periods on the order of 1-30 ms and weak magnetic fields ($\sim 10^9$ G). However, these MSPs have a rotational stability up to 4 orders of magnitude better than that of normal pulsars, as their spin-down rates are very low [57]. Furthermore, the majority ($\sim$ 80%) of MSPs still reside in a binary, which can probe physics, such as GR in the strong-field regime, the EoS of dense matter as it responds to gravity and accreted matter of the companion star, and the strong equivalence principle [11]. These properties make MSPs some of the most consistent objects for studying astrophysical phenomena, as their pulses are so regular. Therefore, we can infer affects from astrophysical sources by looking at deviations from this regular train of radio pulses.

In addition to the highly regular radio pulses, each pulsar also imprints a characteristic shape into the information it broadcasts. The pulses from each pulsar will appear with a shape defined by the plasma properties in the pulsar's magnetosphere, which produce the beams of radio emission. There is still some uncertainty on this matter, however, as various effects from the nature of each pulsar could affect the emission of the beam of radio pulses. Recently, Guillot et al. [17] has suggested that radio profiles from pulsars could originate from the outer edge of the beam of emission instead of from the core. Therefore, we can see that if a pulsar's radio pulses deviate from its intrinsic pulse pattern, then there must be an astrophysical source affecting either the pulsar, the Earth, or the propagating electromagnetic pulse as it propagates through space.

2. PULSAR TIMING

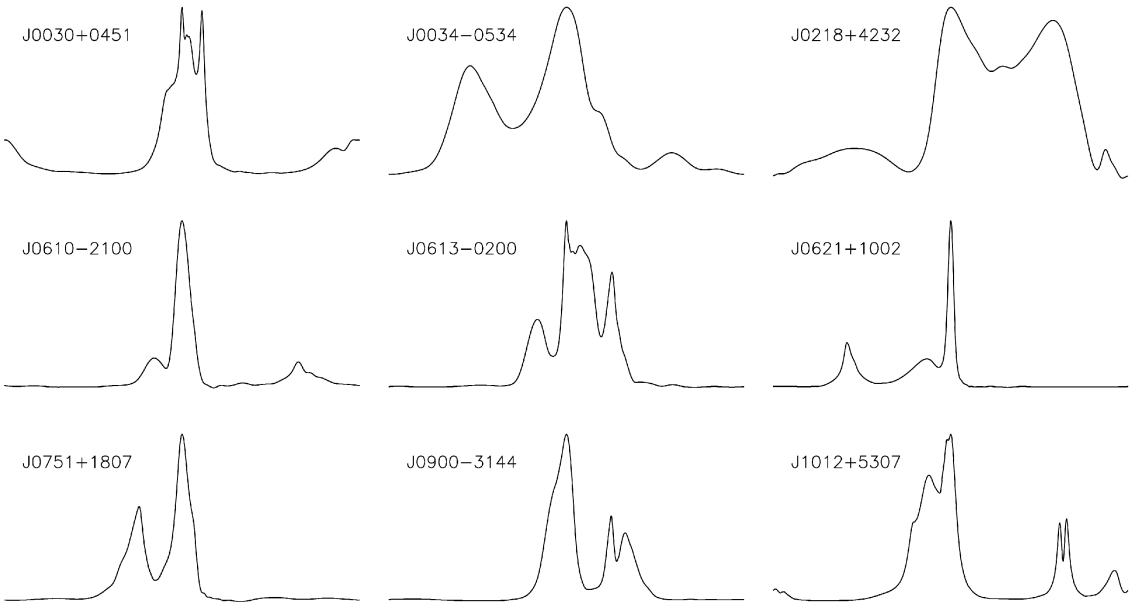

Figure 1. An example of 9 different pulse profiles from pulsars observed by the European Pulsar Timing Array (EPTA). The features of each pulsar's pulse profile is governed by the mechanism that produces it, such as the properties of the pulsar's magnetopshere. (Desvignes et al. [11])

We have seen throughout §1.1 that we can utilize the regularity of pulsars, especially MSPs, to observe certain astrophysical affects on the Earth-pulsar system and the material along our LOS. Pulsar timing is the process of analyzing the deviations from a pulsar's "pulse profile" (also known as its pulse shape) to infer astrophysical scenarios. We have seen already that we need to utilize the pulse's intrinsic shape to account for deviations in its patterned radiation, but there are also various types of noise that will also be accounted for, such as noise from the detector, the motion of the pulsar, and propagation delays along the LOS to the pulsar. Therefore, we must systematically remove all noise and extrinsic/intrinsic affects that serve to alter the regularly observed pulses to give us delays due to astrophysical scenarios.

When observatories measure pulsars, they measure a form of lightcurve called pulse profiles which measure the radio flux of each pulse from the pulsar overtime. Many analyses also consider the change in frequency and polarization of the pulses overtime as well, which will help account for more propagation delays on the pulses. This lightcurve is composed of individual times of arrival of each of the pulses, and the radio flux at the time or arrival (TOA). In order to analyze the delays in the TOAs with respect to the pulsar's pulse profile, timing models are created using parameters of the system and along the LOS towards Earth that affect the travel time of the radio pulse. This "timing model" is then used to calculate the delays in the TOAs of the pulsar by subtracting a measurement of the pulse profile from the timing model. These timeseries of TOA delays are called "timing residuals", and are used to analyze processes intrinsic and extrinsic to the pulsar.

However, creating a timing model for each pulsar would be very time- and data-consuming. Therefore, when timing pulsars, a large number of measurements of the pulse profile are averaged overtime to produce a template, mean profile. This mean profile is produced by integrating the observed flux density of the pulse profile as a function of the pulse phase [55]. The phase is simply the timeseries of the pulses modulo the period of the pulse profile. This is usually called "folding" the data into one pulse phase. However, the individual pulses emitted by the pulsar are much too faint for radio telescopes to measure. Therefore, they need to bin the radio flux over a certain period of time, granting a certain resolution of the pulse profile with respect to time. This translates to having a certain resolution in phase for the template profiles and observations, depending on the bins used. Therefore, to get the most accurate template profiles possible, one should average over a large number of observed pulse profiles [56].

Therefore, we use the timing model to influence the creation of template profiles to be compared to observations of TOAs from pulsars. There are a certain number of parameters used when constructing the timing model to account for delays and noise that all radio pulses are affected by. These parameters include relativistic effects

from the solar system, switching templates when comparing to observations, atmospheric propagation, parallax, interstellar medium (ISM) propagation delays, pulsar binary parameters (if applicable), etc. Seeing as both the pulsar and Earth are moving with respect to each other, there is a delay induced from our relative motion. Therefore, we need to transfer the TOAs we receive from the pulsar into the frame of the Solar System barycenter (SSB), as well as the center of mass frame of the pulsar if it is in a binary or moving throughout the ISM [11]. There is also noise from the template itself, as the start and end to a phase in a template profile is arbitrary. Therefore, using different points of a phase for the "start" and "end" of the profile can cause mismatches when comparing data to templates.

There are many different methods used to create the mathematical models used for the timing models. In the TOAs, we observe a period of rotation P, as well as the derivative of the period $\dot{P}$, also known as spindown. The spindown needs to be split into the pulsar's intrinsic spindown, spindown due to the motion of the galaxy, and spindown due to the transverse velocity of the pulsar. Then, The timing models can fit for the spin parameters of the pulsars using basis polynomial functions [22] or a Taylor series expansion of P and $\dot{P}$. Furthermore, these models take into account constant noise, such as noise from the detector which can be tested beforehand. Various pulsar observatories today use Bayesian inference methods to get distributions of the noise parameters used in the timing model. NANOGrav [23], a pulsar timing array (see §2.2), used the software TEMPO2 [21] to fit their timing models simultaneously with their noise models, while the European Pulsar Timing Array (EPTA) use TEMPONEST [34].

After obtaining a suitably accurate timing model and template profiles, one can then predict the TOAs of a certain pulsar and compare it to the observed pulse profiles, to see certain delays and affects of noise. These TOAs are predicted by cross-correlating the pulse-profiles with a template profile [2]. Then, these observed noise patterns can be used to update the timing models and create more accurate template profiles. This makes the timing models become more accurate after successive observations, that making analyses of timing residuals more accurate and precise, as well as reducing the overall RMS timing residuals solely due to noise.

If the timing model perfectly modeled each contribution to the altering of the TOAs from the observation, the timing residuals would be consistently 0. Any deviation from 0 in the timing residuals highlights effects that weren't taken into account when producing the timing model. In the optimal case, the residuals will only consist of "white" noise, which is time-independent and therefore can be easily measured overtime, instead of "red noise", which is time-dependent, and can have complex affects on the timing residuals that vary overtime. We can however, try to include these effects of noise by finding their unique imprint on the timing residuals, called their "timing signature" and separate them based on their noise type (white or red).

2.1. *Types of Noise*

2.1.1. *White Noise*

White noise is the constant, time-independent noise that stems from effects that imbue a constant value on the timing residuals of a pulsar. This noise is called "white" as its behaves as a constant value on the power spectrum, because it is constant over all times. The white noise that affects timing residuals include the motion of the Earth and the pulsar relative to each other, noise from the detector (also known a radiometer noise), and error in fitting templates/pulse phase. There is also noise due to the fact that template profiles are averages of the pulse profiles of each pulsar. Not every point in the phase is constant across observations of each pulsar. A single pulse from a pulsar at some point in the phase can vary across each observation. Therefore, the pulse will seem to "jitter" around the average pulse profile used for the template.

These sources of noise are constant overtime, and therefore can be measured and placed into the timing model relatively simply. To correct for the motion of the Earth with respect to the pulsar, we transfer the observed TOAs into the frame of the SSB. However, to do this, we need to have three pulsar measurements, where the pulsars are widely separated across the sky [36]. The parameters of the detectors and telescopes can be used to account for the radiometer noise, and the mismatch of the templates can be used to account for template error. We also account for the pulsar's proper motion, spin kinematics, and orbital parameters if it is within a binary.

If white noise were the only source of noise in pulsar timing, it would be simple to extract parameters from timing residuals about a pulsar's spin parameters, kinematics, and binary orbit (if applicable). If these effects are relatively strong, then they can be measured without taking red noise into account, such as for measuring Shapiro delay and advance of periastron to get masses of a pulsar and its companion, or measure the effect of a planet orbiting a pulsar. However, more subtle residuals require an in-depth characterization of red noise.

2.1.2. *Red Noise*

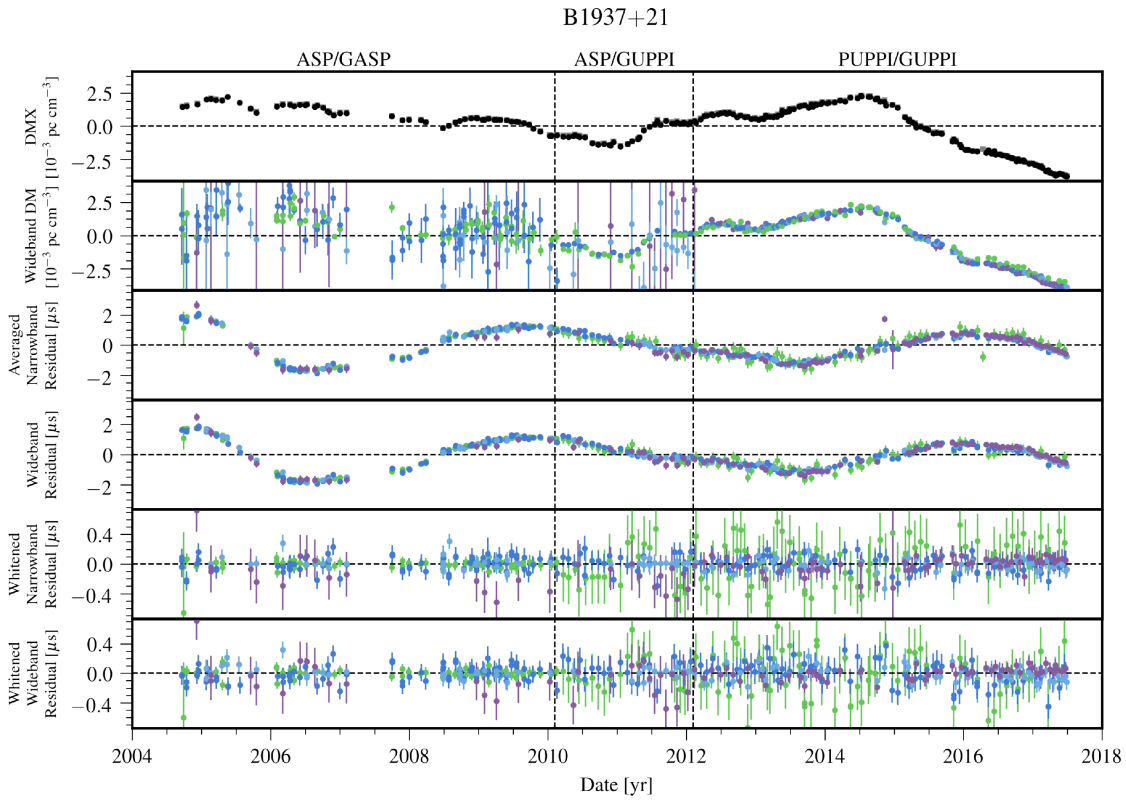

Figure 2. An example of timing residuals from NANOGrav , of the pulsar PSR1937+21. The top two plots display the change in DMX (a parameter used to describe DM) and DM overtime. The four plots below this show the timing residuals for this pulsar using various different analysis schemes. We can see that NANOGrav currently has timing residuals in its pulsars on the order of μs. The partitions in the plots by date correspond to using different instruments to observe the pulsars. (Alam et al. [2])

Red noise is a type of noise that is time-dependent, and therefore has a much more complex effect on timing residuals. This noise is called red, as it peaks at lower frequencies on a power spectrum. There are also many types of red noise that depend on the radio frequency used to observe the pulsar.

Many of the sources of red noise are intrinsic to the pulsar being observed, such as rotational instabilities , spin variations from torque fluctuations, internal NS variability[38], and spin noise [35]. In §2, we assumed that the pulse profiles were constant across time as well as frequency. However, if these parameters intrinsic to the pulsar can change with time, then the pulse profile can change with time. Furthermore, it has been observed that the pulse profile, as well as pulse jitter, differs depending on the frequency the pulsar is observed in [56, 38].

In addition to these sources of red noise intrinsic to the pulsar, there are also sources extrinsic to the pulsar, namely those affecting the radio pulses as they traverse along our LOS. The most significant source of red noise in timing residuals is known as dispersion measure (DM), and results from the change in path from material within the ISM. DM can depend both on the time it is observed, and the frequency it is observed in (see §3). Therefore, this source of noise complicates the timing residuals greatly. There are also other extrinsic sources of red noise, including scintillation and refraction due to structures within the ISM, that vary overtime as well.

In order to measure the effects of these frequency-dependent sources of noise, and accurately account for them, pulsars are observed in two widely separated radio frequency bands. This allows for analysis of DM and structures in the ISM in multiple frequencies, to see patterns that emerge across pulsars and across frequencies [3]. EPTA also measures the first and second derivatives of DM with respect to time in order to place even more constraints on the effect of noise. These derivatives can then be used to infer the magnetic field, age, and distance to the pulsar [11]. In general, DM can be used to infer distances to pulsar, making it useful to measure

DM not only to account for noise, but also to constrain properties of pulsars.

Red noise contributions in the timing residuals are overall the most impactful and cause the most hindrances to extracting parameters of pulsars. It has been shown that red noise in pulse profiles can "pull" a TOA to an earlier or later time as a stochastic contribution to noise. This contribution worsens when observing MSPs, because the data is folded much more frequently compared to canonical pulsars [26]. Seeing as the data is folded more often, and red noise affects each observed phase, the overall timing residuals will be more affected by red noise. One might think it would be easier to average over frequency when determining timing models in order to decrease data consumption, but this will severely hurt the frequency resolution needed to obtain DM and pulse profile variation [2].

2.2. *Pulsar Timing Arrays*

Seeing as there are astrophysical sources that could affect the propagation of emission from one pulsar, there may also be emission that affects emission from multiple pulsars at once. Furthermore, detecting sources of noise that affect multiple pulsars would be able to increase the overall signal-to-noise ratio (STN) of the detection of that noise or astrophysical process. This is what spurred the creation of pulsar timing arrays (PTAs). PTAs utilize the fact that these MSPs are very stable rotators to observe astrophysical source common to all pulsars observed.

Say that you measure two pulsars, and acquire TOAs for both of them. Both of them will have their own independent pulse profiles, and therefore template profiles to compare them to, based off of the timing model used. Each of these pulsars will then have their own timing residuals, differing by the parameters intrinsic to the pulsars, or the LOSs we have with either of them. After subtracting these effects from the residuals, we are left with data correlated between the two pulsars. This is the main concept behind utilizing PTAs. They can correlate timing residuals and signals within them between pulsars to increase the chance of detecting something of astrophysical origin [56, 35].

PTAs are also useful in that they have very high sensitivities to signals that correlate between the timing residuals of multiple pulsars.The sensitivity of PTAs depends on the number of pulsars within the array, the time span of the data observed by the array, the RMS of the timing residuals, and the number of observations for each set of pulsar data [57]. As the time span of the dataset increases, the more information there is to utilize in creating timing models for the observation of pulsars, thus decreasing the effect of noise in timing residuals, thus increasing the sensitivity to astrophysical signals. Increasing the number of pulsars also has a similar effect, in that it increases the number of template profiles that can be used [38, 56]. In general, a PTA needs to have at minimum 5 pulsars in order to achieve a certain level of sensitivity to observe relevant astrophysical signals, which in this case are gravitational waves (GWs) [16] (see §4).

However, there are many limits to the frequency range of the signals in which PTAs can observe. The lowest frequency a PTA dataset is sensitive to is set by the total observing time of the data set. This means that the PTA cannot detect a signal whose period is longer than the time span of the PTA dataset [35]. Furthermore, the sensitivity to lower frequencies is limited by the frequency of observation for each pulsar, which in the best-case scenario is once per week or every few weeks, or on the order of μHz [56]. Though, in general, PTAs are sensitive to frequencies on the order of $1/T$, where T is the length of the dataset. There is another limit on the extent to which PTAs can measure signals based on distances to the astrophysical source. Aggarwal et al. [1] state that PTAs are only sensitive to individual sources in the local Universe, where $(1 + z) \approx z$, where z is redshift.

Currently, there are various PTAs located around the world including the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), the European Pulsar Timing Array (EPTA) [15], and the Parkes Pulsar Timing Array (PPTA) [37]. NANOGrav utilizes the Green Bank Radio Telescope and the Arecibo Radio telescope to obtain its pulse profiles. EPTA uses the Lovell Telescope, Westerbork Synthesis Radio Telescope, Effelsburg Radio Telescope, and Nancay Radio Telescope to obtain its data. PPTA utilizes the Parkes Radio Telescope, located in Australia, to obtain its data. All three of these PTAs collaborate within the International Pulsar Timing Array (IPTA) [20], which utilizes data from all three organizations. Each PTA has a unique way of collecting and analyzing its data, as well as precisions in their measurements.

NANOGrav currently have 47 MSPs within their array, and expect to have $\sim$ 100 MSPs by 2025 [2]. They carry out observations every three weeks with both of their telescopes, with each pulsar being observed for approximately 10-40 min at two widely separated radio frequencies [38]. More recently, they have come up with an analysis scheme to increase the precision of their measurements, called wideband timing. This uses a combination of a frequency-dependent profile model with an augmented TOA measurement algorithm to produce two

measurements regardless of the frequency resolution of the data: one TOA measurement, and one DM measurement. They utilize 2048 bins in the pulse phase, which are time-averaged to have a resolution 2.5% of the phase. Currently, they have approximately 480000 profiles from measurements over the past 12 years.

The IPTA overall, has timing residuals on the order of 60 ns, and TOA uncertainties of 30 ns at best. Of 37 pulsars, 13 have TOA uncertainties less than 250 ns.

3. DISPERSION MEASURE

As stated before, dispersion measure (DM) is the most significant contributor of red noise to the timing residuals of pulsars. Other time-dependent "red" signals will be masked out by the influence of DM on the timing residuals, so it is imperative to characterize and account for DM in each pulsar observed. DM is defined as the integrated column density of free electrons along the LOS between the Earth and the pulsar [56, 2, 49]. As such, the DM is highly dependent upon the LOS that the radio pulse travels along on its way from the pulsar. Therefore, it is related to such dynamics as the ISM, the motion of the pulsar, the motion of the Solar System, etc. It is also imperative that we model DM and its variations in frequency and in time accurately, as modeling it incorrectly could lead to more noise in the timing residuals [49]. However, we cannot average the pulses over frequency to cancel out this effect, as it will smear out the pulse profile [33].

One of the main sources of DM is from the ISM, more specifically the ionized portions of the ISM. As the radio pulse travels through the ISM, it encounters ionized, cold plasma, which causes dispersion [26, 31, 56]. Additionally, radio pulses can be scattered due to inhomogeneities in the ISM. This scattering causes the path of the radio pulse to be broken up into multiple paths, delaying the radio pulse [26]. Furthermore, density variations in the ISM could cause electron density fluctuations in the ISM, altering the DM over time [31]. There may also be regions in the ISM that are more dense or less dense than their surrounding regions. These regions can then refract the radio pulses along our LOS, changing the electron density timeseries we measure [31, 7]. These density fluctuations impose phase perturbations on the radio pulses, which cause them to vary in DM and intensity (also known as scintillation). [30]. Donner et al. [13] showed that DM variations result from small scale ($\sim$ 1 AU) structures in the ISM. The ISM also has a certain degree of turbulence, which will cause the electron content, and therefore the DM, to change as a function of time. In general, ray path averaging of radio emission through different volumes of the turbulent ISM will smooth the DM timeseries, which broadens at lower frequencies [32].

There are also other contributors to the DM that do not involve the ISM. For example, solar winds can alter the electron content in our solar system, which would in turn alter the DM overtime [56, 50]. However, this can be avoided if observations close to the Sun are discarded. Furthermore, the electron content throughout the ionosphere varies depending on your LOS, location on the Earth, and your time of observation [31]. This means that all of these circumstances can change the DM you measure overtime. In addition, pulsars are also moving as we measure them, changing the LOS we view them from, and therefore the path that the radio pulses travel. This causes the DM to change over time as well [31, 56]. The pulsar may also alter the electron content in its local neighborhood through its motion through the ISM [31].

Therefore, in order to mitigate the effect of DM on timing residuals, we need to account for all of these different factors that influence the DM in terms of both frequency and time. In PTAs, this is done by measuring TOAs as a function of frequency, normally with two widely separated frequencies.

Normally, what is fit for in timing models is not the DM itself, but the variation from a constant DM level during a certain time period ΔDM. This constant DM chosen is arbitrary, however, and may introduce noise into the timing residuals if not constructed properly [49]. NANOGrav have a particularly unique way of fitting and accounting for DM in their timing residuals. Instead of using one constant value for a certain time range of observations, they use a piecewise-constant set of DM offsets, that vary per epoch. This parameter is called DMX and is calculated using prior DM measurements [49, 2].

NANOGrav also account for the frequency evolution of the pulse profile, by measuring Frequency-Dependent (FD) parameters, which are polynomial coefficients, and JUMP effects, which account for additional unmodeled profile evolution and effect of using two different frequencies to observe [49]. The number of FD parameters can differ for each MSP, but the requirement is that they must be covariant with DMX and JUMP. FD parameters also account for average scattering broadening over the course of a dataset.

However, seeing as scattering and DM are related, some scattering effects will be absorbed if only DM is fit for in analysis [26]. Furthermore, even if DM is measured at two different frequencies, the DM will change between observations, as it is time-dependent [10]. This

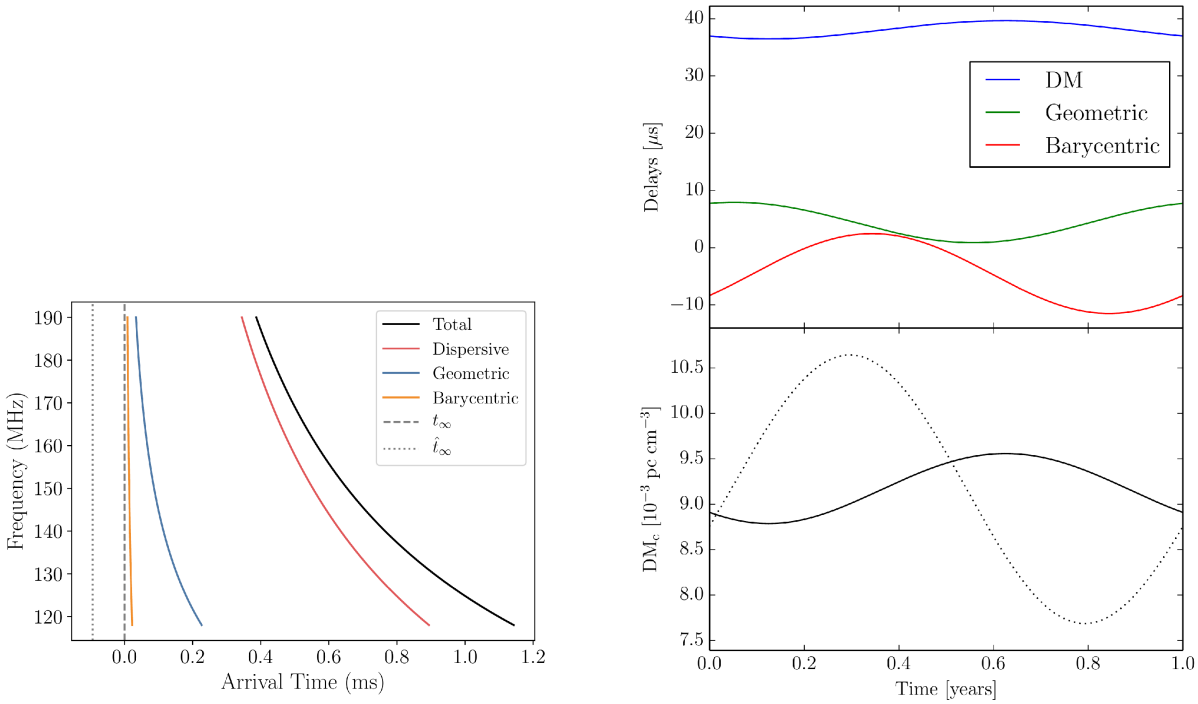

Figure 3. These plots show the variable nature of red noise as a function of both time and frequency. The dispersive delay is DM, the geometric delay is the scattering due to structures in the ISM, and the barycentric delay is due to motions in the solar system and pulsar changing the LOS. *Left:* The delay times due the change in certain red noise processes as a function of observing radio frequency. The dispersive delay shown in red is shown to have a higher amplitude and vary more rapidly with frequency than the other two. This shows that DM is the dominant red noise process. In this case t_∞ represents what delay would be caused if the frequency is infinite. $\hat{t}_\infty$ is simply the estimated value. [32] *Right:* The change in these various delays as a function of time over timescales of years. The solid black line represents the change in DM over time, while the dashed line is the change in all three effects overtime. [31] (Lam et al. [32, 31])

shows that accounting for DM in pulsar observations can be very complex.

There are many results from DM measurements that can explain how DM functions. It is well-known that time delays due to DM are $\propto \nu^{-2}$, where ν is the observing frequency [49, 26]. On the other hand, pulse scatter broadening from inhomogeneities in the ISM causes delays $\propto \nu^{-4}$, which is a property of Kolmogorov turbulence [49, 26]. There are also many forms that the DM can take when it has been measured. The DM can display linear trends dispersed between stochastic variations, piecewise linear trends associated with ISM structures on scales of 1-100 AU, periodic variations, or a power law [31, 13].

There are also many physical properties of the ISM we can infer from the measurement of DM. The DM has been used to infer distances to pulsars, assuming that the Galaxy has a free electron density [9]. The DM can give us information on Galactic models of electron density and variations in electron density on scales $\sim 1-100$ AU [31]. It has been observed that DM can vary on very short timescales from solar wind or structures in the ISM [32]. In general, measurement of how these DM time delays change overtime allow us to build models of the ISM, how it changes, and how it is structured.

DM has also been used to put bounds on precision for pulsar timing in general. Lam et al. [32] showed that a timing precision of $\sim$ 10 ns can be achieved. Lam et al. [30] showed that the same lower bound can be achieved if scintillation timescales were comparable to those of MSPs in PTAs. Overall, DM is a very complicated parameter, and causes the most hindrance in detecting astrophysical signals.

4. MEASURING GRAVITATIONAL WAVES

One of the main hallmarks of Einstein's theory of general relativity was the prediction of gravitational radiation. There are a variety of different ways one can release gravitational radiation, but the most frequent way in which they are produced is from a binary of compact objects. According to Einstein, when two masses are accelerated in an asymmetric fashion, they release gravitational radiation, analogous to electromagnetic radiation [39]. Normally, the gravitational radiation produced by

an object is too faint to be measured, especially over large distances. Therefore, we can only measure gravitational waves (GWs) from objects that have strong gravity and are accelerated very rapidly. Measuring the properties of these waves and how they are polarized can tell us information about the systems they come from. For example, the strength of GWs scales with the compactness of the system they are generated in [59]. The intensity of the strain GWs induce on the spacetime they travel through falls off as $\propto 1/d$, where d is the distance to the source of the GWs. Therefore, measuring the properties of these wave scan tell us about the systems they come from.

4.1. *GWs and PTAs*

Furthermore, GWs can affect the pulses that pulsars send along our LOS. A GW passing through the LOS we have with a pulsar will cause the path of the radio pulse to fluctuate, along with the spacetime it travels through. This strain of the spacetime will perturb the index of refraction for the electromagnetic radiation coming through space, causing its path to be altered [22]. Furthermore, the pulse arrival time is advanced, or retarded, in proportion to the number of cycles of the GW that the radio pulse passes through [22]. This will then cause a redshifting of the pulsar signal. This will imprint a specific quadrupolar signal distinct from other noise signals in the residuals [16]. We need to measure the affect of this delay and the redshift imparted by the GW in order to make a successful detection. There are characteristic ways in which the GW will affect the timing residuals at time t:

$$R(t) = -\int_0^t \frac{\delta\nu(t)}{\nu} dt \tag{1}$$

$$\frac{\delta\nu}{\nu} = H^{ij}(h^e_{ij} - h^p_{ij}) \tag{2}$$

where $R(t)$ is the residual, ν is the frequency of the radio pulse, $\delta\nu$ is the change in frequency, h^e_{ij} is the GW strain at the Earth, h^p_{ij} is the GW strain at the pulsar, and H^{ij} is a geometrical term that depends on the position of the GW source[44, 12, 19]. Furthermore, the fluctuation of the TOA of the pulse is proportional to the strain amplitude:

$$h_0^2\,\Omega_{GW}(f) < (4.8 \times 10^{-9})(f/f_*)^2 \tag{3}$$

where h_0 is the strain amplitude, Ω_{GW} is the fractional energy of GWs in the Universe related to the critical energy density, f is the frequency of the GW, and f_* is an arbitrary characteristic frequency (usually chosen to be 1 yr^{-1}) [36]. Additionally, it has been shown that GWs affect P, but not $\dot{P}$, granting another way of distinguishing the GW signal in timing residuals from other sources of noise [12].

4.2. *Measurable GWs*

Initially, it was only thought that pulsars could be used as point masses to search for passing GWs. GWs would pass by a system with a pulsar in it and be able to affect its timing residuals [16, 44]. However, the first piece of evidence for GWs came from a system of two NSs. One of the most common scenarios involving the release of gravitational radiation involves a binary of two compact objects. As these two bodies orbit closer and closer together, they will accelerate and release gravitational radiation, which will remove energy from the binary. This removal of energy will in turn change the orbital parameters of the system, including the period of the two companions within the binary. Taylor & Weisberg [51] observed a pulsar (PSR 1913+16) orbiting a NS, moving at a high velocity within the strong gravitational field of its companion. When analyzing the timing residuals, they could see a clear decrease of the orbital period of the pulsar, consistent with a loss of energy emitted through quadrupolar gravitational radiation. In addition to proving Einstein's theory, this ruled out many alternate theories of gravity that did not include GWs.

In addition to these GWs emitted from pulsars in binaries with other compact objects, PTAs can also observe GWs within the frequency-space in which they are most sensitive. The sensitivity of PTAs to GWs is limited by the time span of the observed dataset, and the cadence of observations performed, as stated in §2. This means that PTAs can extend GW observations to the very-low-frequency spectrum ($\sim$ 1-100 ns), within the nHz band [59, 3]. The absolute minimum frequency is set by the cadence of observations, which is near $f_{GW} \approx 2.8$ nHz. The maximum frequency is set by the length of the dataset, which is currently $f_{max} \approx 317.8$ Hz [1]. In terms of the characteristic strain of GWs, detectors in PTAs are sensitive to:

$$h_c(f) \sim \epsilon/T$$

for $f = 1/T$ where the frequency f is in the range 1-10 nHz, $f_* = 4.4$ nHz is the limiting frequency, and ϵ is a proportionality constant. [36].

However, detection of low-frequency GWs in PTAs requires a high cadence for observations and long pulsar datasets [2]. Furthermore, timing residuals need high precision, meaning sub-μs accuracy, which is challenging due to all of the forms of noise [26, 30]. However, from

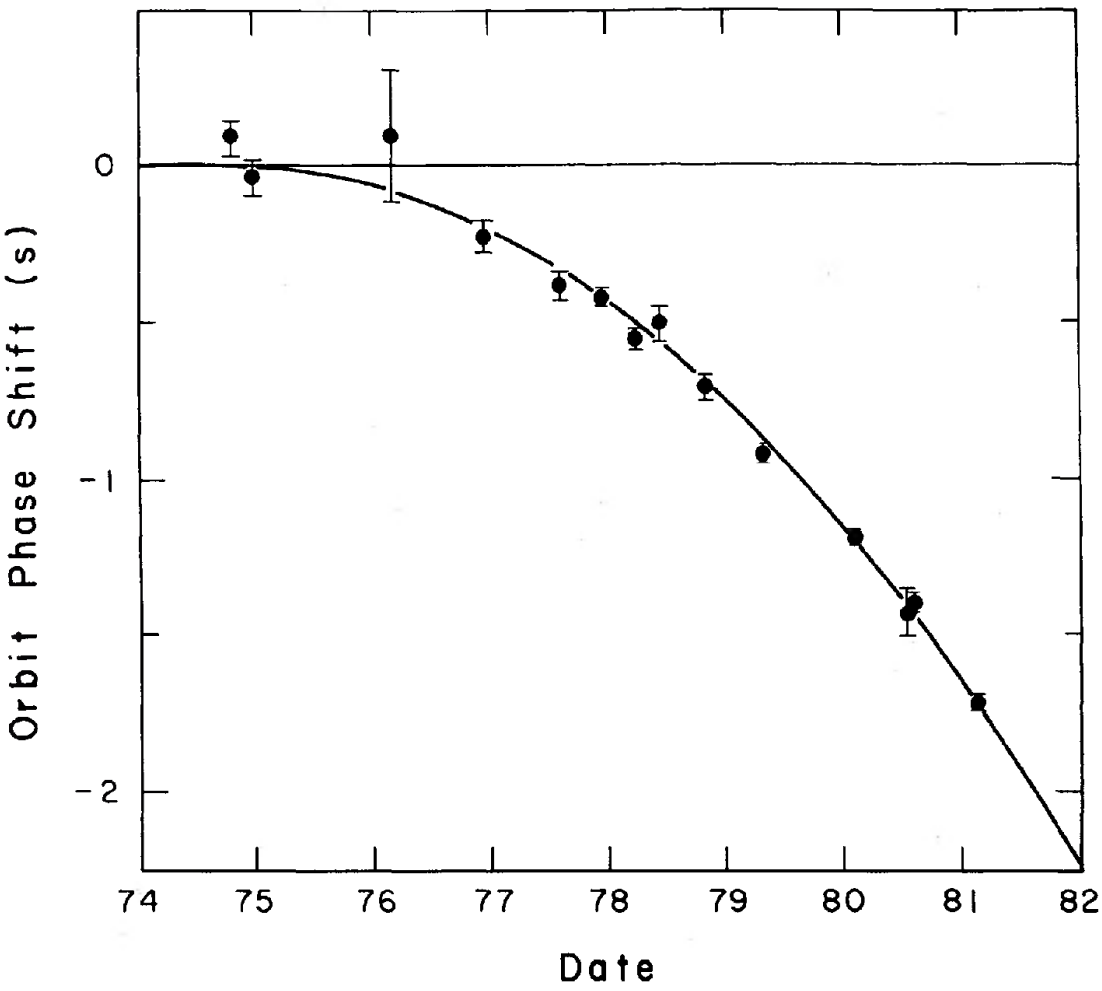

Figure 4. This is the first evidence of GWs detected by Taylor & Weisberg [51] in the observation of a pulsar-NS binary. The release of gravitational radiation caused the orbital period of the binary to decrease in a characteristic fashion, shown by the curve. (Taylor & Weisberg [51])

the estimates of DM in §3, this seems achievable. Furthermore, the STN of GWs are increased by the number of pulsars within the array, meaning the more pulsars, the higher the chance of detection is [35]. Though, it is unlikely that an individual GW source would be detectable with current sensitivities[19].

When discussing GWs detectable by PTAs, theorists will parameterize different aspects of the GWs that can be inferred from PTA measurements. When measuring GWs, one can infer the energy density that that have throughout the whole Universe. Usually, this is presented as Ω_{GW}, which is the fractional energy density of GWs compared to the critical energy density of the Universe [52]. Furthermore, GWs are parameterized by their characteristic strain h_c, which determines by how much they distort the spacetime they travel through. Thrane & Romano [52] parameterizes these variables in the following way:

$$\Omega_{GW} = \Omega_\beta (f/1\,yr^{-1})^\beta$$
$$h_c = A_\alpha (f/1\,yr^{-1})^\alpha$$

where α and β are spectral indices, and A_α and Ω_β are amplitudes corresponding to these indices.

4.3. *Detecting GWs*

GWs will be detected in timing residuals in the same manner as all other forms of noise and delays. GWs will appear as red noise, as they are time dependent, and they will appear in multiple pulsars' TOAs. However, GWs are not accounted for in the timing model, and will therefore provide residuals, even though these residuals will be < 100 ns [19]. GWs also cause a certain type of quadrupolar correlation between the timing residuals of different pulsars, which distinguishes them from other

residuals [16]. Hellings & Downs [18] created a form for the correlation of GW timing residuals between different pulsars within a PTA. They found that they can find traces of GW residuals in fitting for the spin down of a star. Furthermore, they found that the correlations between pulsars from GW fluctuations is a function of the angular separation of pulsars on the sky. This "Hellings-Downs curve" a very useful tool when trying to extract timing residuals of GWs from observations.

4.4. *Types of GWs*

There are a variety of types of GWs that can be detected by PTAs, within their sensitivity range, along with the variety of sources that can generate them. Firstly, PTAs are sensitive to an entirely different frequency range than other GW detectors, such as LIGO [8], meaning that they can measure their own types of GW sources. PTAs can detect the early inspiral of compact binaries, as opposed to LIGO which can only measure the end of compact binary inspirals, and their eventual merger. Because PTAs measure low-frequency GWs, they measure the less energetic events within compact object binaries. Because of this, PTAs do not need to use sophisticated GW waveforms that are generated during coalescence of compact objects like LIGO [59].

There are also other types of GWs that PTAs can observe. The most expected and anticipated type of GW that PTAs expect to observe are those from the stochastic gravitational wave background (GWB). This GWB is not a single GW, but a superposition of a large number of GWs, from a large number of sources, that add incoherently [56]. This GWB most likely comes from binaries of supermassive black holes (SMBHBs) that are created after galaxy mergers, but can also stem from cosmic strings, quantum fluctuations from the early Universe that were amplified by inflation, dark matter, and many other sources [3]. In general, however, the stochastic GWB from SMBHBs is the most likely detectable signal by PTAs [57].

There are also other types of GWs, called continuous gravitational waves (CWs) that emit a roughly constant GW frequency for an extended period of time. These sources are also most likely SMBHBs, which are early on in their inspirals and are continuously emitting low-frequency GWs [38]. The final type of GWs are called burst sources, which have signal durations much shorter than the other GWs. In fact, burst sources have durations much shorter than can be observed currently. These GWs could be due to mergers of SMBHBs, cusps of cosmic strings, and a variety of other sources [36].

5. GRAVITATIONAL WAVE BACKGROUND

We have seen in §4.4, that the main types of GWs we are expecting to observe from PTAs are those that stem from the gravitational wave background, which come from sources such as SMBHBs, cosmic strings, and relic GWs from the early Universe. Now, we wish to see how these GWBs can be observed, what the distribution of GWs will look like, and what limits/constraints we can set from them.

As with normal GWs, the PTAs are most sensitive to the GWB in the range of 1-100 nHz. In order to extract these signals from timing residuals, the TOAs from each pulsar will need to be correlated with each other, and assumingly will follow the Hellings-Downs curve. We can then use this Hellings-Downs curve to see if a possible detection is credible and characterize the GWB. The GWB will also need to display a quadrupolar nature in its timing residuals, as with normal GWs. The significance of a GWB signal can be increased by increasing the number of pulsars observed, decreasing the RMS timing noise, increasing the number of observations, and "whitening" the power spectrum of the residuals, so that no red noise interferes with the GWB signal [24].

Like normal GWs, the GWB will generate a red power spectrum, with excess power at lower frequencies, as most of the GWs forming the stochastic GWB are low-frequency [25]. We have seen that GWs are parameterized by a power-law spectrum, in the form used by Thrane & Romano [52]. However, assuming different sources for the GWB changes the spectral indices α and β. For relic GWs, Ω_β = constant, $\beta = 0$, and $\alpha = -1$. For SMBHBs, $\beta = 2/3$ and $\alpha = -2/3$. Therefore, for the most likely source of the GWB, SMBHBs, the characteristic strain of the GWB falls off with a spectral index of $\alpha = -2/3$ [22, 52]. It has also been shown that the power spectrum of the GWB acts as a power law, then flattens or tips over at frequencies less than 1 yr^{-1} [56]. Additionally, the stochastic GWB will excite both polarizations of GWs, seeing as the GWB is an incoherent superposition of a large number of GWs, each being a linear combination of both GW polarizations [22, 46].

The measurements of the GWB will allow us to place constraints on the sources we assume to generate that GWB. If we assume that cosmic strings are generating the GWB, then we can place constraints on the mechanism within cosmic strings that generate GWs. In this case, the GWB would be able to constrain the tension of the cosmic strings, given as $\frac{G\mu}{c^2}$ [35, 43]. If we assume that relic GWs produce the GWB, then we would be able to place constraints on the relic GW strain and on properties of inflation [25]. There are also a large number of parameters that can be constrained if we assume SMBHBs produce the GWB. We would be able

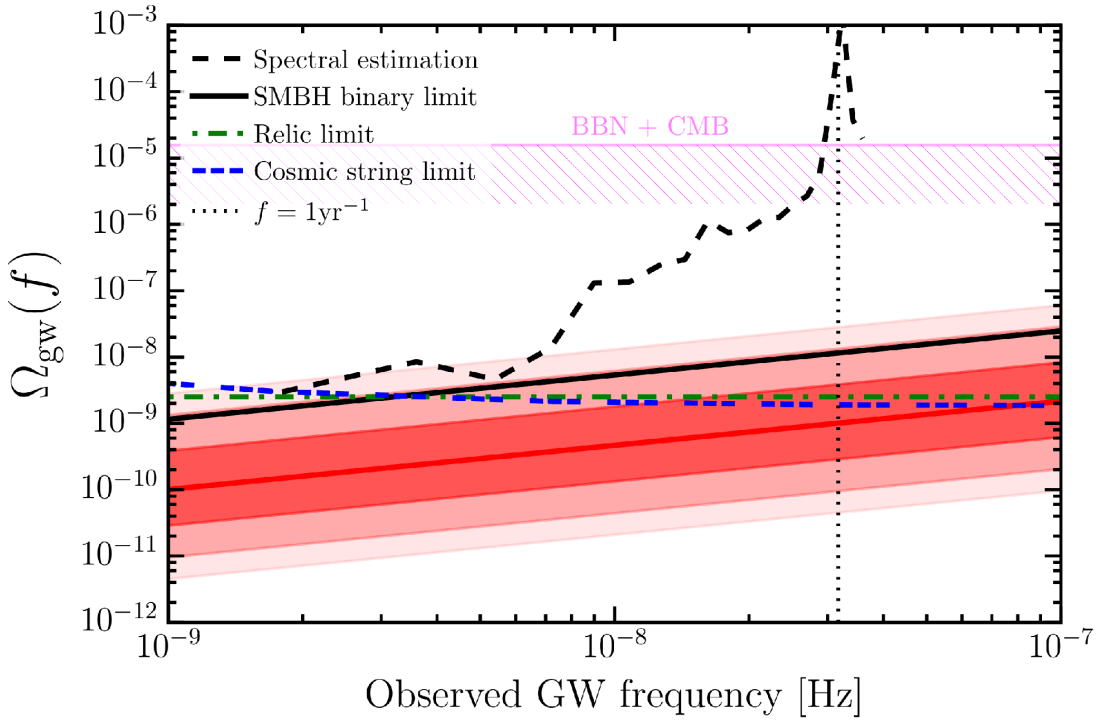

Figure 5. Limits placed on the fractional energy density due to gravitational waves Ω_{GW} assuming various sources for the GWB. These are all based on observational limits from the EPTA. Out of all of the sources for GWs, cosmic strings, SMBHBs, and relic GWs are those that are theorized to be detected by PTAs. The pink region represents the limits placed on the GWB using arguments from nucleosynthesis and measurements of the CMB. The red shaded areas represent the central 68%, 95%, and 99.7% confidence interval of the GWB amplitude according to Sesana [45] assuming that a SMBHB evolves purely due to decay of GWs and binaries are circular. The dashed black line represents the estimation of GWB from EPTA measurements for each frequency, not only utilizing the normal 1 yr^{-1} reference frequency. The limit for the GWB assuming relic GW and cosmic string sources are the green and blue dashed lines. (Lentati et al. [35])

to constrain certain mass relationships known to exist withing galaxies, such as the galaxy-halo mass relationship, the evolution of galaxies and supermassive black holes (SMBHs), the characteristics of the population of SMBHs, the merger rates of galaxies and SMBHs, etc. [25] We would also be able to constrain theories of dark matter, such as fuzzy dark matter [27] and galactic sub-halos [41].

So far, no GW or GWB has been directly observed by PTAs. Data from NANOGrav show a strong, common red noise process consistent with a GWB, but it lacks the quadrupolar corrections necessary to claim a detection [53]. However, this lack of an observation allows us to put limits on the characteristics of the GWB, including its strain and functional form. An upper limit to the GWB can be found by attributing all of the excess noise in timing residuals to GWs [18]. For all sources of the GWB, we find a limit of:

$$t_{emit,GW} = (6.6 \times 10^{-21}) \frac{1}{\epsilon^2} \left(\frac{1\mathrm{Hz}}{f_0}\right)^2 \left(\frac{100}{g_*}\right)^{1/6} \mathrm{sec}$$

where g is the internal degrees of freedom at t_{emit}, ϵ is a proportionality constant between wavelength and the Hubble parameter at t_{emit}, and f_0 is the measured GW frequency [36].

There have been limits imposed on the amplitude of the GWB using nucleosynthesis arguments and the comic microwave background radiation (CMB) [35, 36]. Assuming the GWB is produced by SMBHBs, an observationally-based upper limit on the amplitude is $A_{GWB} = 1.45 \times 10^{-15}$[3]. It was also shown using SMBHB population models that the stochastic GWB is mainly from SMBHBs of redshift $z < 1$ [3].

6. SUPERMASSIVE BLACK HOLE BINARIES

Seeing as SMBHBs are the most likely source of the GWB, which is the most likely source of GWBs to be observed by PTAs, it is fruitful to discuss the problems that could be answered in detecting the GWB. The creation of a GWB by SMBHBs would depend highly on many of the parameters of the SMBHB population, which is governed by many different dynamics of BHs

and galaxies across cosmic time. We will discuss motivations for studying SMBHBs, how they form, how the GWB is formed from SMBHBs, and GWB results gotten from assumptions of SMBHBs.

6.1. *Why SMBHBs?*

It is well known that SMBHs are ubiquitous in galaxies in the local Universe (low z) [58]. We observe them in the centers of almost all galaxies, and if the galaxies are far enough (high z), we observe them as active galactic nuclei (AGN). However, one outstanding problem in astronomy is the formation of these SMBHs, how they evolve over cosmic time, and what their properties are. These BHs are much more massive than those created from CCSNe as a consequence of stellar evolution. Furthermore, there is a disconnect between the population of SMBHs at high and low z, as the abundance of SMBHs in galaxies at large z is not known [58]. There is evidence that present-day SMBHs are dormant remnants of AGN that are present at large z, thus signaling that there is a large population of SMBHs at large z [58, 48]. In fact, there is evidence that SMBHs might be as ubiquitous in high z galaxies as they are today.

Through the use of reverberation mapping techniques, we are able to constrain masses of AGN and therefore get estimates of the SMBH population in general. We can also get these masses through relations such as the $M_{BH} - \sigma_*$ relation, which is related to the velocity dispersion of the galaxy, or the $M_{BH} - M_{bulge}$ relation, which is related to the mass of the bulge of the galaxy. Studying properties of SMBHs could further constrain these relationships and methods used to obtain parameters from the SMBH population. SMBHs also affect the properties of their host galaxies, including their evolution and growth [48]. Thus, studying SMBHs can also lead to constraints on galactic evolution, feedback of AGN, contribution of AGN to SMBH growth, etc.

Furthermore, it is generally accepted that galaxy mergers are responsible for building up the masses of today's galaxies. In this framework, over cosmic time, smaller galaxies merged to form larger ones, who then merged to form even larger ones, etc. This process of hierarchical merging eventually built up the massive galaxies we see today. Seeing as many galaxies at large z are expected to harbor SMBHs, when these galaxies merge, their SMBHs will interact. In most cases, this will form a SMBHB in the center of the merger remnant. Therefore, studying SMBHBs can place constraints on the merger rates of galaxies over cosmic time [22]. Furthermore, studying SMBHB mergers allows us to study galaxy mergers, since the two are so interconnected [3].

6.2. *Formation & Dynamics of SMBHBs*

After the merger of a galaxy, there will be two SMBHs present in the merger remnant. These two SMBHs will eventually approach each other through the process of dynamical friction. In this process, the SMBHs will exchange angular momentum with stars within the merger remnant, slowly bringing them closer to each other in the center. After a time of $\sim 10^8$ years, the two SMBHs will sink to the center if the remnant and become bound to each other [4]. Seeing today's galaxies were built up from hierarchical mergers of smaller galaxies over cosmic time, this means that the Universe should have a large number of inspiraling SMBHBs.

This binary will continue to shrink in orbital separation (or "harden") due to dynamical friction with stars in the center of the galaxy, similar to how the two SMBHBs independently sank to the center of the merger remnant. However, the SMBHB will eventually run out of stars to utilize for dynamical friction in order to harden further. Begelman et al. [4] describes this as the binary sweeping out a "loss cone" of stars near the binary, which, after depleted, cannot allow the exchange of angular momentum to harden. This can cause the merger time of the SMBHB to be larger then a Hubble time, meaning we will not see it merge. However, if by some process the SMBHB exchanges enough energy with an outside object and hardens to a separation $\lesssim .02$ pc, it will begin to release GWs and decay through that process (see §6.3). Eventually, the release of energy through GWs will allow the SMBHB to merge [22, 4]. These GWs will be the most luminous in the Universe, and will only take $\sim 10^4$ years to complete, which is short on cosmic timescales. Therefore, studying GWs can probe the dynamics and formation of SMBHBs.

6.3. *Solutions to the "Final Parsec Problem"*

One of the main issues discussed in the dynamics of SMBHBs is this "final parsec problem", which stems from the lack of stars for the SMBHB to exchange angular momentum with to harden sufficiently, so that it can enter a regime where it releases GWs. Without this angular momentum exchange from stars, the SMBHB cannot radiate GWs, and will therefore not merge within a Hubble time. Thus, if the binary does not merge within a Hubble time, we cannot observe its GWs. However, there are no observed SMBHBs in general, let alone SMBHBs at sub-pc separations [40]. There are various different theories to explain how this final-pc problem could be solved.

The first theory involves gas surrounding the two SMBHBs as they merge. As the second SMBH approaches the primary SMBH, which is assumed to al-

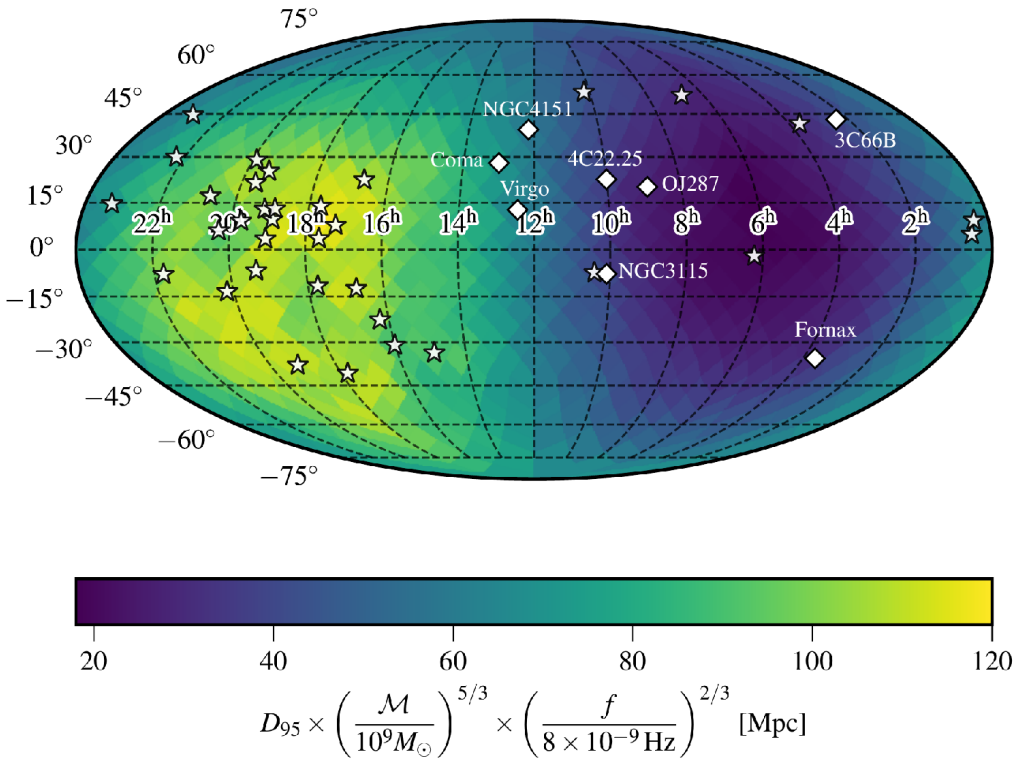

Figure 6. These are 95% lower limits calculated for the luminosity distance to SMBHBs as a function of the sky using data from NANOGrav and assuming that the GWB is from SMBHB mergers. This assumes the SMBHBs have a chirp mass $\mathcal{M} = 10^9 M_\odot$ and $f_{GW} = 8$nHz. The stars indicate pulsars that NANOGrav uses in their array, and the diamonds indicate known observed SMBHB candidates. (Aggarwal et al. [1])

ready be in the center of the galaxy, it opens up a gap in its accretion disk. This disk of gas forms around all BHs, as they are in an environment with large amounts of gas. The secondary BH will then exert a torque on the accretion disk from the primary SMBH, which causes an exchange of angular momentum. This then causes the SMBHB to harden, and merge on a timescale $\sim 10^5$ years.

Another theory involves three-body interactions between the SMBHB and a third SMBH. Seeing as galaxy mergers happen very frequently, it would not be uncommon if two galaxies merged and produced an SMBHB, and then the remnant merged with a third galaxy, letting the SMBHB interact with another SMBH. If we assume that the merger time of an SMBHB is not much less than the timescales between mergers, then this situation is valid. Eventually, interactions between the third SMBH could cause more stars to approach the binary, which would allow the SMBHB to exchange angular momentum and harden. Furthermore, the third SMBH could increase the eccentricity of the SMBHB. It is well-known that GWs will be produced quicker and more readily if the eccentricity of a compact object binary is large. Therefore, increasing the eccentricity of an SMBHB could allow for faster merger times. In general, this occurs through the Kozai mechanism, which induces oscillations in the binary's eccentricity, trading off of the inclination of the third object with respect to the binary plane [5, 29].

6.4. *GWB*

Now that we have found that SMBHBs can merge given certain circumstances, we can analyze the GWs they produce and the GWB that will be formed from the incoherent superposition of their mergers. SMBHB systems with masses $10^{4-9} M_\odot$ are the primary candidates of GW sources in the nHz - μHz frequency range [22]. Furthermore, we expect a large number of SMBHBs emitting in this frequency range [58].

It has been shown that for a GWB produced by SMBHBs, a wide variety of spectral indices α are possible, but that $\alpha = -2/3$ is a practical value to use [57]. Therefore, almost all analyses of the SMBHB GWB use this value for the spectral index. The average strain amplitude of the GWB is shown to have a dependence on frequency of $\langle h_c \rangle \sim f^{4/3}$, due to the fact that is a rapid

decrease in the number density of SMBHB mergers with frequency relative to the spectrum amplitude decrease [22]. In general, the power spectrum for GWB strain flattens at lower frequencies because not many very-low-frequency SMBHBs will coalesce within a Hubble time, due to the final-pc problem, and from interactions between the SMBHBs and their environments. At higher frequencies, α departs from -2/3 and steepens due to loss of power at high frequencies associated with the maximum frequency for inspiral. Furthermore the rapid hardening of binaries can cause the spectrum to steepen as well at higher frequencies [58, 40].

Seeing as this GWB is an incoherent superposition of a large number of SMBHBs, it should be isotropic. However, certain deviations from this isotropy in the GWB could make certain individual SMBHB systems resolvable [35]. These individual, resolvable systems will be shown as sharp peaks in the GWB above the average GWB contribution. Furthermore, these individual events would have to be very luminous, and therefore very massive [46].

There have been a wide variety of constraints placed on the GWB, PTA detectors, galaxy parameters, SMBHB populations, and many more quantities from the assumption that the GWB is formed from the incoherent superposition of GWs from SMBHB mergers. Lentati et al. [35] found limits on the number density of SMBHB mergers per redshift from the current precision of PTAs. They also found that the dominant contribution to the GWB from SMBHBs is from massive, low z systems, which is to be expected because GWs scale as $1/d$. Wyithe & Loeb [58] found that the GWB is dominated by sources of $z < 2$, which increase the validity of this finding. Sesana [45] found that the amplitude of GWs from SMBHBs depends on the rate of SMBHB mergers throughout cosmic time as well as their masses. They also found that the range of GW amplitudes theorized is consistent with observationally based estimates of SMBH properties. Aggarwal et al. [1] were able to place constraints on the chirp mass $\mathcal{M}$, a specific parameter involving the masses of both SMBHs in the binary, and the luminosity distance for SMBHBs in general, as well as depending on the location in the sky. This granted them values of $D_L > 120$ Mpc for $\mathcal{M} = 10^9 M_\odot$ and $D_L > 5.5$ Gpc for $\mathcal{M} = 10^{10} M_\odot$. Sesana et al. [46] found that the brightest GW signals come from SMBHBs with $\mathcal{M} > 5 \times 10^8 M_\odot$ at higher z.

7. FUTURE IMPLICATIONS

Overall, there is a very bright future ahead for PTAs, pulsar timing in general, and the measurement of GWs and the GWB. We have already been able to make significant constraints on the effects of GWs and the sources of GWs simply from not making a detection, Given that a lack of a detection could do so much, a credible detection would be able to enhance our knowledge of the field by a significant amount.

Furthermore, many analyses have been done to infer how we would go about detecting GWs and when we should expect to detect GWs with PTAs, given current sensitivities and such. The main piece of data used for inferring when we will be able to measure GWs with PTAs was given by Jenet et al. [24]: observing 40 pulsars 250 times over 5 years with a precision of 100 ns will allow for a significant detection of the GWB. We already have the number of pulsars and the observing time, so we simply need to increase the precision of the PTAs to be able to measure the GWB. In §3, we saw that precisions of ~ 50 ns were possible, so this means that a detection of the GWB is not very far off in the future. Kelley et al. [28] state that all PTAs could hope to detect GWs in a few years and Rosado et al. [42] state that the GWB would be detected in the next 1-2 decades. McLaughlin [38] state that the detection of GWs will occur by 2023.

Pol et al. [40] show that we will first measure a GWB, then as sensitivity increases we will be able to determine the source of the GWB, then as sensitivity increases we will be able to get properties of the GWB. Similarly, Arzoumanian et al. [3] state that at first, GWs will seem to be red noise with a spectral index consistent with SMBHBs, and then we will see the Hellings-Downs curve come into effect when measuring the GWB as sensitivities increase. Consequently, from these predictions, when we eventually detect a GW from a PTA, we will know how to infer its detection and extract its properties.

There are also many predictions specifically related to SMBHBs if a GWB was detected. Sesana et al. [46] state that the amplitude of a GWB created by SMBHBs in distant ($z \sim 2$) galaxies could be just below observational limits. Pol et al. [40] state that almost immediately after the initial detection of the GWB, we would be able to clearly distinguish between SMBH population models. Furthermore, they state that we could double the STN of the strain amplitude of the GWB in about 5-10 years. We would be able to compare these models predicted from GWBs to models predicted from observed SMBHB candidates. For example, Valtonen et al. [54] show that the AGN OJ287, which is supposedly an SMBHB, is in the final stage of inspiral before merger.

Table 1. Limits on the GWB from SMBHBs

Parameter	Value	Ref.
h_0	$< 7.3 \times 10^{-15}$	Aggarwal et al. [1]
A	$< 1 \times 10^{-15}$	Shannon et al. [48]
Ω_{GW}	$< 2.3 \times 10^{-10}$	Shannon et al. [48]
A_{avg}	2.4×10^{-15}	Shannon et al. [47]
A	$< 7 \times 10^{-15}$	McLaughlin [38]

In general, we are quickly approaching a sensitivity and length of observations where the GWB will be detectable to PTAs. In order to increase the sensitivity of PTAs, we need to be able to increase the number of pulsars observed, decrease the RMS of the timing residuals, increase the length of the dataset, and accurately model red noise. Verbiest et al. [57] show that a residual of 100 ns over 5 years is possible, relating to Jenet et al. [24]'s assertion, and that there is an upper limit of ~ 80 ns for the residual RMS of intrinsic instabilities within a pulsar. Therefore, to get past this limit due to red noise, we need to add more pulsars to the PTAs and keep observing them for long periods of time. In fact, there are multiple PTA and radio telescope projects that will start in the near future that could increase the sensitivity of these measurements: the Australian Square Kilometer Array Pathfinder (ASKAP) is a new radio telescope built in Australia, MeerKAT is a new radio telescope being built in South Africa, and the Square Kilometer Array (SKA) is a a new PTA being built in China [14].

All of these new additions to PTAs will allow their precisions to get better overtime, and also allow for the discovery of many new MSPs to add to the arrays. As time goes on, more data is gotten from all of these instruments, the timing models of MSPs will get better, and the sensitivities to GWs and other timing delays will be much better than they are now. The near future is expected to be filled with new discoveries ranging from the mergers of SMBHs to the turbulence of the ISM, simply from timing pulsars.

REFERENCES

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

[2]Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2020, ApJS, 252, 5, doi: 10.3847/1538-4365/abc6a1

[3]Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018, arXiv:1801.02617 [astro-ph, physics:gr-qc], doi: 10.3847/1538-4357/aabd3b

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

[5]Blaes, O., Lee, M. H., & Socrates, A. 2002, ApJ, 578, 775, doi: 10.1086/342655

[6]Camilo, F., Ransom, S. M., Halpern, J. P., & Reynolds, J. 2007, ApJ, 666, L93, doi: 10.1086/521826

[7]Clegg, A. W., Fey, A. L., & Lazio, T. J. W. 1998, ApJ, 496, 253, doi: 10.1086/305344

[8]Collaboration, T. L. S. 2015, Class. Quantum Grav., 32, 074001, doi: 10.1088/0264-9381/32/7/074001

[9]Cordes, J. M., & Lazio, T. J. W. 2003, arXiv:astro-ph/0207156. http://arxiv.org/abs/astro-ph/0207156

[10]Cordes, J. M., Shannon, R. M., & Stinebring, D. R. 2016, ApJ, 817, 16, doi: 10.3847/0004-637X/817/1/16

[11]Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, Mon. Not. R. Astron. Soc., 458, 3341, doi: 10.1093/mnras/stw483

[12]Detweiler, S. 1979, The Astrophysical Journal, 234, 1100, doi: 10.1086/157593

[13]Donner, J. Y., Verbiest, J. P. W., Tiburzi, C., et al. 2019, A&A, 624, A22, doi: 10.1051/0004-6361/201834059

[14]Feng, Y., Li, D., Zheng, Z., & Tsai, C.-W. 2020, Phys. Rev. D, 102, 023014, doi: 10.1103/PhysRevD.102.023014

[15]Ferdman, R. D., van Haasteren, R., Bassa, C. G., et al. 2010, Class. Quantum Grav., 27, 084014, doi: 10.1088/0264-9381/27/8/084014

[16]Foster, R. S., & Backer, D. C. 1990, The Astrophysical Journal, 361, 300, doi: 10.1086/169195

[17]Guillot, S., Kerr, M., Ray, P. S., et al. 2019, ApJL, 887, L27, doi: 10.3847/2041-8213/ab511b

[18]Hellings, R. W., & Downs, G. S. 1983, The Astrophysical Journal Letters, 265, L39, doi: 10.1086/183954

[19]Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2009, arXiv:0911.5206 [astro-ph], doi: 10.1088/0264-9381/27/8/084013

[20]—. 2010, Class. Quantum Grav., 27, 084013, doi: 10.1088/0264-9381/27/8/084013

[21]Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, Monthly Notices of the Royal Astronomical Society, 369, 655, doi: 10.1111/j.1365-2966.2006.10302.x

[22]Jaffe, A. H., & Backer, D. C. 2003, arXiv:astro-ph/0210148, doi: 10.1086/345443

[23]Jenet, F., Finn, L. S., Lazio, J., et al. 2009, arXiv:0909.1058 [astro-ph]. http://arxiv.org/abs/0909.1058

[24]Jenet, F. A., Hobbs, G. B., Lee, K. J., & Manchester, R. N. 2005, ApJ, 625, L123, doi: 10.1086/431220

[25]Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, ApJ, 653, 1571, doi: 10.1086/508702

[26]Jones, M. L., McLaughlin, M. A., Roy, J., et al. 2021, arXiv:2009.08409 [astro-ph]. http://arxiv.org/abs/2009.08409

[27]Kato, R., & Soda, J. 2020, J. Cosmol. Astropart. Phys., 2020, 036, doi: 10.1088/1475-7516/2020/09/036

[28]Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2017, Monthly Notices of the Royal Astronomical Society, 471, 4508, doi: 10.1093/mnras/stx1638

[29]Kozai, Y. 1962, The Astronomical Journal, 67, 591, doi: 10.1086/108790

[30]Lam, M. T., Cordes, J. M., Chatterjee, S., & Dolch, T. 2015, arXiv:1411.1764 [astro-ph], doi: 10.1088/0004-637X/801/2/130

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

[32]Lam, M. T., Lazio, T. J. W., Dolch, T., et al. 2020, ApJ, 892, 89, doi: 10.3847/1538-4357/ab7b6b

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

[34]Lentati, L., Alexander, P., Hobson, M. P., et al. 2014, Monthly Notices of the Royal Astronomical Society, 437, 3004, doi: 10.1093/mnras/stt2122

[35]Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, Mon. Not. R. Astron. Soc., 453, 2577, doi: 10.1093/mnras/stv1538

[36]Maggiore, M. 2000, Physics Reports, 331, 283, doi: 10.1016/S0370-1573(99)00102-7

[37]Manchester, R. N. 2008, AIP Conference Proceedings, 983, 584, doi: 10.1063/1.2900303

[38]McLaughlin, M. 2013, Class. Quantum Grav., 30, 224008, doi: 10.1088/0264-9381/30/22/224008

[39]Pitkin, M., Reid, S., Rowan, S., & Hough, J. 2011, Living Rev. Relativ., 14, 5, doi: 10.12942/lrr-2011-5

[40]Pol, N. S., Taylor, S. R., Kelley, L. Z., et al. 2021, arXiv:2010.11950 [astro-ph, physics:gr-qc]. http://arxiv.org/abs/2010.11950

[41]Ramani, H., Trickle, T., & Zurek, K. M. 2020, J. Cosmol. Astropart. Phys., 2020, 033, doi: 10.1088/1475-7516/2020/12/033

[42]Rosado, P. A., Sesana, A., & Gair, J. 2015, Monthly Notices of the Royal Astronomical Society, 451, 2417, doi: 10.1093/mnras/stv1098

[43]Sanidas, S. A., Battye, R. A., & Stappers, B. W. 2012, Phys. Rev. D, 85, 122003, doi: 10.1103/PhysRevD.85.122003

[44]Sazhin, M. V. 1978, Soviet Astronomy, 22, 36. http://adsabs.harvard.edu/abs/1978SvA....22...36S

[45]Sesana, A. 2012, arXiv:1211.5375 [astro-ph, physics:gr-qc], doi: 10.1093/mnrasl/slt034

[46]Sesana, A., Vecchio, A., & Volonteri, M. 2009, Monthly Notices of the Royal Astronomical Society, 394, 2255, doi: 10.1111/j.1365-2966.2009.14499.x

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

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

[49]Shapiro-Albert, B. J., Hazboun, J. S., McLaughlin, M. A., & Lam, M. T. 2021, arXiv:2010.07301 [astro-ph], doi: 10.3847/1538-4357/abdc29

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

[51]Taylor, J. H., & Weisberg, J. M. 1982, The Astrophysical Journal, 253, 908, doi: 10.1086/159690

[52]Thrane, E., & Romano, J. D. 2013, Phys. Rev. D, 88, 124032, doi: 10.1103/PhysRevD.88.124032

[53]Turner, J. E., McLaughlin, M. A., Cordes, J. M., et al. 2020, arXiv:2012.09884 [astro-ph]. http://arxiv.org/abs/2012.09884

[54]Valtonen, M. J., Mikkola, S., Merritt, D., et al. 2010, ApJ, 709, 725, doi: 10.1088/0004-637X/709/2/725

[55]van Straten, W. 2006, ApJ, 642, 1004, doi: 10.1086/501001

[56]Verbiest, J. P. W., Oslowski, S., & Burke-Spolaor, S. 2021, arXiv:2101.10081 [astro-ph]. http://arxiv.org/abs/2101.10081

[57]Verbiest, J. P. W., Bailes, M., Coles, W. A., et al. 2009, Monthly Notices of the Royal Astronomical Society, 400, 951, doi: 10.1111/j.1365-2966.2009.15508.x

[58]Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 590, 691, doi: 10.1086/375187

[59]Yunes, N., & Siemens, X. 2013, arXiv:1304.3473 [astro-ph, physics:gr-qc, physics:hep-ph, physics:hep-th], doi: 10.12942/lrr-2013-9