10-20-2014

Gravitational waves from individual supermassive black hole binaries in circular orbits: Limits from the north american nanohertz observatory for gravitational waves

Z. Arzoumanian

A. Brazier

S. Burke-Spolaor

S. J. Chamberlin

S. Chatterjee

See next page for additional authors

Recommended Citation

Z. Arzoumanian, et. al., (2014) Gravitational waves from individual supermassive black hole binaries in circular orbits: Limits from the north american nanohertz observatory for gravitational waves.Astrophysical Journal794:2. DOI: http://doi.org/10.1088/0004-637X/794/2/141

Authors

Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, J. M. Cordes, P. B. Demorest, X. Deng, T. Dolch, J. A. Ellis, R. D. Ferdman, N. Garver-Daniels, F. Jenet, G. Jones, V. M. Kaspi, M. Koop, M. T. Lam, T. J.W. Lazio, A. N. Lommen, D. R. Lorimer, J. Luo, R. S. Lynch, D. R. Madison, M. A. McLaughlin, S. T. McWilliams, D. J. Nice, N. Palliyaguru, T. T. Pennucci, S. M. Ransom, and A. Sesana

The Astrophysical Journal, 794:141 (19pp), 2014 October 20 doi:10.1088/0004-637X/794/2/141

GRAVITATIONAL WAVES FROM INDIVIDUAL SUPERMASSIVE BLACK HOLE BINARIES IN CIRCULAR ORBITS: LIMITS FROM THE NORTH AMERICAN NANOHERTZ OBSERVATORY FOR GRAVITATIONAL WAVES

Z. Arzoumanian[1], A. Brazier[2], S. Burke-Spolaor[3], S. J. Chamberlin[4], S. Chatterjee[2], J. M. Cordes[2], P. B. Demorest[5], X. Deng[6], T. Dolch[2], J. A. Ellis[4], R. D. Ferdman[7], N. Garver-Daniels[8], F. Jenet[9], G. Jones[10], V. M. Kaspi[7], M. Koop[6], M. T. Lam[2], T. J. W. Lazio[11], A. N. Lommen[12], D. R. Lorimer[8], J. Luo[9], R. S. Lynch[7], D. R. Madison[2], M. A. McLaughlin[8], S. T. McWilliams[8], D. J. Nice[13], N. Palliyaguru[8], T. T. Pennucci[14], S. M. Ransom[5], A. Sesana[15], X. Siemens[4], I. H. Stairs[16], D. R. Stinebring[17], K. Stovall[18], J. Swiggum[8], M. Vallisneri[11], R. van Haasteren[11,19], Y. Wang[9], and W. W. Zhu[15]

NANOGrav Collaboration

[1] Center for Research and Exploration in Space Science and Technology and X-Ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA
[2] Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[3] California Institute of Technology, Pasadena, CA 91125, USA
[4] Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA; justin.ellis18@gmail.com
[5] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[6] Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA 16802, USA
[7] Department of Physics, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada
[8] Department of Physics, West Virginia University, P.O. Box 6315, Morgantown, WV 26505, USA
[9] Center for Gravitational Wave Astronomy, University of Texas at Brownsville, Brownsville, TX 78520, USA
[10] Department of Physics, Columbia University, New York, NY 10027, USA
[11] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91106, USA
[12] Department of Physics and Astronomy, Franklin and Marshall College, P.O. Box 3003, Lancaster, PA 17604, USA
[13] Department of Physics, Lafayette College, Easton, PA 18042, USA
[14] University of Virginia, Department of Astronomy, P.O. Box 400325, Charlottesville VA, 22904, USA
[15] Max-Planck-Institut für Gravitationsphysik, Albert Einstein Institut, Am Mülenber 1, D-14476 Golm, Germany
[16] Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[17] Department of Physics and Astronomy, Oberlin College, Oberlin, OH 44074, USA
[18] Physics and Astronomy Department, University of New Mexico, Albuquerque, NM, USA
Received 2014 April 21; accepted 2014 August 28; published 2014 October 1

ABSTRACT

We perform a search for continuous gravitational waves from individual supermassive black hole binaries using robust frequentist and Bayesian techniques. We augment standard pulsar timing models with the addition of time-variable dispersion measure and frequency variable pulse shape terms. We apply our techniques to the Five Year Data Release from the North American Nanohertz Observatory for Gravitational Waves. We find that there is no evidence for the presence of a detectable continuous gravitational wave; however, we can use these data to place the most constraining upper limits to date on the strength of such gravitational waves. Using the full 17 pulsar data set we place a 95% upper limit on the strain amplitude of $h_0 \lesssim 3.0 \times 10^{-14}$ at a frequency of 10 nHz. Furthermore, we place 95% sky-averaged lower limits on the luminosity distance to such gravitational wave sources, finding that $d_L \gtrsim 425$ Mpc for sources at a frequency of 10 nHz and chirp mass $10^{10}\ M_\odot$. We find that for gravitational wave sources near our best timed pulsars in the sky, the sensitivity of the pulsar timing array is increased by a factor of ∼four over the sky-averaged sensitivity. Finally we place limits on the coalescence rate of the most massive supermassive black hole binaries.

Key words: gravitation – pulsars: general

Online-only material: color figures

1. INTRODUCTION

The direct detection of gravitational waves (GWs) is a major goal of experimental physics and astrophysics. One of the most promising means of detecting GWs is through the precise timing of an array of millisecond pulsars (MSPs). The concept of a pulsar timing array (PTA) was first conceived of over two decades ago (Sazhin 1978; Detweiler 1979; Hellings & Downs 1983; Romani 1989; Foster & Backer 1990). Twenty years later, three main PTAs are in full operation around the world: the North American Nanohertz Observatory for Gravitational waves (NANOGrav; Jenet et al. 2009), the Parkes Pulsar Timing Array (PPTA; Manchester 2008), and the European Pulsar Timing Array (EPTA; Janssen et al. 2008). The three PTAs collaborate to form the International Pulsar Timing Array (IPTA; Hobbs et al. 2010), which will result in increased sensitivity to GWs through more data and longer time spans than any single PTA.

PTAs are most sensitive to GWs with frequencies in the nanohertz regime (i.e., 10^{-9} Hz–10^{-7} Hz). Potential sources of GWs in this frequency range include supermassive black hole binary systems (SMBHBs; Sesana et al. 2008), cosmic (super)strings (Olmez et al. 2010), inflation (Starobinsky 1979), and a first-order phase transition at the QCD scale (Caprini et al. 2010). The community has thus far focused mostly on stochastic

[19] Einstein Fellow.

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

In this paper we will use current-generation frequentist (Ellis et al. 2012) and Bayesian (Ellis 2013) data analysis pipelines to compute upper limits on the strain amplitude of continuous GWs from SMBHBs in circular orbits. In Section 2 we briefly review the radio observations and timing analysis. In Section 3 we describe the signal model used to describe the continuous GWs in the PTA band. In Section 4 we describe, in detail, the time domain likelihood function, the noise model, and the frequentist and Bayesian search pipelines. In Section 5 we apply our search and upper limit pipelines to the NANOGrav data set and report our findings. In Section 6 we present some predictions for future PTA sensitivities to continuous GWs and summarize our results. In Appendices A–C we derive the form of the frequency evolution of SMBHBs, and give full details on the computational implementation of our Bayesian code.

2. OBSERVATIONS AND TIMING ANALYSIS

The observational data used for this analysis are the same as those presented by Demorest et al. (2013)[20]; the reader is referred to that paper for a detailed description of the observations and timing analysis. Here we present a brief review of the relevant features of the data set. The timing data used here were acquired during 2005–2010 using two radio telescopes, the 305 m Arecibo telescope, and the 100 m Robert C. Byrd Green Bank Telescope (GBT). A total of 17 pulsars (8 at Arecibo, 10 at the GBT, with J1713 + 0747 observed by both telescopes) were monitored using a typical observational cadence of four to six weeks between sessions. At each observing epoch, every pulsar was observed using two separate receiver systems operating at widely separated radio frequencies ranging from 327 MHz to 2.3 GHz. The typical observation length was 30 minutes per pulsar per receiver. All data were recorded using the identical ASP (at Arecibo) and GASP (at the GBT) pulsar backend systems (Demorest 2007). These systems processed a typical radio bandwidth of 64 MHz using real-time coherent dedispersion and pulse period folding, resulting in 2048 bin full-Stokes pulse profiles averaged over one to three minutes in 4 MHz channels.

Pulse profile calibration, integration, and time of arrival (TOA) determination was done using standard techniques via the PSRCHIVE[21] software package (Hotan et al. 2004). For each pulsar all profiles in a given epoch were integrated in time to form a single set of profiles across radio frequency. From these, TOAs were measured *separately* in each 4 MHz radio frequency channel. This resulted in a set of ∼20–30 multi-frequency TOAs at each epoch, or ∼500–2000 TOAs total for each pulsar over the full data set. Before searching for the presence of GW in these data, the rotational, orbital, astrometric, and interstellar medium (ISM) properties specific to each pulsar—effects collectively known as the timing model—must first be determined from the TOA data. For this we analyzed the TOAs using both the TEMPO[22] and TEMPO2[23] (Hobbs et al. 2006) timing software packages and obtained identical results with both. Notable features of the timing models used here include: spin frequency and spin-down rate, but no higher frequency derivatives, were fit for all pulsars; all five astrometric parameters (sky position, proper motion, and parallax) were fit for all pulsars[24]; time-variable dispersion measure (DM) was included by fitting for an *independent* DM value at each epoch, using the multi-frequency TOAs;[25] intrinsic profile shape evolution with frequency was included as a constant-in-time offset for each frequency channel, and Keplerian and relativistic orbital elements, as appropriate for pulsars in binary systems. All TOA data and final timing solutions for this data set are publicly available online.[26]

3. GWs FROM SUPERMASSIVE BLACK HOLE BINARIES

Pulsar timing residuals are defined as the difference between observed TOAs of radio pulses and a deterministic timing model. In this section we review the form of the residuals induced by SMBHBs consisting of non-spinning black holes in a circular orbit (e.g., Wahlquist 1987) and introduce our notation. Spin effects are not likely to play any measurable role in the orbital dynamics (Sesana & Vecchio 2010) and eccentric systems (Roedig & Sesana 2012; Sesana 2013) will be addressed in a future work, as we choose not to include eccentricity here. The GW is a metric perturbation to flat spacetime defined in terms of its two polarizations as

$$h_{ab}(t,\hat{\Omega}) = e^{+}_{ab}(\hat{\Omega})h_{+}(t,\hat{\Omega}) + e^{\times}_{ab}(\hat{\Omega})h_{\times}(t,\hat{\Omega}), \tag{1}$$

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

$$e^{+}_{ab}(\hat{\Omega}) = \hat{m}_a\hat{m}_b - \hat{n}_a\hat{n}_b, \tag{2}$$

$$e^{\times}_{ab}(\hat{\Omega}) = \hat{m}_a\hat{n}_b + \hat{n}_a\hat{m}_b, \tag{3}$$

where

$$\hat{\Omega} = -(\sin\theta\cos\varphi)\hat{x} - (\sin\theta\sin\varphi)\hat{y} - (\cos\theta)\hat{z}, \tag{4}$$

$$\hat{m} = -(\sin\varphi)\hat{x} + (\cos\varphi)\hat{y}, \tag{5}$$

$$\hat{n} = -(\cos\theta\cos\varphi)\hat{x} - (\cos\theta\sin\varphi)\hat{y} + (\sin\theta)\hat{z}, \tag{6}$$

where $\hat{x}$, $\hat{y}$, and $\hat{z}$ are the usual Cartesian coordinate unit vectors. In this coordinate system, $\theta = \pi/2 - \delta$ and $\varphi = \alpha$ are the polar and azimuthal angles of the source, respectively, where δ and α are declination and right ascension in the usual equatorial

[20] Currently, NANOGrav is timing 43 pulsars with baselines as long as nine years. Full timing solutions and GW analysis papers will be forthcoming.
[21] http://psrchive.sourceforge.net
[22] http://tempo.sourceforge.net
[23] http://tempo2.sourceforge.net
[24] The best fit parallax for PSR J1640 + 2224 is consistent with zero and is therefore not fit in the final timing model.
[25] Models for pulsars J1853 + 1308, J1910 + 1256, and B1953 + 29 did not include DM variation measurement as only single-frequency data were available for these.
[26] http://data.nanograv.org

coordinates, where the north celestial pole is in the $\hat{z}$ direction and the vernal equinox is in the $\hat{x}$ direction.

We will write our GW induced pulsar timing residuals in the following form:

$$s(t, \hat{\Omega}) = F^{+}(\hat{\Omega})\Delta s_{+}(t) + F^{\times}(\hat{\Omega})\Delta s_{\times}(t), \tag{7}$$

where

$$\Delta s_A(t) = s_A(t_p) - s_A(t_e), \tag{8}$$

and t_e and t_p are the times at which the GW passes the Earth[27] and pulsar, respectively, and the index $A \in \{+, \times\}$ labels polarizations. The functions $F^A(\hat{\Omega})$ are known as antenna pattern functions and are defined by

$$F^{+}(\hat{\Omega}) = \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}} \tag{9}$$

$$F^{\times}(\hat{\Omega}) = \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})}{1+\hat{\Omega}\cdot\hat{p}}, \tag{10}$$

where $\hat{p}$ is the unit vector pointing from the Earth to the pulsar. Also, from geometry we can write[28]

$$t_p = t_e - L(1+\hat{\Omega}\cdot\hat{p}), \tag{11}$$

where L is the distance to the pulsar. Given these definitions, we can write the GW contributions to the timing residuals at 0-PN (Post-Newtonian) order as (Wahlquist 1987; Corbin & Cornish 2010)

$$s_{+}(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}[-\sin[2\Phi(t)](1+\cos^2\iota)\cos 2\psi - 2\cos[2\Phi(t)]\cos\iota\sin 2\psi] \tag{12}$$

$$s_{\times}(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}}[-\sin[2\Phi(t)](1+\cos^2\iota)\sin 2\psi + 2\cos[2\Phi(t)]\cos\iota\cos 2\psi], \tag{13}$$

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

$$\Phi(t) = \Phi_0 + \frac{1}{32\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega(t)^{-5/3}\right) \tag{14}$$

and

$$\omega(t) = \omega_0\left(1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}t\right)^{-3/8}, \tag{15}$$

where Φ_0 and ω_0 are the initial values at the time of our first observation, the chirp mass is defined by $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1+m_2)^{1/5}$, where m_1 and m_2 are the masses of the two SMBHs, and d_L is the luminosity distance to the SMBHB source. See Appendix A for a more complete derivation of the frequency evolution of the binary including important approximations that can be made. For our purposes here we will just note that orbital frequency evolution over our observing time span is very unlikely while frequency evolution over the Earth–pulsar light travel time is almost certain for sources with reasonably large chirp masses (i.e., $\mathcal{M} \sim 3\times 10^8\, M_\odot$; see, e.g., Sesana & Vecchio 2010). We can relate the GW frequency to the orbital frequency of the binary by $\omega_{\rm gw} = 2\omega_0$ for circular orbits. Note that we use the observed redshifted values. For example, the chirp mass and orbital angular frequency in the rest frame are $\mathcal{M}_r = \mathcal{M}/(1+z)$ and $\omega_r = \omega_0(1+z)$, respectively, where z is the cosmological redshift.

From the signal model presented above, we see that our parameter space is eight-dimensional and the continuous wave parameter space vector is

$$\boldsymbol{\lambda}_0 = \{\theta, \varphi, \Phi_0, \psi, \iota, \mathcal{M}, d_L, \omega_0\}. \tag{16}$$

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

$$h_0 = 2\frac{\mathcal{M}^{5/3}(\pi f_{\rm gw})^{2/3}}{d_L}. \tag{17}$$

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

4. SEARCH TECHNIQUES

4.1. Likelihood Function for PTAs

It was shown in van Haasteren & Levin (2013) (and modified in Ellis 2013 to include deterministic sources) that the likelihood function for the residuals, marginalized over the timing model parameters, can be written as

$$p(\delta t|\boldsymbol{\phi}, \boldsymbol{\lambda}) = \frac{\exp\left[-\frac{1}{2}(\delta t - s)^T G(G^T C G)^{-1} G^T (\delta t - s)\right]}{\sqrt{(2\pi)^{(N_{\rm TOA}-m)}\det(G^T C G)}}, \tag{18}$$

where δt is a vector of the pulsar residuals, $\boldsymbol{\phi}$ are parameters that describe the noise in the pulsar residuals, $\boldsymbol{\lambda}$ are the parameters that characterize the continuous GW signal,[29] and C is the covariance matrix of the noise. In practice, the noise in pulsar timing residuals is non-Gaussian due to ISM scintillation effects which are manifested through a time varying pulse intensity. Nonetheless, the noise in each residual is modeled very well by a Gaussian with zero mean and standard deviation equal to the uncertainty on the TOA. In other words, the noise in the *weighted* (by the individual TOA errors) residuals is very well approximated as a Gaussian. As will be detailed in the next section, we include these error bar weights in our noise covariance matrix. Thus, assuming the noise is Gaussian[30] is justified. Here G is an $N_{\rm TOA}\times(N_{\rm TOA}-m)$ matrix with $N_{\rm TOA}$ the

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

[28] Here we use geometrized units where $G = c = 1$.

[29] We have dropped the α subscript on the parameter vector here as we are only considering a single pulsar.

[30] Note that here Gaussian noise simply means that the data obey a multivariate Gaussian probability distribution function. This does not mean that we assume the data is white.

number of TOAs and m the number of fitted parameters in the timing model. The derivation of the G-matrix approach can be found in van Haasteren & Levin (2013) and will not be explored here. The matrix G^T is a projection operator that projects our data δt onto the null space of the timing model design matrix M, that is, it projects the data into a subspace orthogonal to the linearized timing model. In the timing analysis used here, DM variation and profile frequency evolution effects are part of the timing model, and these terms are included when constructing the G matrix. In this way we have fully taken into account the timing model fitting procedure.

For this work we will assume that the noise in the residuals between pulsars is uncorrelated. In other words, we are assuming that the stochastic GW background will be negligible compared to the intrinsic noise in each pulsar. In general this is not likely to be a good assumption when we would expect a detection of a single GW source. Furthermore, terrestrial clock errors (Hobbs et al. 2012) and errors in solar system ephemerides (Champion et al. 2010)[31] can also cause correlations between the noise in the residuals from different pulsars, with however different angular correlation properties on the sky than are expected from GWs. The effects of omitting the correlations in the likelihood function are unknown and will be the subject of future work. Under these assumptions, the likelihood function for the full PTA can be written as

$$p(\delta t|\boldsymbol{\lambda}) = \prod_{\alpha=1}^{N_{\rm psr}} p(\delta t_\alpha|\boldsymbol{\lambda}_\alpha), \tag{19}$$

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

$$\ln\Lambda = \sum_{\alpha}^{N_{\rm psr}} \left[(\delta t_\alpha|s(\boldsymbol{\lambda}_\alpha)) - \frac{1}{2}(s(\boldsymbol{\lambda}_\alpha)|s(\boldsymbol{\lambda}_\alpha)) \right], \tag{20}$$

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

$$(x|y) = x^T G(G^T C G)^{-1} G^T y. \tag{21}$$

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

$$\rho = \sqrt{2\langle\ln\Lambda\rangle} = \left(\sum_{\alpha=1}^{N_{\rm psr}} (s(\boldsymbol{\lambda}_\alpha)|s(\boldsymbol{\lambda}_\alpha)) \right)^{1/2}, \tag{22}$$

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

4.2. Noise Model

In the above section we have derived the likelihood function used for our analysis; however, we have not specified the form of the noise covariance matrix C. Previous Bayesian analysis schemes (van Haasteren et al. 2009, 2011; van Haasteren & Levin 2010, 2013; Ellis et al. 2012; van Haasteren 2013; Ellis et al. 2013; Ellis 2013) have used a power-law red noise model and EFAC (constant multiplier on the TOA uncertainties) and EQUAD (additional Gaussian white noise added in quadrature to EFAC noise) parameters to describe the white noise, with a covariance matrix of the form

$$C = E^2 W + \mathcal{Q}^2 \mathbb{I} + C^{\rm red}(A_{\rm red}, \gamma_{\rm red}), \tag{23}$$

where E is the EFAC parameter, $W = {\rm diag}\{\sigma_i^2\}$, with σ_i the uncertainty on the ith TOA, $\mathcal{Q}$ is the EQUAD parameter, and $C_{\rm red}$ is an analytic expression of the red noise amplitude $A_{\rm red}$ and spectral index $\gamma_{\rm red}$. It is worth noting that we use no EFAC or EQUAD parameters in our pulsar timing model fit but instead include them directly in our noise model. The EFAC is simply a parameter that quantifies any additional uncertainty in the TOA uncertainties and the EQUAD parameter quantifies any additional white noise that is not related to the formal TOA uncertainties. In principle, a different EFAC value should be used for each pulsar timing backend as this parameter is related to intrinsic receiver noise; however, in this five-year NANOGrav data set, only one backend per telescope was used.[32] Therefore, we are justified in only using one EFAC parameter per pulsar. This noise model is quite general and works well for many pulsars; however, the size of the matrices is quite large (on the order of $10^3 \times 10^3$) and inversion is a large bottleneck in the analysis pipelines. Furthermore, current NANOGrav observing schemes produce large sets of multifrequency observations that are essentially simultaneous. One may be tempted to simply perform a weighted average of the TOAs and work with the new reduced data sets but in the Bayesian scheme we must marginalize over the timing model parameters analytically and it is unclear how to carry out this process for epoch-averaged TOAs. Because of this, we have developed a framework to essentially work backward from the marginal likelihood to derive a nearly exact averaging scheme. First we re-write our noise covariance matrix

$$C = N + U\tilde{C}U^T, \tag{24}$$

where $\tilde{C}$ is a $q \times q$ reduced covariance matrix with q the number of epochs[33] in our data set, N is a white noise covariance matrix of the EFAC and EQUAD terms, and U is the "exploder" matrix that maps epochs (columns) to the full set of TOAs (rows). If we now make use of this new formalism, the likelihood function is then

$$p(\delta t|\boldsymbol{\phi}, \boldsymbol{\lambda}) = \frac{\exp\left[-\frac{1}{2}((\delta t - s)^T \tilde{N}^{-1}(\delta t - s) - d^T \Sigma^{-1} d)\right]}{\sqrt{(2\pi)^{n-m}\det(\tilde{C})\det(G^T N G)\det(\Sigma)}}, \tag{25}$$

where we have used the Woodbury Lemma[34] to compute the inverse and determinant of C, $\tilde{N}^{-1} = G(G^T N G)^{-1} G^T$, $d = U^T \tilde{N}^{-1}(\delta t - s)$, and $\Sigma = (\tilde{C}^{-1} + U^T \tilde{N}^{-1} U)$. Note that d here are essentially daily averaged residuals. For NANOGrav data sets the number of epochs per pulsar is on the order of 30–100, while the total number of TOAs per pulsar is on the order of 10^3, thus the inversions (here N is diagonal and $\tilde{N}^{-1}$ can be pre-computed, thus the only dense matrix inversion is

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

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

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

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

Σ^{-1}) required in this likelihood function scale as q^3 as opposed to n^3, resulting in computational speed ups of several orders of magnitude. Furthermore, the epoch-averaged covariance matrix $\tilde{\mathcal{C}}$ can take on several forms depending on the red noise model used; however, as long as it is a slowly varying function of the TOAs (i.e., a truly red process) then this formalism is completely valid.

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

$$\delta t = M\delta\boldsymbol{\xi} + n_{\rm white} + n_{\rm red} + s, \tag{26}$$

where M is the design matrix (an $N_{\rm TOA} \times m$ matrix composed of the functional form of the linearized timing model (column) evaluated at each TOA (row)), $n_{\rm white}$ and $n_{\rm red}$ are the white and red components of the residuals, respectively, and $\delta\boldsymbol{\xi}$ are the timing model parameter offsets. It is possible to expand the red noise piece in a Fourier series

$$n_{\rm red} = \sum_{j=1}^{N_{\rm mode}} \left[a_j \sin\left(\frac{2\pi jt}{T}\right) + b_j \cos\left(\frac{2\pi jt}{T}\right)\right] = F\mathbf{a}, \tag{27}$$

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

$$\varphi_{ij} = \langle \mathbf{a}\mathbf{a}^T\rangle_{ij} = \delta_{ij}\varphi_i, \tag{28}$$

where there is no sum over i and the elements of φ, denoted $\{\varphi_i\}$ are the coefficients of the model power spectrum of the red noise process in the residuals. If the red noise process is wide-sense stationary, then this relation is always true irrespective of the sampling as all information about the uneven sampling here comes from the Fourier design matrix F. Thus, we can write the covariance and epoch-averaged covariance matrices, respectively, as

$$C = N + F\varphi F^T \tag{29}$$

$$\tilde{\mathcal{C}} = \tilde{F}\varphi\tilde{F}^T, \tag{30}$$

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

$$\varphi_i = \frac{1}{T}\frac{A_{\rm red}^2}{12\pi^2}\left(\frac{f_i}{f_{\rm yr}}\right)^{3-\gamma_{\rm red}} f_i^{-3}, \tag{31}$$

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

In the course of our single pulsar noise analysis (J. A. Ellis et al. 2014, in preparation) we found that the addition of an extra white noise parameter was needed to accurately describe the data. This new white noise term incorporates a correlation among frequency channels (within a given epoch) while still remaining independent of other epochs. In other words, this white noise term accounts for epoch-to-epoch fluctuations as opposed to fluctuations within an epoch. We defer to another paper the inclusion of pulse-jitter noise from pulsar magnetospheric activity but point out that our inferred extra term may be the same as jitter noise known to be present in all well-studied pulsars (Cordes & Shannon 2010). This parameter is quite easy to incorporate as it is simply an EQUAD-like parameter in the epoch-averaged sense, that is,

$$J = U\tilde{J}U^T = \mathscr{J}^2 U\mathbb{I}_q U^T, \tag{32}$$

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

$$C = N + U\left(\tilde{F}\varphi\tilde{F}^T + \mathscr{J}^2\mathbb{I}_q\right)U^T, \tag{33}$$

and noise parameter vector

$$\boldsymbol{\phi} = \{E, \mathcal{Q}, \mathscr{J}, A_{\rm red}, \gamma_{\rm red}\}. \tag{34}$$

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

4.3. $\mathcal{F}_p$-Statistic

The $\mathcal{F}_p$ statistic was first derived from the likelihood function of Equation (18) in Ellis et al. (2012, hereafter ESC12) as a "total-power" frequentist detection statistic. We will not derive the full expression here (see Appendix B for an alternate derivation to ESC12), but rather we will explain its functional form and discuss its statistics. First we define the following harmonic basis functions:

$$B_\alpha^1(t) = \frac{1}{\omega_0^{1/3}}\sin(2\omega_0 t) \tag{35}$$

$$B_\alpha^2(t) = \frac{1}{\omega_0^{1/3}}\cos(2\omega_0 t), \tag{36}$$

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

$$2\mathcal{F}_p = \sum_{\alpha=1}^{N_{\rm psr}} P_\alpha^i Q_{ij}^\alpha P_\alpha^j, \tag{37}$$

where we have assumed Einstein summation notation over latin indices, $P_\alpha^i = (\delta t|B_\alpha^i(t))$, $Q_{ij}^\alpha = (B_i^\alpha|B_j^\alpha)$ and the formal sum is over all pulsars in the array. An intuitive way to think of this

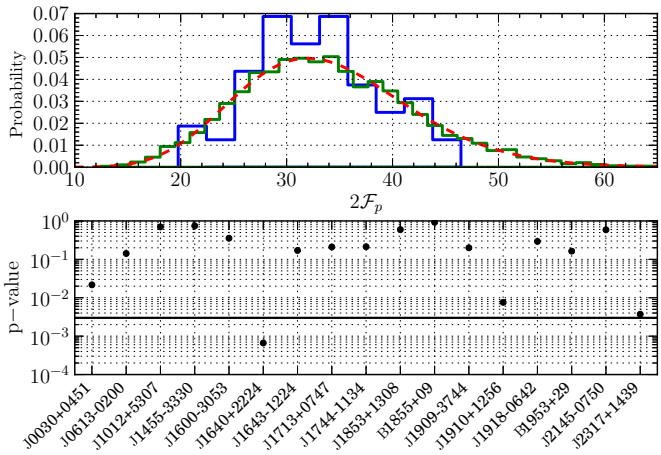

Figure 1. Histogram of $\mathcal{F}_p$ statistic values (top panel) across all independent frequencies (black (blue) histogram) and for 100,000 realizations of simulated data with noise parameters measured from the real data (gray (green) histogram). The red dashed curve is the probability distribution function for a χ^2 distribution with 34 (i.e., $2N_{\rm psr}$) degrees of freedom. The lower panel shows the p-value from a KS test comparing the $\mathcal{F}_p$ statistic for each pulsar to a χ^2 distribution with two degrees of freedom. The solid line represents the 3σ threshold for the p-value. (A color version of this figure is available in the online journal.)

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

$$\langle 2\mathcal{F}_p\rangle = 2N_{\rm psr} + \rho^2. \tag{38}$$

Figure 1 shows the distribution of the $\mathcal{F}_p$ statistic (top panel) for both real and simulated data as well as a p-value test (bottom panel) for each pulsar, where we compare the single-pulsar $\mathcal{F}_p$ distribution to the expected χ^2 distribution. To compute the $\mathcal{F}_p$-statistic, we have used the maximum a posteriori noise values obtained in a previous single-pulsar noise analysis to construct the noise covariance matrix. Since we do not have independent realizations of our data, we compute the $\mathcal{F}_p$-statistic for each independent[35] frequency bin and then construct a histogram of the results. If our noise model is a good description of the true noise in our data and there is no GW present in the data then this distribution should follow the correct χ^2 distribution. We see from Figure 1 that the $\mathcal{F}_p$-statistic values do indeed follow a χ^2 distribution with $2N_{\rm psr}$ degrees of freedom. The black (blue) curve in the top panel of Figure 1 shows the aforementioned histogram along with the χ^2 distribution in the dashed gray (red) line. The p-value that results from a Kolmogorov–Smirnov (KS) test comparing the $2\mathcal{F}_p$ and χ^2 (with 34 degrees of freedom) distributions is 0.33, showing good agreement between our data and the expected χ^2 distribution. As a cross-check, we have also simulated 100,000 data sets with the measured noise parameters and have evaluated the $\mathcal{F}_p$ statistic for each. This distribution is plotted as a gray (green) histogram in the figure and it is obvious that this distribution follows a χ^2 distribution with 34 degrees of freedom nearly perfectly. We have also performed a similar test but for each pulsar separately. In the lower panel of Figure 1 we carry out the same KS test mentioned above but now compute the $\mathcal{F}_p$ statistic values for each pulsar individually and then compare to a χ^2 with two degrees of freedom. The solid line corresponds to the p-value at which we should reject the null hypothesis that the two distributions are the same with 99.7% confidence. We see that with the exception of one pulsar, J1640 + 2224, all others lie above this threshold value. This indicates that our noise model for all pulsars, except for J1640 + 2224, provides a good description of the true noise in the data set. Better noise models for this pulsar are currently being explored (J. A. Ellis et al. 2014, in preparation). Since our full 17 pulsar $\mathcal{F}$-statistic distribution is totally consistent with the expected χ^2 distribution and PSR J1640+2224 does not bias the overall PTA distribution, we use the standard noise model described in Section 4.2.

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

$$p_0(\mathcal{F}_p) = \frac{\mathcal{F}_p^{n/2-1}}{(n/2-1)!}\exp(-\mathcal{F}_p), \tag{39}$$

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

$$P_F(\mathcal{F}_{p,0}) = \int_{\mathcal{F}_{p,0}}^{\infty} p_0(\mathcal{F}_p)d\mathcal{F}_p = \exp(-\mathcal{F}_{p,0})\sum_{k=0}^{n/2-1}\frac{\mathcal{F}_{p,0}^k}{k!}. \tag{40}$$

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

$$P_F^T(\mathcal{F}_{p,0}) = 1 - [1 - P_F(\mathcal{F}_{p,0})]^{N_f}, \tag{41}$$

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

4.4. Bayesian Method

The Bayesian search pipeline in this work is very similar to that of Ellis (2013, hereafter E13). Here we use an MPI-enabled Parallel Tempered Markov Chain Monte Carlo (PTMCMC) sampler[36] (see E13 and Appendix C.1 for details on the implementation). In this work we use two "modes" of operation for the Bayesian search. The first is the most general in which we evaluate the full likelihood function of Equation (18) and allow both the GW parameters, $\boldsymbol{\lambda}$, and the noise model parameters, $\boldsymbol{\phi}$ to vary simultaneously. In principle, this is the more desirable setup as it allows the uncertainty in our noise model to propagate into the measured GW parameters and also accounts for any correlations between the noise and GW parameters. This mode does require significantly more computational power as the number of search parameters in the MCMC is quite large. The total parameter space consists of eight GW parameters, $N_{\rm psr}$ pulsar distances, and $5 \times N_{\rm psr}$ noise parameters; this comes to 110 parameters for the full 17 pulsar array.

The second mode is when we fix the noise parameters to their maximum a posteriori values obtained from a previous single

[35] Note that the frequencies are not completely independent since our data are irregularly sampled. The frequency bins were chosen here assuming a cadence of two observing sessions per month, giving a maximum frequency of $f_{\rm max} = 8 \times 10^{-7}$ Hz and minimum frequency of $f_{\rm min} = 1/T_{\rm obs} = 6 \times 10^{-9}$ Hz.

[36] https://github.com/jellis18/PAL

pulsar analysis. All previous GW searches for single sources have been performed in this manner (Jenet et al. 2004; Yardley et al. 2010; Babak & Sesana 2012; Ellis et al. 2012; Petiteau et al. 2013; Ellis 2013), which is justified *if* the noise model only contains white noise and the GW signal present in any single data set is weak. If the noise model contains only white noise, there will be little to no correlation between the GW parameters and the noise parameters, and if the signal is weak then it will not affect the single pulsar noise analysis. However, there is some evidence of red noise in our pulsars and because of the highly varying noise levels among pulsars, it is likely that a *detectable* source would be seen in the best timed pulsars individually. Therefore, this type of Bayesian analysis is not robust and could possibly lead to biased results; nonetheless, we will carry out this mode for comparison purposes in this study. Note that we will have the same problem with the $\mathcal{F}_p$ statistic. Possible methods to ameliorate this problem in fixed-noise searches are being explored and will be the subject of a future paper.

In a Bayesian sense, the detection problem is an application of model selection and the upper limit problem is an application of parameter estimation. The upper limit problem will be discussed in the next section. The model selection technique is very similar to that discussed in E13 and we refer the reader there for more details (see Appendix C.2 for more details on the implementation). If we want to perform model selection to claim a detection or compare different waveforms then we can make use of the Bayesian odds ratio between models "A" and "B"

$$\mathcal{O} = \frac{p(d|\mathcal{H}_A)}{p(d|\mathcal{H}_B)}\frac{p(\mathcal{H}_A)}{p(\mathcal{H}_B)}, \tag{42}$$

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

In this work we wish to compare the signal model $\mathcal{H}_1$ parameterized by $\{\boldsymbol{\lambda}, \boldsymbol{\phi}\}$ to the null-hypothesis model $\mathcal{H}_0$ parameterized by $\boldsymbol{\phi}$. When we allow the noise and GW parameters to vary simultaneously we will need to compute the evidence for both models, $\mathcal{H}_1$ and $\mathcal{H}_0$, separately. On the other hand, when we fix the noise parameters we can simply evaluate the likelihood-ratio in (20) and use this to compute the evidence. In this case, since the null-hypothesis model has no free parameters, the Bayes factor is simply given by the evidence computed using the likelihood ratio. In practice, to compute the evidence we make use of thermodynamic integration as detailed in E13 and Appendix C.2. The results of our Bayesian search and verification on injections will be discussed in the next section.

4.4.1. Priors

In a Bayesian analysis, especially when using parallel tempering and thermodynamic integration, it is very important to choose reasonable priors so that we are not exploring areas of parameter space that have been ruled out by previous experiments. We choose isotropic priors on all angular parameters and uniform priors in the log of the chirp mass with $\mathcal{M} \in [10^8, 10^{10}]\, M_\odot$, log of the luminosity distance with $d_L \in [1, 10^4]$ Mpc, and log of the frequency of the GW with $f_{\rm gw} \in [6 \times 10^{-9}, 4 \times 10^{-7}]$ Hz. We impose an additional condition on the strain amplitude such that $h_0(\mathcal{M}, d_L, f_{\rm gw}) \leqslant h_{\rm ref}(f_{\rm gw}/f_0)^{2/3}$, where $h_{\rm ref} = 1 \times 10^{-13}$ and $f_0 = 10^{-8}$ Hz. This value is chosen so that the maximum strain is well above the level of detection. Essentially this is a cheap way to impose a correlated prior on chirp mass, luminosity distance, and GW frequency. The normalization is computed through Monte Carlo integration. For the pulsar distance prior we use the current electromagnetic (EM) measurements either from timing parallax or very long baseline interferometry corresponding to the best measured values taken from Verbiest et al. (2012; 10 pulsars) if available, otherwise, we use the values from the Australia National Telescope Facility (ATNF) pulsar catalog (Manchester et al. 2005)[37] which have distances based on dispersion measure and the NE2001 Galactic electron-density model (Cordes & Lazio 2002, 2003). For pulsars without parallax distances we assume a 20% uncertainty on the distance. Using this information, we write the distance prior as follows

$$p(\boldsymbol{L}) = \prod_{\alpha=1}^{N_{\rm psr}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left(-\frac{\left(L_\alpha - L_\alpha^{\rm EM}\right)^2}{2\sigma_\alpha^2}\right), \tag{43}$$

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

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

$$\begin{aligned}\sigma_{\rm red} &= \int_{1/T}^{\infty} df\, P(f) \\ &= 2.05 \frac{1}{\sqrt{\gamma_{\rm red} - 1}} \left(\frac{A_{\rm red}}{10^{-15}}\right)\left(\frac{T}{1\ {\rm yr}}\right)^{\frac{\gamma_{\rm red}-1}{2}}\ {\rm ns},\end{aligned} \tag{44}$$

with T the total observation time and $P(f)$ the power spectrum of the red noise. This prior essentially restricts the model from considering red-noise-dominated residuals, which is a very good approximation (Perrodin et al. 2013; J. A. Ellis et al. 2014, in preparation). This prior is chosen because it leads to much more computationally efficient runs by allowing us to run fewer high temperature chains in the Thermodynamic Integration scheme (see Appendix C.2 for more details). In principle this red noise prior is illegal in the sense that it uses the data (i.e., the variance

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

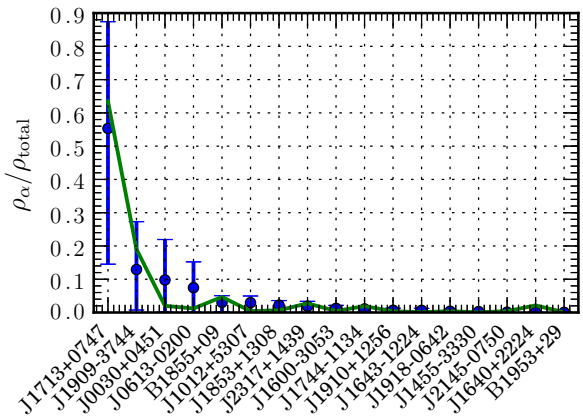

Figure 2. Fraction of S/N that each pulsar contributes (black (blue) points). We see that PSR J1713 + 0747 dominates the total S/N. The gray (green) curve is a simple $1/\sigma_\alpha$ scaling which matches the measured S/N values quite well showing that the overall variance of the noise for each pulsar is the dominating factor in determining the overall S/N.

(A color version of this figure is available in the online journal.)

of the residuals); however, this prior restricts access to an area of parameter space that is not supported by the likelihood function. Therefore, by omitting this area of parameter space the evidence calculation for each model, $\mathcal{H}_1$ and $\mathcal{H}_0$, will be biased slightly low. However, the likelihood function evaluated at this area of parameter space is essentially zero, and this slight bias will be negligible.

5. RESULTS

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

5.1. Verification

First, it is interesting to determine how much each pulsar in the 17 pulsar array will contribute to the overall S/N (signal-to-noise ratio) when a GW is present. In Figure 2 we plot the fraction $\rho_\alpha/\rho_{\rm total}$, where ρ_α is the single pulsar S/N, for each pulsar in the array. To compute this fraction we simulate 5000 S/N = 10 GW realizations (with parameters drawn from isotropic distributions in all angles and distributions uniform in the log of chirp mass and frequency) and calculate the single pulsar and total PTA S/N from Equation (22). The black (blue) points in the plot show the mean and standard deviation of the aforementioned ratio for each pulsar and the gray (green) curve is a simple naive scaling of $1/\sigma_\alpha^2$, where σ_α is the weighted rms of the αth pulsar's TOA uncertainties. It is obvious that J1713+0747 contributes more than 55% of the S/N on average, and PSRs J1909−3744, J0030+0451, and J0613−0200 contribute ∼10% on average. As we see from the gray (green) curve, this is very consistent with the overall scaling with the inverse of the variance of the noise; however, PSRs J0030+0451 and J0613−0200 carry a higher percentage because they are located opposite the bulk of other pulsars on the sky, and therefore will contribute more to the S/N for GWs coming from that side of the sky due to the antenna pattern response. This calculation does not mean that we advocate only timing the pulsars with the highest timing precision. Although many of the lower timing precision pulsars do not help with continuous GW detection or parameter estimation, they are *essential* for detection and parameter estimation of a stochastic GW background (see, e.g., Siemens et al. 2013).

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

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

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

We have carried out this analysis and the results are shown in Figure 3 where we plot $\mathcal{F}_p$ versus GW frequency when using the measured noise values (black) and the true injected noise values (gray). The vertical dashed line indicates the injected frequency

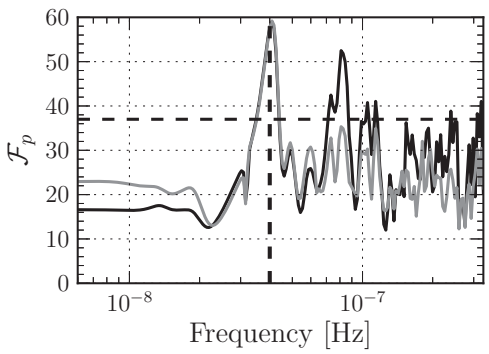

Figure 3. $\mathcal{F}_p$ statistic evaluated over the frequency range $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ Hz. The horizontal dashed line corresponds to our detection threshold of FAP $= 10^{-4}$ and the vertical dashed line denotes the injected frequency. The black and gray curves are the $\mathcal{F}_p$-statistic values when using the measured and true noise parameters, respectively. See text for more details.

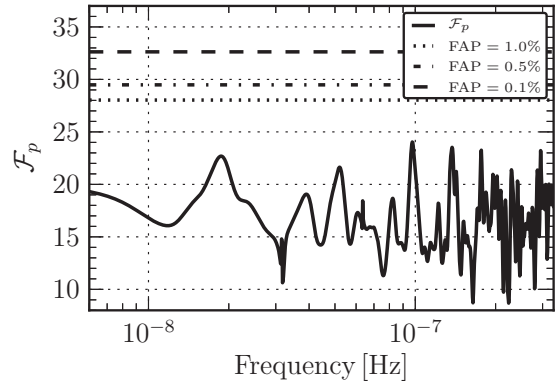

Figure 4. $\mathcal{F}_p$ statistic evaluated over the frequency range $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ Hz. These frequencies were chosen based on the fact that the approximate cadence is 2.5 week^{-1}. The dashed, dash-dotted, and dotted lines represent the value of $\mathcal{F}_p$ that gives an FAP of 0.1%, 0.5%, and 1%, respectively. Here we note that there is no evidence for a detection and the data are consistent with the null hypothesis.

and the horizontal dashed line represents our detection threshold corresponding to an FAP of 10^{-4}. To compute the total FAP over all frequencies we make use of Equation (41) and choose $N_f = 324$, resulting in a total FAP of 1.6×10^{-8} which is a decisive detection. The number of independent frequencies is difficult to calculate when we are using many data sets with very irregular sampling. In this work we have chosen $N_f = 324$ as this corresponds to $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ and $\Delta f_{\rm gw} = 10^{-9}$ Hz. The upper limit on frequency was chosen because our approximate observing cadence is 2.5 weeks^{-1} and the frequency spacing was chosen by imposing the condition that the autocorrelation function of $\mathcal{F}_p$ when no signal is present drops to half of its maximum value at that frequency lag. This analysis shows us that we can indeed detect a continuous GW if it is present in our data by conducting a fully blind search; however, we also see that our results will not be conclusive as there are several frequencies at which the FAP is above our threshold value. From comparison with the true-noise case, we see that the uncertainty (and residual correlations between GW and noise parameters) in the noise parameters can lead to confusing results. This again is mostly due to the fact that our sensitivity is dominated by a small number of pulsars. Because of this, we caution against using a fixed-noise method to make final detection statements but instead advocate these methods as a first round in a suite of analyses.

Both Bayesian pipelines (with and without varying noise parameters) were run on this synthetic data set. For both runs we have used PTMCMC and thermodynamic integration as discussed in Appendix C.2. Due to the large parameter spaces when using the full GW and noise model, we have chosen to use only the pulsars that contribute more than 1% to the injected S/N, resulting in six pulsars, J1713+0747, J1909−3744, B1855+09, J0030+0451, J0613−0200, and J1012+5307. Here we use the same noise parameters as mentioned above for our fixed-noise search. Even though these estimates are different from the true noise parameters, we nonetheless achieve a log-Bayes factor of 27.4 for the fixed-noise search (a log-Bayes factor greater than five is considered decisive evidence). However, as we mentioned earlier, we should not totally trust this level of evidence as it does not fully incorporate our uncertainty in the noise model. When we run an analysis where we allow the noise and GW parameters to vary simultaneously, we only achieve a log-Bayes factor of 5.35. While still decisive, this search is much less sensitive to the GW; nonetheless, this search is the most robust and will be the real test as to whether or not one can trust a real GW detection candidate. Of course these results could change depending on the noise realization or GW parameter combinations. A more detailed study of this is warranted but beyond the scope of this paper. Nonetheless, the large spread of overall noise levels in modern PTAs will most likely make the confident detection of a continuous GW very difficult.

5.2. Search Results

First we will discuss the results of the $\mathcal{F}_p$ statistic search on the real 17 pulsar NANOGrav data. To carry out the analysis we have computed $\mathcal{F}_p$ for many frequencies with $f_{\rm gw} \in [1/T, 3.3 \times 10^{-7}]$ Hz. These frequencies were chosen based on the fact that the approximate cadence is 2.5 week^{-1}. The results of this search are shown in Figure 4, where the solid black line is the value $\mathcal{F}_p$ at each frequency; and the dotted, dash-dotted, and dashed lines are the value of $\mathcal{F}_p$ corresponding to a 1.0%, 0.5%, and 0.1% FAP, respectively, where these values are calculated from Equation (40). Furthermore if we maximize $\mathcal{F}_p$ over frequencies then the total FAP, accounting for the trials factor N_f, is very nearly 1, indicating that we should accept the null hypothesis (no visible GW signal) with very high confidence.

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

5.3. Upper Limits

In this section we will outline the procedures used to compute both the frequentist and Bayesian upper limits on the strain amplitude, h_0. First we wish to state that an $x\%$ upper limit on the strain amplitude does not mean that we would have detected a signal with that amplitude with 95% confidence, it simply means that the *true* value of the amplitude is less than the upper limit with 95% probability (this probability measures the frequency of measuring that value of the amplitude in the frequentist case and the degree of belief in the true value of the amplitude in the Bayesian case). In the following sections we will discuss the mathematics of upper limit computation in the frequentist and Bayesian frameworks, and then we will lay out our computational procedure.

5.3.1. Frequentist Approach

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

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

$$p(\mathcal{F}_p|h_0) = \int p(\mathcal{F}_p|h_0, \tilde{\lambda}, \mathbf{n}) p(\tilde{\lambda}) p(\mathbf{n})\, d\tilde{\lambda}\, d\mathbf{n}, \tag{45}$$

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

$$p(\mathbf{n}) = \frac{1}{\sqrt{\det 2\pi C}} \exp\left(-\frac{1}{2}\mathbf{n}^T C^{-1}\mathbf{n}\right), \tag{46}$$

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

$$\begin{aligned}\alpha &= \int_{\mathcal{F}_{p,0}}^{\infty} p(\mathcal{F}_p|h_0)\, d\mathcal{F}_p \\ &= \left\langle \frac{1}{N}\sum_{i=1}^{N} \begin{Bmatrix} 1 & \text{if } \mathcal{F}_{p,i} \geqslant \mathcal{F}_{p,0} \\ 0 & \text{otherwise} \end{Bmatrix} \right\rangle,\end{aligned} \tag{47}$$

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

In practice, we carry out the following computational procedure.

1. Measure the value $\mathcal{F}_{p,0}$ from the real 17 pulsar NANOGrav data set as described in Section 4.3.
2. Simulate a synthetic noise vector $\mathbf{n} = L\mathbf{w}$ for each pulsar, where L is the Cholesky decomposition of the noise covariance matrix C, and $\mathbf{w}$ is a unit variance, zero-mean, vector.
3. Choose strain amplitude, h_0 and construct a GW waveform $\mathbf{s}(t, h_0, \tilde{\lambda})$ for each pulsar where the parameters $\tilde{\lambda}$ are drawn from the distribution $p(\tilde{\lambda})$.
4. Construct a new set of residuals for each pulsar $\delta t_{\rm sim} = R(\mathbf{n}+\mathbf{s}(t, h_0, \tilde{\lambda}))$, where R is the so-called fitting projection matrix introduced in Demorest et al. (2013) and Ellis et al. (2013).[38]
5. Now measure the value $\mathcal{F}_{p,i}$ for the simulated data set.
6. Repeat steps 2–5 10,000 times and measure the number of realizations that result in $\mathcal{F}_{p,i} > \mathcal{F}_{p,0}$.
7. Repeat steps 2–6 with different values of h_0 until 95% of simulations result in $\mathcal{F}_{p,i} > \mathcal{F}_{p,0}$.

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

5.3.2. Bayesian Approach

As mentioned above, in the Bayesian framework we do not rely on simulations as we treat the data as fixed and integrate the posterior PDF of the parameter of interest to compute the upper limit. In principle, a Bayesian upper limit is much more simple and intuitive than a frequentist upper limit. To compute a Bayesian upper limit we compute an integral that is analogous to Equation (47)

$$\begin{aligned}\alpha &= \int_0^{h_{\rm up}} dh_0 \int d\tilde{\lambda}\, d\boldsymbol{\phi}\; p(\delta t|h_0, \tilde{\lambda}, \boldsymbol{\phi}) p(h_0) p(\tilde{\lambda}) p(\boldsymbol{\phi}) \\ &= \int_0^{h_{\rm up}} dh_0\; p(\delta t|h_0) p(h_0),\end{aligned} \tag{48}$$

where $p(\delta t|h_0, \tilde{\lambda}, \boldsymbol{\phi})$ is the likelihood function; and $p(h_0)$, $p(\tilde{\lambda})$, $p(\boldsymbol{\phi})$ are the prior probability distributions on h_0, $\tilde{\lambda}$, and $\boldsymbol{\phi}$, respectively, where $\boldsymbol{\phi}$ denotes the noise model parameters. In words, we simply integrate the marginalized posterior distribution of h_0 until the desired credible region corresponding to a probability of α is reached at $h = h_{\rm up}$. As in the frequentist case, we want upper limits on the strain amplitude as a function of GW frequency or sky location. In this case we simply fix the parameters and then marginalize over the others. In practice, to compute the Bayesian upper limits we carry out a separate MCMC run for fixed values of frequency and/or fixed sky locations and then compute the 95% upper limit for each. The choice

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

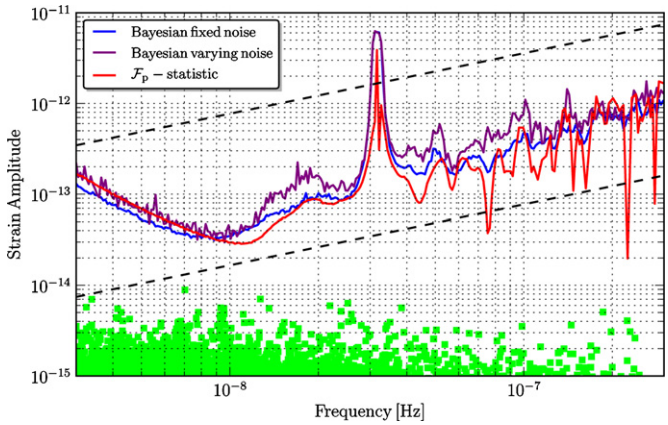

Figure 5. Sky-averaged upper limit on the strain amplitude h_0 as a function of GW frequency. The Bayesian upper limits are computed using a fixed-noise model (thick black (blue)) and a varying noise model (thin black (purple)) and the frequentist upper limit (gray (red)) is computed using the $\mathcal{F}_p$ statistic. The dashed curves indicate lines of constant chirp mass for a source with a distance to the Virgo cluster of 16.5 Mpc and chirp mass of $10^9\ M_\odot$ (lower) and $10^{10}\ M_\odot$ (upper). The gray (green) squares show the strain amplitude of the loudest GW sources in 1000 Monte Carlo realizations using an optimistic phenomenological model of Sesana (2013). See text for more details.

(A color version of this figure is available in the online journal.)

of prior on h_0 is very important and can lead to very different upper limits. Such a detailed analysis of priors is beyond the scope of this work but will be addressed in a future paper. In principle, our prior distribution should come from population synthesis models (Sesana 2013); however, since we wish our upper limits to be informed by our data and not dominated by our prior distribution we use a very conservative[39] prior that is uniform in h_0 with $h_0 \in [0, 10^{-11}]$.

5.3.3. Sky-averaged Strain Upper Limits

In Figure 5 we report the 95% upper limits on the strain amplitude, h_0, as a function of GW frequency computed using the methods described above for the frequentist and Bayesian pipelines. The gray (red), thick black (blue), and thin black (purple) curves are the 95% upper limits on strain amplitude computed using the $\mathcal{F}_p$ statistic, Bayesian method with fixed-noise values, and Bayesian method with varying noise values, respectively. There are several features in the plot that require explanation. First, the decrease in sensitivity at $f_{\rm gw} = 1\ {\rm yr}^{-1}$ and $f_{\rm gw} = 2\ {\rm yr}^{-1}$ is due to the sky position and parallax fitting in the timing model, respectively. The upward trend at lower frequencies is due to the quadratic spin-down model fit. The noisiness of the frequentist upper limit is due to the fact that the $\mathcal{F}_p$ statistic distribution at higher frequencies is indeed quite noisy when computed using the real data, and since our upper limits compare the value measured in real data to values measured in simulated data, this noisiness is to be expected.

If we compare our results to those of Yardley et al. (2010), we see that the upper limits using the five-year NANOGrav data sets are a factor of two to three times more constraining. The main reason for this improvement is the higher timing precision of the NANOGrav data set as compared to the older PPTA data sets (Verbiest et al. 2009). Although the procedures for setting frequentist and Bayesian upper limits are quite different, our results are very similar. In part, this is due to the fact that we have used a uniform prior on the strain amplitude, h_0, making our Bayesian analysis very similar to a pure likelihood analysis. Since the $\mathcal{F}_p$ statistic is just the likelihood (ratio) maximized over amplitudes, we would expect a likelihood analysis to give similar results. Note that the Bayesian upper limits when varying the noise parameters are somewhat less constraining than the fixed-noise case. This is to be expected since at lower frequencies the GW amplitude and red noise amplitude are somewhat correlated and at higher frequencies the GW amplitude and jitter parameter are somewhat correlated. Both correlations will result in slightly worse upper limits on the GW amplitude when allowing the noise parameters to vary.

In Figure 5, the dashed curves indicate lines of constant chirp mass for a source with a distance to the Virgo cluster of 16.5 Mpc and chirp mass of 10^9 and 10^{10}, respectively, and the gray (green) squares are the strain amplitude of the loudest GW events in 1000 Monte Carlo realizations using an optimistic phenomenological model of Sesana (2013). The model used here produces a stochastic GW background with dimensionless strain amplitude of $\sim 2 \times 10^{-15}$, just below the current upper limits presented in Shannon et al. (2013). Astrophysically, these upper limits tell us that we can essentially rule out any source with $\mathcal{M} \geqslant 10^{10}\ M_\odot$ at the distance to the Virgo cluster (16.5 Mpc); however, our horizon distance falls just short of the Virgo cluster for sources with $\mathcal{M} \leqslant 10^9\ M_\odot$. This limit is not particularly constraining for the Virgo cluster, which generally does not contain galaxies with such large implied masses; however, it is interesting to assess this limit against other nearby massive galaxies (see, Lommen & Backer 2001). Moreover, at our most sensitive frequency of $\sim$10 nHz, we can convert our upper limit on the GW strain to a lower limit on the luminosity distance of $d_L \leqslant 425$ Mpc for sources with $\mathcal{M} = 10^{10}\ M_\odot$.

Finally, we see that all sources from Monte Carlo realizations (gray (green) squares) have strain amplitudes below our upper limits indicating that it is very unlikely that we will see a resolvable source at the current sensitivity (consistent with our search results). It is important to note, however, that these strain amplitude upper limits are averaged over sky location and inclination angle (either through marginalization in the Bayesian case, or from Monte Carlo sampling in the frequentist case), both of which play a large part in the *overall* amplitude of the signal. Therefore, these results have the caveat that they make statements about the *average* sensitivity to such GW sources; however, it is still unlikely (i.e., probability of detection $\lesssim$50%) that we could detect even the loudest optimally oriented source shown in Figure 5. For face-on systems (i.e., $\iota = \pi/2$) and sky location near the best timed pulsars, the overall amplitude of the GW can be $\sim$five times larger than the averaged strain amplitudes reported here.

5.3.4. Angular Upper Limits

In Figures 6 and 7 we report the 95% *lower* limit on the luminosity distance as a function of sky location computed using the frequentist and Bayesian techniques, respectively. We have chosen to present our results in terms of the luminosity distance instead of the strain amplitude as it is a true *physical* parameter and it gives a more intuitive feel as to what the data can constrain. To compute this lower limit we carry out the same procedure as above but we fix the frequency to $f_{\rm gw} = 10^{-8}$ Hz and compute an upper limit on the strain amplitude as a function of sky location; we can then use Equation (17) to convert an upper limit on

[39] On a logarithmic scale this prior prefers higher strain values a priori; however, it is conservative in the sense that the corresponding upper limit will not overestimate our sensitivity and the limit will not depend on the lower bound of the prior as is the case for logarithmic priors.

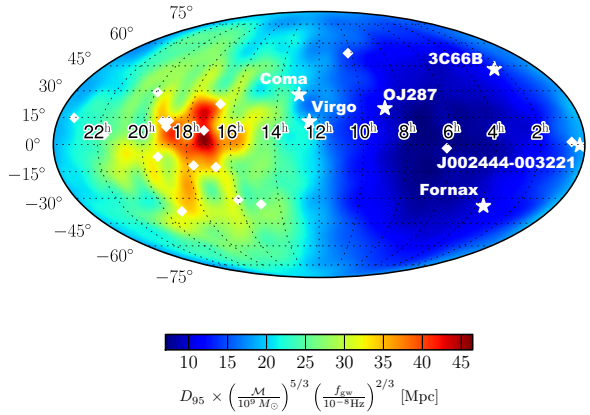

Figure 6. 95% lower limit on the luminosity distance as a function of sky location computed using the $\mathcal{F}_p$ statistic plotted in equatorial coordinates. The values in the colorbar are calculated assuming a chirp mass of $\mathcal{M} = 10^9\,M_\odot$ and a GW frequency $f_{gw} = 1 \times 10^{-8}$ Hz. The white diamonds denote the locations of the pulsars in the sky and the black (white) stars denote possible SMBHBs or clusters possibly containing SMBHBs. As expected from the antenna pattern functions of the pulsars, we are most sensitive to GWs from sky locations near the pulsars. The luminosity distances to the potential sources are 92.3, 1575.5, 2161.7, 16.5, 104.5, and 19 Mpc for 3C 66B, OJ 287, J002444−003221, Virgo Cluster, Coma Cluster, and Fornax Cluster, respectively.

(A color version of this figure is available in the online journal.)

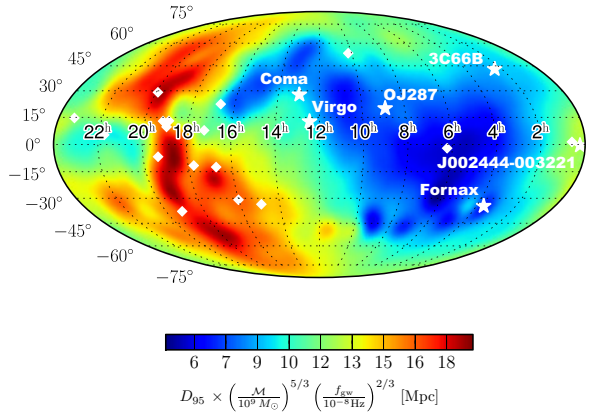

Figure 7. 95% lower limit on the luminosity distance as a function of sky location computed using the Bayesian method including the noise model. The values in the colorbar are calculated assuming a chirp mass of $\mathcal{M} = 10^9\,M_\odot$ and a GW frequency $f_{gw} = 1 \times 10^{-8}$ Hz. The white diamonds denote the locations of the pulsars in the sky and the black (white) stars denote possible SMBHBs or clusters possibly containing SMBHBs.

(A color version of this figure is available in the online journal.)

strain amplitude into a lower limit on luminosity distance. The values in the color bar are calculated assuming a chirp mass of $\mathcal{M} = 10^9\,M_\odot$ and a frequency of $f_{gw} = 10^{-8}$ Hz but this can be scaled to determine the minimum luminosity distance for any chirp mass value and GW frequency. In Figures 6 and 7 the white diamonds represent the locations of the 17 NANOGrav pulsars used in the analysis and the black (white) stars are the sky locations of potential GW hotspots (Simon et al. 2014) and possible GW source candidates (Valtonen et al. 2008; Iguchi et al. 2010; Ju et al. 2013).

We will now discuss the features of this sky-dependent upper limit computed using the frequentist $\mathcal{F}_p$ statistic. Firstly, we notice that the overall distribution is quite similar to the antenna pattern response (i.e., $1 + \cos\mu$) as is to be expected in the case of no detection. Due to this, we are most sensitive (larger lower limit on luminosity distance) at sky locations near the best timed pulsars (i.e., J1713+0747, B1855+09, J1909−3744) and least sensitive in the opposite direction. More quantitatively, we note that in the most sensitive areas of the sky we can constrain the luminosity distance $d_L \gtrsim 47$ Mpc for $\mathcal{M} = 10^9\,M_\odot$. Furthermore, it is possible to constrain the luminosity distance $d_L \gtrsim \sim 2$ Gpc in the most sensitive sky locations if we consider $10^{10}\,M_\odot$ chirp mass sources. It should be noted that the Bayesian fixed-noise search gives nearly identical results to the fixed-noise frequentist search.

We now move to the sky-dependent upper limit computed using the full Bayesian technique where the GW and noise parameters are varied simultaneously. The first observation that we make is that the overall scale is about a factor of two lower than the fixed-noise frequentist or Bayesian upper limit. At first this may be surprising given the general agreement of the sky-averaged upper limits of Figure 5; however, full Bayesian sky-dependent upper limits exacerbate the problem of relatively few pulsars contributing to the overall PTA sensitivity as shown in Figure 2. Another difference in this upper limit, as opposed to the frequentist upper limit, is that it does not quite match the expected antenna pattern response function. These differences are due to the fact that we are simultaneously varying the GW and noise parameters, and when only one or a few pulsars contribute to the PTA sensitivity, there is a degeneracy between intrinsic red-noise processes in the pulsar and a common GW among all pulsars. In other words, it is very difficult to distinguish between a low-frequency continuous GW and a red noise process if only a small number of pulsars have sufficiently low noise levels to resolve the GW.

Because Bayesian upper limits marginalize or integrate over all parameters except the amplitude, the correlations between the GW and the red noise amplitude will broaden the 1D PDF of the amplitude and thus will result in larger upper limits as opposed to the fixed-noise case. As is clear from Figure 7, the aforementioned effect is very strong for GW sky locations near our best timed pulsars. For example, we are not most sensitive to GWs around the sky location of PSR J1713 + 0747 because this pulsar contributes a very large percentage of the overall S/N of the GW in this case and thus results in a very large correlation between the GW and red noise amplitudes.

Since, at the moment, we have no way of measuring the noise properties of the pulsars independently of any GWs that may be present in the data, to perform a completely robust upper limit or search we must allow both to vary simultaneously. Given this reality, we must view any fixed-noise results with the caveat that they assume that the noise parameters are measured perfectly and are independent of any GWs in the data.

Unfortunately, many of the GW hotspots and potential SMBHB sources are located at insensitive sky locations, for both frequentist and Bayesian analyses, where our lower limit on distance only allows us to constrain $10^{10}\,M_\odot$ sources. This fact is a great argument for aggressive pulsar search campaigns and the addition of new pulsars to the PTA at sky locations that are currently insensitive (Burt et al. 2011).

5.3.5. Constraints on the SMBHB Coalescence Rate

A non-detection of continuous GW, as we have presented here, allows us to compute an upper limit on the rate of SMBH coalescences using methods presented in Wen et al. (2011). Since we have made no detections, we assume Poisson statistics for the probability of an event (i.e., a detectable signal) occurring, that is, the probability of no events is $e^{-\langle N\rangle}$, where $\langle N\rangle$ is the expected number of events. We use this probability

distribution function to place a 95% upper limit on the expected number of events such that $\exp(-N_{95}) = 0.05$, telling us that $\langle N \rangle \leqslant \langle N_{95} \rangle = 3$. Therefore, if the expected number of events were greater than 3, at least one source would have been detected with 95% probability. Now, following Wen et al. (2011), the expected number of events is

$$\langle N \rangle = \int \frac{d^2 R}{d\log_{10}(1+z) d\log_{10}(\mathcal{M}_r)} \left(\frac{df}{dt}\right)^{-1} \times P_d(\mathcal{M}_r, z, f)\, d\log_{10}(1+z)\, d\log_{10}(\mathcal{M}_r)\, df, \tag{49}$$

where $P_d(\mathcal{M}_r, z, f)$ is the probability of detecting an SMBHB with chirp mass $\mathcal{M} = \mathcal{M}_r(1+z)$, redshift z, and observed GW frequency f. Following the derivation in Wen et al. (2011) and making the assumption that the differential coalescence rate does not vary significantly over the range $\Delta \log_{10} \mathcal{M}_r = 1$ and $\Delta \log_{10}(1+z) = 0.2$, it is possible to show that

$$\frac{d^2 R}{d\log_{10}(1+z) d\log_{10}(\mathcal{M}_r)} < \frac{15}{\int \left(\frac{df}{dt}\right)^{-1} P_d(\mathcal{M}_r, z, f)\, df}. \tag{50}$$

In order to compute the detection probability P_d, we make use of the $\mathcal{F}_p$ statistic. We use the same method that we have used for the upper limits, except now we compare the value of $\mathcal{F}_p$ computed using simulated data with injections to a specified threshold based on an FAP of 10^{-4}. We use 10,000 realizations at each value of z and f. After the probability of detection is computed, we numerically integrate the above expression to obtain a limit on the differential coalescence rate. It should be noted that we will be able to place more meaningful constraints on the coalescence rate using upper limits on the amplitude of a stochastic background of SMBHBs; however, this is beyond the scope of this work and will be addressed in a future paper. In Figure 8 we plot our constraints on the differential coalescence rate as a function of redshift. Since we have made the assumption that this differential coalescence rate does not vary significantly over an order of magnitude in chirp mass, the results presented here are for the $\mathcal{M}_r = 10^{10}\, M_\odot$ case. We are unable to place meaningful constraints on less massive systems. The light gray (red) shaded area is constructed using the model presented in Jaffe & Backer (2003) along with measurements from the Sloan Digital Sky Survey (Wen et al. 2009). The medium gray (blue) shaded area is constructed by considering the different galaxy merger rates based on observations (Sesana 2013) along with the most recent MBH–sigma relation from McConnell & Ma (2013). The dashed line comes from an a posteriori implementation of the McConnell & Ma (2013) MBH–sigma relation into the semi-analytic model of Guo et al. (2011) assuming accretion onto both SMBHs before merger. The black (green) shaded region is constructed by using the observed evolution of the galaxy mass function combined with the MBH–M-stars relation from McConnell & Ma (2013) to calibrate an analytical model for evolving the mass function via mergers (McWilliams et al. 2012). The figure shows that the coalescence rate for MBHs of $\sim 10^{10}\, M_\odot$ is poorly constrained. This is mostly because of the steepness of the galaxy mass function at such high masses: a small change in the slope results in a large variation in the sparse population of $10^{10}\, M_\odot$ black holes. The intrinsic scattering (e.g., Gültekin et al. 2009) and poor knowledge of the behavior of the MBH–galaxy relations at the high mass end (e.g., Lauer et al. 2007) add further uncertainties, making the coalescence rate estimate problematic. As is clear from the figure, we are unable to place any constraints on the physical models mentioned above; however, as our GW sensitivity improves with time, we will begin to place meaningful constraints on physical models pertaining to the coalescence rate of SMBHBs.

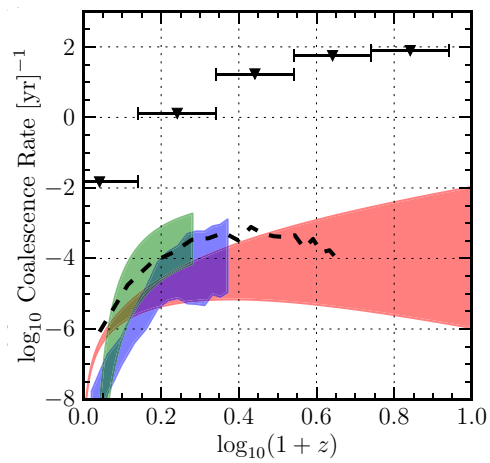

Figure 8. Differential coalescence rate of SMBHBs per redshift per chirp mass with mass bin centered on $10^{10}\, M_\odot$ and width 1 dex. We have chosen to explore only the highest masses since these high mass sources are the ones likely to be detected by GW searches in the future. The black triangles represent our upper 95% upper limits, the light gray (red) shaded area show expected coalescence rate estimates obtained from Jaffe & Backer (2003) as well as data from the Sloan Digital Sky Survey (Wen et al. 2009). The medium gray (blue) shaded region comes from the phenomenological models of Sesana (2013) and the black dashed line comes from an a posteriori implementation of the McConnell & Ma (2013) MBH–sigma relation into the semi-analytic model of Guo et al. (2011). The black (green) shaded region is constructed by using the observed evolution of the galaxy mass function combined with the MBH–M-stars relation from McConnell & Ma (2013) to calibrate an analytical model for evolving the mass function via mergers (McWilliams et al. 2012).

(A color version of this figure is available in the online journal.)

6. DISCUSSION

6.1. Future Improvements

Predicting the future sensitivity of PTAs to continuous GWs is quite difficult and depends on a number of poorly constrained factors that make up the entire noise budget for each pulsar (Cordes & Shannon 2010). A detailed study of these effects on the future performance of a PTAs sensitivity to continuous GWs is beyond the scope of this work. Here we simply derive a rough scaling law for the S/N of a continuous GW measured by a PTA and make a few statements about expected future performance. The square of the S/N is defined to be[40]

$$\rho^2 = \sum_\alpha (s_\alpha | s_\alpha) = \sum_\alpha 2\mathrm{Re} \int_{-\infty}^{\infty} df\, \frac{|\tilde{s}_\alpha(f)|^2}{S_\alpha^n(f)}, \tag{51}$$

where $\tilde{s}_\alpha(f)$ is the Fourier transform of the GW induced timing residuals and $S_\alpha^n(f)$ is the power spectral density of the noise for the αth pulsar. As we mentioned above, the frequency of the GW will not vary over the observation time and the waveform is approximated by a sine wave at a single frequency. We will further assume that the pulsar term is at the same frequency for

[40] We ignore timing model fitting here and make use of the Fourier domain for ease of computation.

this scaling law computation. The S/N then becomes

$$\rho^2 \approx \sum_\alpha 2\mathrm{Re} \int_{-\infty}^{\infty} df \frac{A(f)^2 a_\alpha^2 \delta(f-f')^2}{S_\alpha^{\mathrm{red}}(f) + \sigma_\alpha^2/c_\alpha}$$
$$\approx A(f')^2 \sum_\alpha \frac{T_\alpha}{2} \frac{a_\alpha^2}{S_\alpha^{\mathrm{red}}(f') + \sigma_\alpha^2/c_\alpha}, \quad (52)$$

where $\delta(f - f')$ is the Dirac delta function, $A(f')$ is the pulsar independent amplitude of the GW, a_α is a geometric factor that depends on the antenna pattern functions for each pulsar, c_α is the observing cadence for each pulsar, and $S_\alpha^{\mathrm{red}}(f')$ and σ_α are the red noise power spectral density at $f = f'$ and white noise rms, respectively. Our sensitivity to continuous GWs is proportional to the S/N, thus this expression for the S/N can serve as a proxy for how our upper limits and sensitivity will improve with various quantities. It is interesting to examine this scaling law in the white noise and red noise dominated regime

$$\rho_{\mathrm{white}} \propto \left(\sum_\alpha \frac{T_\alpha a_\alpha^2 c_\alpha}{\sigma_\alpha^2} \right)^{1/2} \quad (53)$$

$$\rho_{\mathrm{red}} \propto \left(\sum_\alpha \frac{T_\alpha a_\alpha^2}{S_\alpha^{\mathrm{red}}(f')} \right)^{1/2}. \quad (54)$$

The above scaling laws tell us that many pulsars distributed across the sky with high timing precision, high observing cadences, and low red noise levels observed over a long baseline will result in the best possible sensitivity to continuous GWs. New pulsar timing backends at Arecibo and the GBT (DuPlain et al. 2008) give roughly a factor of two higher timing precision for many pulsars, which will translate into an expected upper limit on the amplitude of continuous GWs that is a factor of two more constraining. As we acquire more data on our currently timed and newly discovered pulsars we will gain more sensitivity and will be able to probe to lower frequencies. Access to IPTA data would essentially serve to increase the observing cadence, and thus our sensitivity, since we would have complementary data from many different observatories measured at different times. Furthermore, current pulsar search campaigns (Lynch & Green Bank North Celestial Cap Survey Collaborations 2013) are discovering new MSPs in our least sensitive sky locations (see Figures 6 and 7), which will dramatically increase our overall sky coverage and will allow for better distinction between GW and noise models. Finally, advanced detectors, such as the Square Kilometer Array (SKA; Lazio 2013) are expected to time tens of pulsars at or below the 100 nanosecond level, which will likely solve many of the current problems that we face with poor angular sensitivity and inability to distinguish between single GW source and intrinsic pulsar noise.

In the red noise dominated regime, the cadence of observations and the overall white noise level is negligible and we essentially only gain sensitivity through the addition of new pulsars and continued timing. Here we note again that there is very little evidence for red noise in the five-year NANOGrav data set but this may change in the future as we become more sensitive to the stochastic GW background which would induce a common red noise signal in all pulsars. Of course, red noise from a stochastic background of GWs or from intrinsic pulsar spin noise (Shannon & Cordes 2010) is likely to have a steep spectrum and this red-noise dominated regime would only apply at the lowest frequencies.

For many pulsars in the five-year data set we are in the white noise dominated regime and since the S/N consists of the sum of the inverses of the white noise rms values, we see that only the best timed pulsars will contribute (as we have seen throughout this paper) and one may argue that we should focus all observing time on the best pulsars. However, as was shown in Siemens et al. (2013), our sensitivity to the stochastic background has a significantly weaker dependence on observing cadence and white noise rms but has a *linear* dependence on the number of pulsars in the array. Thus, it is difficult to realistically optimize a PTA for both continuous and stochastic GW sources.

6.2. Summary

In this paper we have performed various searches for continuous GWs from non-spinning SMBHBs in circular orbits using both frequentist and Bayesian techniques. Specifically, we have run a fixed-noise frequentist and Bayesian pipeline, as well as a varying noise Bayesian pipeline. In the absence of any detections we have placed upper limits on the strain amplitude of continuous GWs as a function of GW frequency. We have also computed a lower limit on the distance to such SMBHBs as a function of sky location, as well as placing constraints on the differential coalescence rate of such SMBHBs. Our sky-averaged upper limits on strain amplitude as a function of frequency are a factor of ∼three times more constraining than the previously published upper limits (Yardley et al. 2010) and we see good agreement between all three data analysis methods. Although improving, our limits still lie well above the amplitudes of individual sources produced from several realizations of an optimistic SMBHB population. We have shown that with good estimates of the intrinsic noise we can rule out any sources with luminosity distance <425 Mpc and a chirp mass of $\sim 10^{10}\, M_\odot$. Furthermore, this limit becomes about a factor of four more constraining for GW sources that are near our most precisely timed pulsars in the sky. Unfortunately we are not yet able to place any constraints on predictions for the coalescence rate of SMBHBs obtained from both theory and observations.

Throughout the paper we have made several statements about what is needed for completely *robust* data analysis techniques and what will be required from future PTAs in order to secure a *confident* detection of a continuous GW. These statements can be summarized as follows.

1. Currently we have no way to confidently separate intrinsic noise in the residuals from any GW that may be present. Therefore, it is necessary to include both noise and GW parameters in any data analysis pipeline that aims to be truly robust. This is not to say that fixed-noise methods should not be used; instead we advocate a hierarchical approach where the faster fixed-noise methods are used as a first pass and then followed up with a full GW plus noise search. Lastly, a signal with more information, such as that from an eccentric system, could help break this degeneracy between signal and noise models and will be the subject of a future paper.
2. Even with simultaneous noise and GW characterization, unless we have several well-timed pulsars (with very similar timing precision on all) with decent sky coverage, a *confident* detection of a continuous GW is unlikely even if the signal is loud.

In summary, we note that while not as likely as a detection of a stochastic GW background, with continually improving timing precision, the addition of new pulsars to PTAs, and improved data analysis techniques, prospects are good for obtaining astrophysically constraining GW limits, or possibly even a detection of a continuous GW, over the next decade.

The authors thank Neil Cornish, Jolien Creighton, and Stephen Taylor for many useful discussions regarding the data analysis methods presented in this work. The work of Z.A., A.B., S.B.-S., S.J.C., S.C., J.M.C., P.B.D., T.D., J.A.E., N.G.-D., F.J., G.J., M.T.L., T.J.W.L., A.N.L., D.R.L., J.L., D.R.M., M.A.M., D.J.N., N.P., T.T.P., S.M.R., X.S., D.R.S., K.S., J.S., and Y.W. was partially supported through the National Science Foundation (NSF) PIRE program award number 0968296. All computational work was performed on the Nemo cluster at UWM supported by NSF grant No. 0923409. NANOGrav research at UBC is supported by an NSERC Discovery Grant and Discovery Accelerator Supplement and by the Canadian Institute for Advanced Research. X.S. and J.E. were partially funded through an NSF CAREER award number 0955929 and through NSF grant No. PHY-1307429. J.E. was partially funded through the Wisconsin Space Grant Consortium. Portions of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Data for this project were collected using the facilities of the National Radio Astronomy Observatory and the Arecibo Observatory. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. The Arecibo Observatory is operated by SRI International under a cooperative agreement with the NSF (AST-1100968), and in alliance with Ana G. Méndez-Universidad Metropolitana, and the Universities Space Research Association. R.vH. is supported by NASA Einstein Fellowship grant PF3-140116.

APPENDIX A

GRAVITATIONAL WAVE FREQUENCY EVOLUTION

Since GWs will radiate power away from a SMBHB source, to compensate for this loss of energy the orbital separation must decrease with time. Equivalently, though Kepler's third law; GW radiation will cause the orbital frequency to *increase* with time. By setting the power radiated in GWs equal to the change of orbital energy due to increasing orbital frequency, $-dE_{\rm orbit}/dt$, we obtain

$$\dot{\omega} = \frac{96}{5}\mathcal{M}^{5/3}\omega^{11/3}, \tag{A1}$$

where $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1+m_2)^{1/5}$ is the chirp mass and ω is the orbital frequency of the binary system (note $\omega_{\rm gw} = 2\omega$). We can now use this expression to analytically solve for the orbital frequency as a function of time

$$\begin{aligned}\int_{t_0}^{t} dt &= \frac{5}{96}\mathcal{M}^{-5/3}\int_{\omega(t=t_0)}^{\omega(t)} d\omega\,\omega^{-11/3}\\ t - t_0 &= \frac{5}{256}\mathcal{M}^{-5/3}\left(\omega_0^{-8/3} - \omega(t)^{-8/3}\right)\\ \therefore \omega(t) &= \omega_0\left(1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}(t-t_0)\right)^{-3/8},\end{aligned} \tag{A2}$$

where t_0 is the time at which the first measurement was made on Earth and $\omega_0 = \omega(t = t_0)$ is the initial orbital frequency. For a circular orbit, we define the phase to be

$$\frac{d\Phi}{dt} = \omega. \tag{A3}$$

We can solve this equation similarly

$$\begin{aligned}\int_{\Phi(t=t_0)}^{\Phi(t)} d\Phi &= \int_{t=t_0}^{t} dt'\omega(t')\\ \Phi(t) - \Phi_0 &= \int_{\omega(t=t_0)}^{\omega(t)} d\omega\,\frac{\omega}{\dot{\omega}}\\ &= \frac{5}{96}\mathcal{M}^{-5/3}\int_{\omega(t=t_0)}^{\omega(t)} d\omega\,\omega^{-8/3}\\ \therefore \Phi(t) &= \Phi_0 + \frac{1}{32\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega(t)^{-5/3}\right),\end{aligned} \tag{A4}$$

where, again, $\Phi_0 = \Phi(t = t_0)$.

Equations (A2) and (A4) are true in general and can be applied when the frequency evolves appreciably over the total observing time. However, it is very useful to work under the assumption of slowly evolving binaries where $T_{\rm chirp} \gg T$, with T the observing time and

$$\begin{aligned}T_{\rm chirp} &= \frac{\omega_0}{\dot{\omega}}\\ &= 3.2\times 10^5\,{\rm yr}\left(\frac{\mathcal{M}}{10^8\,M_\odot}\right)^{-5/3}\left(\frac{f_0}{1\times 10^{-8}\,{\rm Hz}}\right)^{-8/3}.\end{aligned} \tag{A5}$$

Since typical PTA observations are on the order of 10–20 yr and $T/T_{\rm chirp} \sim 10^{-4}$, this is a safe assumption for a broad range of masses and initial orbital frequencies of interest. With this approximation we can write the orbital frequency and phase for the Earth term simply as

$$\Phi_e(t) = \Phi_0 + \omega_0(t - t_0) \tag{A6}$$

$$\omega_e(t) = \omega_0. \tag{A7}$$

However, for the pulsar term we are "seeing" the phase and frequency at a retarded time $t_p = t - L(1-\cos\mu)$, where L is the pulsar distance and μ is the angle between the GW and the pulsar on the sky. Because pulsar distances are on the order of a few kiloparsecs, this means that the total time baseline is on the order of thousands of years and we would expect frequency evolution over those timescales. However, just because the pulsar "sees" a different frequency than the Earth, this does not mean that the frequency at the pulsar changes over the observation time. For this reason we can write the phase and frequency at the pulsar in a similar manner

$$\Phi_p(t) = \Phi_{p,0} + \omega_p t \tag{A8}$$

$$\omega_p(t) = \omega_p. \tag{A9}$$

We can determine the "pulsar-term frequency" by evaluating Equation (A2) and setting $t = t_p$

$$\begin{aligned}\omega_p(t) &= \omega_0\left(1 - \frac{256}{5}\mathcal{M}^{5/3}\omega_0^{8/3}(t_p - t_0)\right)^{-3/8}\\ &= \omega_0\left(1 + \frac{8}{3}\frac{\dot{\omega}_0}{\omega_0}L(1-\cos\mu) + \frac{8}{3}\frac{\dot{\omega}_0}{\omega_0}(t - t_0)\right)^{-3/8}\\ &\approx \omega_0\left(1 + \frac{8}{3}\frac{\dot{\omega}_0}{\omega_0}L(1-\cos\mu)\right)^{-3/8} \equiv \omega_p,\end{aligned} \tag{A10}$$

In the above, we can safely ignore the last term in the second line with the reasoning that the frequency does not evolve over the observation time. Notice that the pulsar term frequency is always less than the Earth term frequency as we are observing the dynamics of the SMBHB in the past when the orbital separation was larger. Determining the pulsar-term phase in this approximation is a bit trickier. Re-writing Equation (A4), we get

$$\begin{aligned}\int_{\Phi(t_{p,0})}^{\Phi(t_p)} d\Phi &= \int_{t_{p,0}}^{t_p} dt'\omega(t')\\ \Phi_p(t) - \Phi_{p,0} &= \omega_p(t - L(1-\cos\mu)) + \omega_p L(1-\cos\mu)\\ \therefore \Phi_p(t) &= \Phi_{p,0} + \omega_p t, \end{aligned} \tag{A11}$$

where we have used the fact that $\omega(t) = \omega_p$ in the region of integration and we have adopted a notation in which $\Phi_p(t) \equiv \Phi(t_p)$. To determine the initial phase at $t = -L(1-\cos\mu)$ we use Equation (A4) to obtain

$$\begin{aligned}\Phi_{p,0} &= \Phi(t = -L(1-\cos\mu))\\ &= \Phi_0 + \frac{1}{32\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega_p^{-5/3}\right)\end{aligned} \tag{A12}$$

Although the above expressions for $\Phi_e(t)$ and $\Phi_p(t)$ are approximations, they hold true for nearly all values of $\mathcal{M}$ and ω_0 that we would expect in nature.

APPENDIX B

ALTERNATIVE $\mathcal{F}_p$ STATISTIC DERIVATION

The $\mathcal{F}_p$ statistic was introduced in Ellis et al. (2012) as a continuous GW detection statistic. However, when applied to single pulsars (as opposed to the full PTA) it essentially acts as a noise-weighted periodogram. While a specific notation was introduced in the original work we will use a consistent notation here to avoid confusion. In the same manner as above, let the pulsar timing residuals be denoted as

$$\delta t = M\delta\boldsymbol{\xi} + n + F\mathbf{a}, \tag{B1}$$

where now n encapsulates all noise in the data (parameterized by some parameters $\boldsymbol{\phi}$) and $F\mathbf{a}$ is the Fourier decomposition of a single GW source and *not* the non-white noise components. Here we will assume a flat prior for the Fourier coefficients $\mathbf{a}$. We now write the log of the marginalized likelihood ratio

$$\begin{aligned}\ln\Lambda &= \ln\, p(\delta t|\mathbf{a},\boldsymbol{\phi}) - \ln\, p(\delta t|\boldsymbol{\phi})\\ &= \delta t^T\tilde{C}^{-1}F\mathbf{a} - \frac{1}{2}\mathbf{a}^T F^T\tilde{C}^{-1}F\mathbf{a},\end{aligned} \tag{B2}$$

where, in a similar manner as above, $\tilde{C}^{-1} = G(G^TCG)^{-1}G^T$ and C is the covariance matrix of the noise in the residuals. Finding the maximum likelihood values of $\mathbf{a}$ and plugging it into the likelihood ratio we arrive at the $\mathcal{F}_p$ statistic for a single pulsar:

$$\mathcal{F}_p = \frac{1}{2}\delta t^T\tilde{C}^{-1}F(F^T\tilde{C}^{-1}F)^{-1}F^T\tilde{C}^{-1}\delta t. \tag{B3}$$

Note that we are projecting the noise-weighted residuals onto a Fourier basis and then taking the square. Essentially this is a noise-weighted power spectrum and is identical to previous expressions for the $\mathcal{F}_p$ statistic.

APPENDIX C

MCMC IMPLEMENTATION DETAILS

Here we will go over the specific implementation of the PTMCMC algorithm. The algorithm is based on that presented in E13 but has been slightly modified to be more efficient and robust. Specifically we will detail the jump proposals used in this work and the setup of the parallel tempering chains and thermodynamic integration calculation.

C.1. Jump Proposals

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

C.1.1. Correlated Jumps

As described in E13, we use an Adaptive Metropolis (AM) scheme to make correlated jump proposals. In essence, this jump uses the previous points in the chain's history to construct a sample covariance matrix, C_n (Haario et al. 2001), which is then updated throughout the run. In practice we update this covariance matrix every 1000 iterations. Primarily the sample covariance matrix is multiplied by a scale parameter $s_d = 2.4^2/n_{\rm dim}$, which gives an optimal 25% acceptance rate in the case of Gaussian posterior distributions; however, we will occasionally make small jumps (scale by 0.01) or large jumps (scale by 10). This jump is used in ∼20% of our total jump cycle.

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

$$y = U^T x, \tag{C1}$$

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

$$\langle yy^T\rangle = U^T\langle xx^T\rangle U = U^TUSU^TU = S. \tag{C2}$$

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

$$y_{i+1}^j = y_i^j + 2.4\mathcal{N}\left(0, \sigma_s^j\right), \tag{C3}$$

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

$$x_{i+1} = Uy_{i+1}. \tag{C4}$$

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

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

$$x_{i+1} = x_i + s_{DE}(x_m - x_n), \tag{C5}$$

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

Finally, we use a special jump proposal for the chirp mass and luminosity distance. This jump makes use of the fact that there is likely to be a large correlation between chirp mass and luminosity distance for nearly non-evolving sources. If the frequency evolution is negligible then the waveform only has information about the combination $\mathcal{M}^{5/3}/d_L$, with a very weak mass dependence in the frequency derivative term. For this jump, we draw a random luminosity distance value from the prior and then resolve for what value of chirp mass will keep the combination $\mathcal{M}^{5/3}/d_L$ constant. This jump is used in ∼5% of our total jump cycle.

C.1.2. Uncorrelated Jumps

Although we use mostly correlated jump proposals, about 15% of our jumps consist of uncorrelated jumps. In many cases, these uncorrelated jump proposals are simple draws from the prior distribution. For prior draws, we have four different jump proposals. Since all pulsars will have a strong white noise component but a weak red noise component and likely no visible GW signal, we draw from the white noise, red noise, and GW prior distributions separately with different weights. Red noise and GW (including the pulsar distance) prior draws account for ∼12% of our total jump cycle. Finally, we also occasionally make white noise and full parameter space prior draws, which account for ∼3% of our total jump cycle. Although this is quite a large percentage of jumps that draw from the prior, it greatly improves mixing in our case when we have many search parameters with broad posterior distributions. It should also be noted that when doing injections, we reduce the amount of prior draws to about ∼5% of the total jump cycle.

C.1.3. Auxiliary Pulsar Mode Jump

In E13, we discussed the difficulty posed by including the pulsar distance as a search parameter, showing that a very small change to the pulsar distance (⩽1 pc) can result in a phase shift in the GW waveform of order 2π. In that work we sidestepped this problem by breaking the pulsar term into a "phase" term and an "evolution" term. The phase term corresponds to very small jumps in the pulsar distance that will change the constant phase of the pulsar term, whereas the evolution term corresponds to large jumps in the pulsar distance that will change the frequency evolution. We used separate parameters to jump in the phase and evolution. More explicitly, we introduce a pulsar-term phase for each pulsar that is used in the phase term and *also* include the pulsar distance that is only used in the evolution term. While this method allows for good mixing and acceptance rates, it adds an extra $N_{\rm psr}$ parameters to the search.

Here we will describe a new method that does not require any additional parameters. This jump technique is summarized as follows.

1. Perform initial jump (either correlated or uncorrelated as described above).
2. Construct pulsar-term phase of Equation (A12) using the new parameters. This phase is likely several radians from the pre-jump pulsar-term phase due to the pulsar distance jump.
3. We desire a small Gaussian jump in the pulsar initial phase. To accomplish this we will slightly modify the pulsar distance such that

$$\Phi_{p,0}(L^1 + \delta L) = \Phi^0_{p,0} + \delta\phi, \tag{C6}$$

where the 1 and 0 superscripts denote post and pre-jump values, respectively, δL is a small pulsar distance offset, and $\delta\phi$ is a small Gaussian phase jump. We can re-write the above expression

$$\Phi^1_{p,0} + \left.\frac{d\Phi_{p,0}}{dL}\right|_{L=L^1} \delta L = \Phi^0_{p,0} + \delta\phi, \tag{C7}$$

where $\Phi^1_{p,0} = \Phi_{p,0}(L^1)$ and we have simply used a Taylor expansion. Making use of Equation (A12) we solve for δL

$$\delta L = \frac{\Phi^1_{p,0} - \Phi^0_{p,0} + \delta\phi}{\omega_p(1 - \cos\mu^1)}. \tag{C8}$$

4. Now let $L_{\rm new} = L^1 + \delta L$.

Essentially what we have done is to turn a pulsar distance jump into a pulsar-term phase jump. So in essence we are not breaking detailed balance as we are simply using the pulsar distance as an *auxiliary* parameter and initial pulsar-term phase as the actual *search* parameter. This auxiliary jump is called after every jump proposal in the cycle to ensure reasonable acceptance rates.

C.2. Parallel Tempering and Evidence Evaluation

Here we will describe our parallel tempering and thermodynamic integration techniques used to calculate the Bayesian evidence. We want our algorithm to then quickly locate the global maxima in the parameter space. To accomplish this in

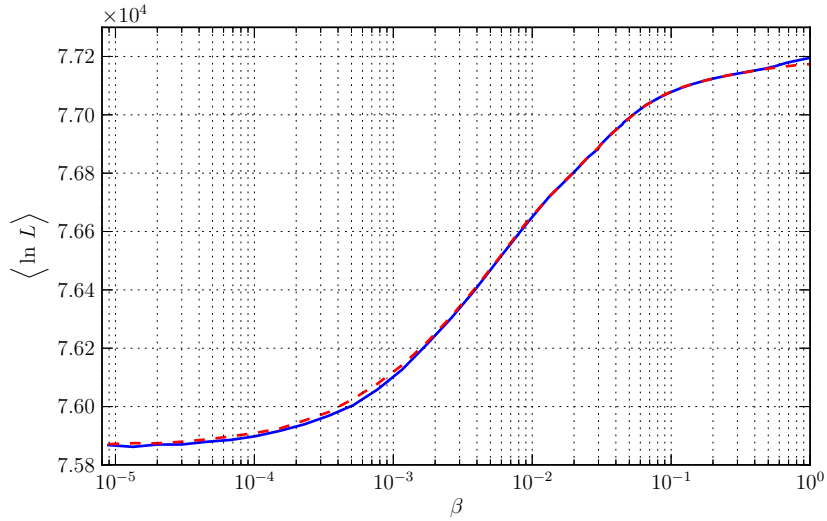

Figure 9. Mean log-likelihood vs. β for GW plus noise (gray (blue)) and noise (black (green)) models. Here we see that we have indeed explored a sufficient range of temperatures based on the fact that both curves become constant at small β.

(A color version of this figure is available in the online journal.)

a way that satisfies detailed balance we make use of parallel tempering. This technique involves different chains exploring the parameter space simultaneously, each with a different target distribution

$$p(\Theta|d, \beta) = p(\Theta)p(d|\Theta)^{\beta}, \tag{C9}$$

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

$$H_{i\to j} = \frac{p(d|\Theta_i, \beta_j)p(d|\Theta_j, \beta_i)}{p(d|\Theta_i, \beta_i)p(d|\Theta_j, \beta_j)}. \tag{C10}$$

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

$$\begin{aligned} Z(\beta) &= \int d\Theta\, p(d|\Theta, \mathcal{H}, \beta)p(\Theta|\mathcal{H}) \\ &= \int d\Theta\, p(d|\Theta, \mathcal{H})^{\beta}\, p(\Theta|\mathcal{H}). \end{aligned} \tag{C11}$$

Since the prior is independent of β, we can take the log and integrate over β to obtain

$$\ln\, p(d|\mathcal{H}) = \int_0^1 d\beta\, \langle \ln\, p(d|\Theta, \mathcal{H})\rangle_{\beta}, \tag{C12}$$

where $\langle \ln\, p(d|\Theta, \mathcal{H})\rangle_{\beta}$ is the expectation value of the likelihood for the chain with temperature $1/\beta$. The expectation values are calculated over the post burn-in chains.

In practice, it is important to choose a temperature ladder such that we explore the entire likelihood surface and recover the full integrand of Equation (C12). Here we will closely follow Littenberg & Cornish (2010) in the construction of our temperature ladder and diagnostic techniques. In constructing a temperature ladder to be used with thermodynamic integration it is important to understand that there are two regimes that we are interested in (at least in the GW detection problem). The first regime is the range of temperatures in which the (tempered) likelihood is still in "contact" with the GW, that is, the data still inform on the GW parameters. Since this is where the bulk of the integrand is concentrated when a signal is present, it is very important that we choose a fine temperature spacing here to resolve the point at which the likelihood loses contact with the GW. To do this we choose a geometrically spaced temperature ladder with temperature spacing

$$\Delta T = 1 + \sqrt{\frac{2}{n_{\rm dim}}}, \tag{C13}$$

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

$$T_{\rm max} = \left(\frac{\rho_{\rm max}}{\rho_{\rm eff,max}}\right)^2, \tag{C14}$$

where $\rho_{\rm eff,max}$ is the S/N at which we lose contact with the GW signal. For this work we have chosen $\rho_{\rm max} = 10$ and $\rho_{\rm eff,max} = 3$, corresponding to $T_{\rm max} = 11.1$. These values were chosen based on the fact that if we indeed have a signal with $\rho \approx 10$, it would have likely been detected in previous PTA data sets, and $\rho_{\rm eff,max}$ is chosen based on trial and error and simulations.

If we were only interested in parameter estimation then we would cut off the temperature ladder here; however, for evidence evaluation we must explore the full parameter space. This is the second temperature regime of evidence evaluation via thermodynamic integration. This is most important when

varying the noise parameters as well as the GW parameters. Here, we must choose an overall maximum temperature such that we are effectively sampling from the prior. In other words, the temperature must be sufficiently high such that the average log-likelihood has become constant with respect to increasing temperature. In this regime we choose a more coarse temperature spacing with $\Delta T = 1.5$ and geometric spacing. As noted in Littenberg & Cornish (2010) a good diagnostic to ensure that we are using a high enough temperature is to plot the mean log-likelihood for each temperature chain versus β.

In Figure 9 we plot the mean log-likelihood versus β for GW plus noise (gray (blue)) and noise (black (green)) models. First we notice that at low temperatures (high β) the GW plus noise model fits the data better based on the higher likelihood values (the data has an S/N 10 GW injection) but that it has slightly lower values at high temperature (low β) because of the expanded prior volume due to the GW parameter space. Since the Bayesian evidence is the area under these curves the question that is being answered by computing a Bayes factor is "does the fact that the GW plus noise model fit the data better (low temperature regime) overcome the fact that that model has a larger prior volume (high temperature regime)?" Because of this, it is crucial that we include temperatures high enough so that the average log likelihood becomes constant, indicating that we are sampling the prior distribution. For this work, we find that a maximum temperature of $\sim 10^5$ is sufficient.

REFERENCES

Babak, S., & Sesana, A. 2012, PhRvD, 85, 044034
Braak, C. 2006, Statistics and Computing, 16, 239
Burt, B. J., Lommen, A. N., & Finn, L. S. 2011, ApJ, 730, 17
Caprini, C., Durrer, R., & Siemens, X. 2010, PhRvD, D82, 063511
Champion, D. J., Hobbs, G. B., Manchester, R. N., et al. 2010, ApJL, 720, L201
Corbin, V., & Cornish, N. J. 2010, arXiv:1008.1782
Cordes, J. M., & Lazio, T. J. W. 2002, arXiv:astro-ph/0207156
Cordes, J. M., & Lazio, T. J. W. 2003, arXiv:astro-ph/0301598
Cordes, J. M., & Shannon, R. M. 2010, arXiv:1010.3785
Demorest, P. B. 2007, PhD thesis, Univ. of California, Berkeley
Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94
Detweiler, S. 1979, ApJ, 234, 1100
DuPlain, R., Ransom, S., Demorest, P., et al. 2008, Proc., SPIE, 7019
Ellis, J. A. 2013, CQGra, 30, 224004
Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012, ApJ, 756, 175
Ellis, J. A., Siemens, X., & van Haasteren, R. 2013, ApJ, 769, 63
Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300
Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, ApJ, 698, 198
Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101
Haario, H., Saksman, E., & Tamminen, J. 2001, Bernoulli, 7, 223
Haario, H., Saksman, E., & Tamminen, J. 2005, Computational Statistics, 20, 265
Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39
Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010, CQGra, 27, 084013
Hobbs, G., Coles, W., Manchester, R. N., et al. 2012, MNRAS, 427, 2780
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655
Hotan, A. W., van Straten, W., & Manchester, R. N. 2004, PASA, 21, 302
Iguchi, S., Okuda, T., & Sudou, H. 2010, ApJL, 724, L166
Jaffe, A. H., & Backer, D. C. 2003, ApJ, 583, 616
Janssen, G. H., Stappers, B. W., Kramer, M., et al. 2008, in AIP Conf. Ser. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, ed. C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi (Melville, NY: AIP), 633
Jenet, F., et al. 2009, arXiv:0909.1058
Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, ApJ, 653, 1571
Jenet, F. A., Lommen, A., Larson, S. L., & Wen, L. 2004, ApJ, 606, 799
Ju, W., Greene, J. E., Rafikov, R. R., Bickerton, S. J., & Badenes, C. 2013, ApJ, 777, 44
Kaspi, V. M., Taylor, J. H., & Ryba, M. F. 1994, ApJ, 428, 713
Lauer, T. R., Tremaine, S., Richstone, D., & Faber, S. M. 2007, ApJ, 670, 249
Lazio, T. J. W. 2013, CQGra, 30, 224011
Lentati, L., Alexander, P., Hobson, M. P., et al. 2013, PhRvD, 87, 104021
Littenberg, T. B., & Cornish, N. J. 2009, PhRvD, 80, 063007
Littenberg, T. B., & Cornish, N. J. 2010, PhRvD, 82, 103007
Lommen, A. N., & Backer, D. C. 2001, ApJ, 562, 297
Lynch, R. S., Green Bank North Celestial Cap Survey Collaborations. 2013, in IAU Symp. 291, The Hunt for New Pulsars with the Green Bank Telescope, ed. J. van Leeuwen (Cambridge: Cambridge Univ. Press), 41
Manchester, R. N. 2008, in AIP Conf. Ser. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, ed. C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi (Melville, NY: AIP), 584
Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993
McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184
McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2012, arXiv:1211.5377
Olmez, S., Mandic, V., & Siemens, X. 2010, PhRvD, 81, 104028
Perrodin, D., Jenet, F., Lommen, A., et al. 2013, arXiv:1311.3693
Petiteau, A., Babak, S., Sesana, A., & de Araújo, M. 2013, PhRvD, 87, 064036
Roedig, C., & Sesana, A. 2012, JPhCS, 363, 012035
Romani, R. 1989, in NATO ASI Series, Series C, Vol. 262, Timing Neutron Stars, ed. H. Ögelman & E. van den Heuvel (Dordrecht: Kluwer), 113
Sazhin, M. V. 1978, SvA, 22, 36
Sesana, A. 2013, MNRAS, 433, L1
Sesana, A., & Vecchio, A. 2010, PhRvD, 81, 104008
Sesana, A., Vecchio, A., & Colacino, C. N. 2008, MNRAS, 390, 192
Sesana, A., Vecchio, A., & Volonteri, M. 2009, MNRAS, 394, 2255
Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607
Shannon, R. M., Ravi, V., Coles, W. A., et al. 2013, Sci, 342, 334
Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, CQGra, 30, 224015
Simon, J., Polin, A., Lommen, A., et al. 2014, ApJ, 784, 60
Starobinsky, A. A. 1979, JETPL, 30, 682
Vallisneri, M. 2012, PhRvD, 86, 082001
Valtonen, M. J., Lehto, H. J., Nilsson, K., et al. 2008, Natur, 452, 851
van Haasteren, R. 2013, MNRAS, 429, 55
van Haasteren, R., & Levin, Y. 2010, MNRAS, 401, 2372
van Haasteren, R., & Levin, Y. 2013, MNRAS, 428, 1147
van Haasteren, R., Levin, Y., Janssen, G. H., et al. 2011, MNRAS, 414, 3117
van Haasteren, R., Levin, Y., McDonald, P., & Lu, T. 2009, MNRAS, 395, 1005
Verbiest, J. P. W., Bailes, M., Coles, W. A., et al. 2009, MNRAS, 400, 951
Verbiest, J. P. W., Weisberg, J. M., Chael, A. A., Lee, K. J., & Lorimer, D. R. 2012, ApJ, 755, 39
Wahlquist, H. 1987, GReGr, 19, 1101
Wen, Z. L., Jenet, F. A., Yardley, D., Hobbs, G. B., & Manchester, R. N. 2011, ApJ, 730, 29
Wen, Z. L., Liu, F. S., & Han, J. L. 2009, ApJ, 692, 511
Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 590, 691
Yardley, D. R. B., Hobbs, G. B., Jenet, F. A., et al. 2010, MNRAS, 407, 669