# On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array

Boris Goncharov[1,2,3] , R. M. Shannon[1,2] , D. J. Reardon[1,2] , G. Hobbs[4], A. Zic[4,5] , M. Bailes[1,2] , M. Curyło[6], S. Dai[4,7] , M. Kerr[8] , M. E. Lower[1,4] , R. N. Manchester[4] , R. Mandow[1,4], H. Middleton[1,2,9] , M. T. Miles[1,2], A. Parthasarathy[10] , E. Thrane[2,11] , N. Thyagarajan[4] , X. Xue[12,13], X.-J. Zhu[2,11] , A. D. Cameron[1,2] , Y. Feng[14] , R. Luo[4] , C. J. Russell[15] , J. Sarkissian[16], R. Spiewak[1,17] , S. Wang[4,18] , J. B. Wang[18] , L. Zhang[19] , and S. Zhang[20]

[1] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia; boris.goncharov@me.com
[2] OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Australia
[3] Gran Sasso Science Institute (GSSI), I-67100 L'Aquila, Italy
[4] Australia Telescope National Facility, CSIRO, Space and Astronomy, P.O. Box 76, Epping, NSW 1710, Australia
[5] Department of Physics and Astronomy, and Research Centre in Astronomy, Astrophysics and Astrophotonics, Macquarie University, NSW 2109, Australia
[6] Astronomical Observatory, University of Warsaw, Al. Ujazdowskie 4, 00-478 Warsaw, Poland
[7] Western Sydney University, Locked Bag 1797, Penrith South DC, NSW 1797, Australia
[8] Space Science Division, Naval Research Laboratory, Washington, DC 20375-5352, USA
[9] School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
[10] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany
[11] School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia
[12] CAS Key Laboratory of Theoretical Physics, Insitute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
[13] School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[14] CAS Key Laboratory of FAST, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
[15] CSIRO Scientific Computing, Australian Technology Park, Locked Bag 9013, Alexandria, NSW 1435, Australia
[16] CSIRO Space and Astronomy, Australia Telescope National Facility, P.O. Box 276, Parkes, NSW 2870 Australia
[17] Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK
[18] Xinjiang Astronomical Observatory, Chinese Academy of Sciences, 150 Science 1-Street, Urumqi, Xinjiang 830011, People's Republic of China
[19] National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing 100101, People's Republic Of China
[20] Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, People's Republic Of China

## Abstract

A nanohertz-frequency stochastic gravitational-wave background can potentially be detected through the precise timing of an array of millisecond pulsars. This background produces low-frequency noise in the pulse arrival times that would have a characteristic spectrum common to all pulsars and a well-defined spatial correlation. Recently the North American Nanohertz Observatory for Gravitational Waves collaboration (NANOGrav) found evidence for the common-spectrum component in their 12.5 yr data set. Here we report on a search for the background using the second data release of the Parkes Pulsar Timing Array. If we are forced to choose between the two NANOGrav models—one with a common-spectrum process and one without—we find strong support for the common-spectrum process. However, in this paper, we consider the possibility that the analysis suffers from model misspecification. In particular, we present simulated data sets that contain noise with distinctive spectra but show strong evidence for a common-spectrum process under the standard assumptions. The Parkes data show no significant evidence for, or against, the spatially correlated Hellings–Downs signature of the gravitational-wave background. Assuming we did observe the process underlying the spatially uncorrelated component of the background, we infer its amplitude to be $A = 2.2^{+0.4}_{-0.3} \times 10^{-15}$ in units of gravitational-wave strain at a frequency of $1\,\mathrm{yr}^{-1}$. Extensions and combinations of existing and new data sets will improve the prospects of identifying spatial correlations that are necessary to claim a detection of the gravitational-wave background.

*Unified Astronomy Thesaurus concepts:* Gravitational waves (678); Gravitational wave astronomy (675); Millisecond pulsars (1062); Pulsar timing method (1305); Bayesian statistics (1900)

## 1. Introduction

While detections of gravitational waves (e.g., Abbott et al. 2016) have been made with ground-based interferometers that are sensitive to hertz–kilohertz gravitational waves, experiments that operate at lower frequencies have yet to identify a signal. Pulsar timing array (PTA) experiments, which monitor and measure arrival times from millisecond pulsars (MSPs), have been established to search for signals in the nanohertz band. This is the domain of the stochastic background from supermassive black hole binaries, which is expected to be the first gravitational-wave signal detected with PTAs (Rosado et al. 2015). The background manifests as a temporally and spatially correlated process in the MSP arrival times. The strain

spectrum of such a background is predicted to have the power-law form $h(f) = A(f/1\,\mathrm{yr}^{-1})^{-2/3}$, where $f$ is the gravitational-wave frequency and $A$ is the strain amplitude at $f = 1\,\mathrm{yr}^{-1}$ (Phinney 2001). The amplitude, $A$, will depend on the demographics of the supermassive black hole population. Astrophysical effects relating to supermassive binary black hole evolution may cause deviations from a power law (e.g., Ravi et al. 2014; Sampson et al. 2015; Taylor et al. 2017b).

A definitive detection of the gravitational-wave background is the presence of specific spatial correlations in the arrival times (Hellings & Downs 1983). Other processes can produce signals with similar temporal properties, with either no (Shannon & Cordes 2010) or different spatial correlation. Tiburzi et al. (2016) and Taylor et al. (2017a) showed the

challenges in distinguishing between sources that produce different spatial correlations.

Previous searches for the gravitational-wave background from supermassive black hole binaries have reported limits on the strain amplitude, $A$, ranging between $1.1 \times 10^{-14}$ and $1.0 \times 10^{-15}$ at 95% of either confidence or credibility, where appropriate (Jenet et al. 2006; van Haasteren et al. 2011; Demorest et al. 2013; Shannon et al. 2013; Lentati et al. 2015; NANOGrav Collaboration et al. 2015; Shannon et al. 2015; Arzoumanian et al. 2018). The limits are now known to be affected by systematic uncertainties in the ephemeris of the solar system, which impacts pulsar timing because the arrival times are necessarily referenced to an inertial frame located at solar system barycenter (e.g., Arzoumanian et al. 2018).

In a recent analysis of the NANOGrav 12.5 yr data set, a common-noise process was reported having a Bayes factor greater than $10^4$ (this corresponds to 9 on the commonly used natural logarithmic scale), with the signal persisting even when accounting for uncertainties in the solar system ephemeris (Arzoumanian et al. 2020). We discuss the meaning of the term "common process" in detail later. Here we define the symbol CP to represent the common process as obtained by the hypotheses used in the NANOGrav analysis. Evidence for Hellings–Downs correlations was insignificant.

We have carried out a similar analysis using the Parkes Pulsar Timing Array (PPTA; Manchester et al. 2013) second data release (Kerr et al. 2020). The observations and methodology are described in Section 2. In Section 3 we discuss the results of the searches. In Section 4, we discuss limitations in the methodology. In particular, we demonstrate through simulation that the search methods can spuriously detect a common red process in timing array data sets in which it is absent.

## 2. The Data Set and Methodology

The PPTA project monitors an ensemble of MSPs with the 64 m Parkes radio telescope (also named *Murriyang*) in New South Wales, Australia. The data used, namely pulse arrival times from the observations, were acquired between 2003 and 2018 and were published as part of the second data release of the project (PPTA-DR2; Kerr et al. 2020). Observations were taken at a cadence of approximately three weeks. At each epoch, data were usually recorded in bands centered at three different radio frequencies in order to correct variations in pulsar dispersion measures (Keith et al. 2013). Data were recorded with a series of digital processing systems, with quality having improved over the course of the project.

We analyzed the data set using methodology that was based on that applied to the NANOGrav 12.5 yr data set (Arzoumanian et al. 2020), which itself was based on NANOGrav Collaboration et al. (2015) and Taylor et al. (2017a). Stochastic signals were modeled as being correlated (red) or uncorrelated (white) in time. We had previously characterized the noise processes for individual pulsars in the PPTA sample (Goncharov et al. 2021). That analysis showed that the PPTA data sets contain a wide variety of noise processes, including instrument-dependent or band-dependent processes. In this work we included red-noise processes in all pulsars, even for those pulsars that showed no evidence for such noise in previous analyses.

As in Goncharov et al. (2021), we assume that the power spectral density of all red processes follows a power law,

parameterized such that the amplitude, $A$, is in units of gravitational-wave strain at $1 \, \text{yr}^{-1}$:

$$P(f|A, \gamma) = \Gamma(\zeta_{ab}) \frac{A^2}{12\pi^2} \left( \frac{f}{\text{yr}^{-1}} \right)^{-\gamma} \text{yr}^3. \quad (1)$$

The fluctuation frequency of the pulse arrival time power spectrum is denoted $f$ and the spectral index is $-\gamma$. The noise terms are modeled using a Fourier series, starting with a fundamental frequency that is the inverse of the observation span corresponding to the entire pulsar data set, $T_{\text{obs}}$. We use $n_c = 30$ harmonics if $\gamma > 1.5$ (Goncharov et al. 2020), otherwise we use $n_c$ from the single-pulsar analysis (Goncharov et al. 2021). The overlap reduction function ($\Gamma(\zeta_{ab})$; Finn et al. 2009) characterizes the spatial correlation of the signal, and depends on the angular separation, $\zeta_{ab}$, of two pulsars $a$ and $b$ with respect to the observer. For an isotropic stochastic background from the gravitational waves of general relativity (Hellings & Downs 1983; Jenet & Romano 2015),

$$\Gamma_{\text{GWB}}(\zeta_{ab}) = \frac{1}{2} - \frac{1}{4} \left( \frac{1 - \cos \zeta_{ab}}{2} \right)$$
$$+ \frac{3}{2} \left( \frac{1 - \cos \zeta_{ab}}{2} \right) \ln \left( \frac{1 - \cos \zeta_{ab}}{2} \right), \quad (2)$$

when $a \neq b$.

Our analysis proceeded through these steps:

1. We first searched for a common power-law, red-spectrum stochastic process (CP1), with an identical power spectrum and unrelated temporal evolution or spatial correlation across pulsars. We emphasize that the timing spectrum of the process is assumed to be a statistically identical ensemble-average power spectrum among pulsars, which would be the case for a gravitational-wave background.[21] Throughout our analysis we marginalized over deterministic terms in the timing model (Reardon et al. 2021). This included instrument-dependent offsets ("jumps") of unknown value, as identified by Kerr et al. (2020). Offsets with a priori measured values were held fixed. We trialed both marginalizing over the white-noise parameters and also holding them fixed at their maximum a posteriori values. We obtained consistent results between these approaches.

2. Following the NANOGrav analysis we also assumed a power-law model with $\gamma = 13/3$. This value has astrophysical interest as it is the expected value for a gravitational-wave background caused by supermassive binary black holes (Phinney 2001). The resulting common power-law, red-noise process is here labeled as CP2. Based on a factorized-likelihood approach, we performed a dropout analysis to evaluate the consistency of individual PPTA DR2 pulsars with the signal identified by CP2 (see Arzoumanian et al. 2020).

3. We measured the amplitude of individual Fourier components $P_i(f_i|\rho_i) = \rho_i^2 \, T_{\text{obs}}$ of a common process, at each frequency $f_i$ separately in order to determine whether the power-law assumption of CP1,2 is valid.

---

[21] This term was first introduced in Arzoumanian et al. (2018), and we refer the reader to that paper for further discussion of its meaning.

4. We searched for evidence that CP2 exhibits spatial correlations from a gravitational-wave background, a monopolar signal, MP, or a dipolar signal, DP. In this analysis we held the white-noise stochastic components fixed at their maximum a posteriori value to reduce the number of parameters in the search and reduce computation time.

5. To assess the shape of any spatial correlations, we measured $\Gamma(\zeta_{ab})$ at seven equally separated "node" angles between $0°$ and $180°$ inclusive, using linear interpolation to determine $\Gamma(\zeta_{ab})$ from pulsar pairs between the nodes. The interpolant modeling of the PTA correlation curve was first done in Taylor et al. (2013).

### 2.1. Comparison with the NANOGrav Data Set and Processing Methods

The NANOGrav analysis included the timing data from 45 pulsars in their analysis of their 12.5 yr data set (Arzoumanian et al. 2020). The PPTA-DR2 analysis is based on the data from 26 pulsars spanning up to 15 yr.

While the data sets were obtained with different telescopes at a different range of frequencies, the two data sets have 11 pulsars in common, including some of the most precisely timed pulsars. The sources in common are PSRs J0613−0200, J1024 −0719, J1600−3053, J1643−1224, J1713+0747, J1832 −0836, J1857+0943, J1909−3744, J1939+2134, and J2145 −0750. Even for these pulsars we independently determined noise models. Our instrumental noise terms are necessarily independent from NANOGrav. However, we also included extra noise terms into the modeling for specific pulsars. In particular for PSR J1713+0747 the NANOGrav analysis included timing noise and dispersion-measure noise terms as well as the inclusion of two exponential dips attributed to rapid dispersion-measure variations. The PPTA analysis is similar. In the second exponential dip we allowed for a different chromaticity as there is evidence that it is not caused by dispersion-measure variations (Goncharov et al. 2021).

Both analyses made use of TEMPO2. Bayesian inference was performed with ENTERPRISE (Ellis et al. 2019). Preferred models were selected based on the Bayes factor calculated using a product-space sampling method (Carlin & Chib 1995; Taylor et al. 2020). We denote a Bayes factor for model A over model B as $\mathcal{B}_B^A$. The null model, with no common or correlated noise processes in the data set, is denoted $\varnothing$. We referenced pulse arrival times to the solar system barycenter using the ephemeris DE436, to maintain consistency with PPTA-DR2 (Kerr et al. 2020) and the single-pulsar analyses (Goncharov et al. 2021; Reardon et al. 2021). The more recent DE438 ephemeris was used in the NANOGrav analysis.

### 3. Results

Following the process described above, we obtained strong evidence for CP1, with $\ln \mathcal{B}_\varnothing^{\mathrm{CP1}} > 15.0$. This implies an odds ratio in favor of CP1 of $>3 \times 10^6$: 1, if both models had even odds a priori. The results from the parameter estimation are shown in the left panel in Figure 1. We obtain[22] $\log_{10} A_{\mathrm{CP1}} = -14.55^{+0.10}_{-0.23}$ and $\gamma = 4.11^{+0.52}_{-0.41}$. The results from the NANOGrav analysis are overlaid and have significant

overlap, although the NANOGrav analysis preferred a steeper spectral exponent, when the latter analysis was conducted with five Fourier components. Unlike in Arzoumanian et al. (2020), our measurements are consistent when we change the numbers of fluctuation frequencies used to model the common process.

The right panel in Figure 1 represents the parameter estimation for the free-spectral model. It is challenging to obtain a complete noise model for the brightest MSP, PSR J0437−4715 (Goncharov et al. 2021), and so we show this spectrum with, and without, the inclusion of this pulsar. We overlay the astrophysically interesting spectrum corresponding to $\gamma = 13/3$, along with the frequencies corresponding to the orbital periods of the planets.

In the left-hand panel of Figure 2, we show the posterior distribution for $\log_{10} A$, assuming a power-law model with $\gamma = 13/3$. The measured CP2 amplitude is $\log_{10} A = -14.66 \pm 0.07$ corresponding to $A = 2.2^{+0.4}_{-0.3} \times 10^{-15}$, which is consistent with that measured in the NANOGrav 12.5 yr data set.

The NANOGrav analysis attempted to determine which pulsars contributed to this signal by calculating dropout factors, which for our sample are displayed in the right-hand panel of Figure 2. The pulsars with the smallest dropout factors (PSRs J1824−2452A and J1939+2134) are known to have strong timing noise inconsistent in strength with that of other pulsars. However, the pulsars with the highest dropout factors include pulsars with high (PSR J0437−4715) and low timing precision (PSR J1022+1001), and pulsars with shorter timing baselines (PSR J2241−5236). The meaning and use of dropout factors is further discussed in Section 4.3.

The results from the model-independent parameter estimation of the overlap reduction function (obtained assuming $\gamma = 13/3$) are provided in Figure 3. The left-hand panel shows the inferred spatial correlations. They were sampled at seven node angles, whereas spatial correlations for other angles were obtained with linear interpolation. The Hellings–Downs relation is overplotted. The right-hand panel shows inferred amplitude of the Hellings–Downs spatially correlated noise and that for CP2. With the entire PPTA sample of pulsars the Bayes factor is $\ln \mathcal{B}_{\mathrm{CP2}}^{\mathrm{HD}} = 0.3$, which provides no significant evidence for, or against, Hellings-and-Downs correlations. We note that if PSR J0437−4715 is removed from the sample then the Bayes factor increases to $\ln \mathcal{B}_{\mathrm{CP2}}^{\mathrm{HD}} = 1.0$. The data strongly disfavor CP2 having monopole or dipole spatial correlations ($\ln \mathcal{B}_{\mathrm{CP2}}^{\mathrm{MP}}$ and $\ln \mathcal{B}_{\mathrm{CP2}}^{\mathrm{DP}}$ both $< -10$).

### 4. Discussion

Under the same assumptions the Bayesian analyses of both the PPTA and NANOGrav data sets show a preference for models that include a common-noise process in addition to individual noise terms. The CP1 model has consistent spectral index and amplitude between the data sets and therefore we can exclude this signal from being telescope dependent. Given the different strategies employed in mitigating interstellar propagation effects by the two projects (both in terms of choice of observing band and methods for correcting for dispersion-measure variations), it is also unlikely that the noise is associated with the interstellar medium.

However, we are attempting to detect a common-noise process from a single realization of the process in each pulsar. The noise process is strongest at the lowest fluctuation frequency, so the process is being characterized on a timescale

---

[22] Unless otherwise specified, throughout the paper uncertainties provide $1\sigma$ credible levels.
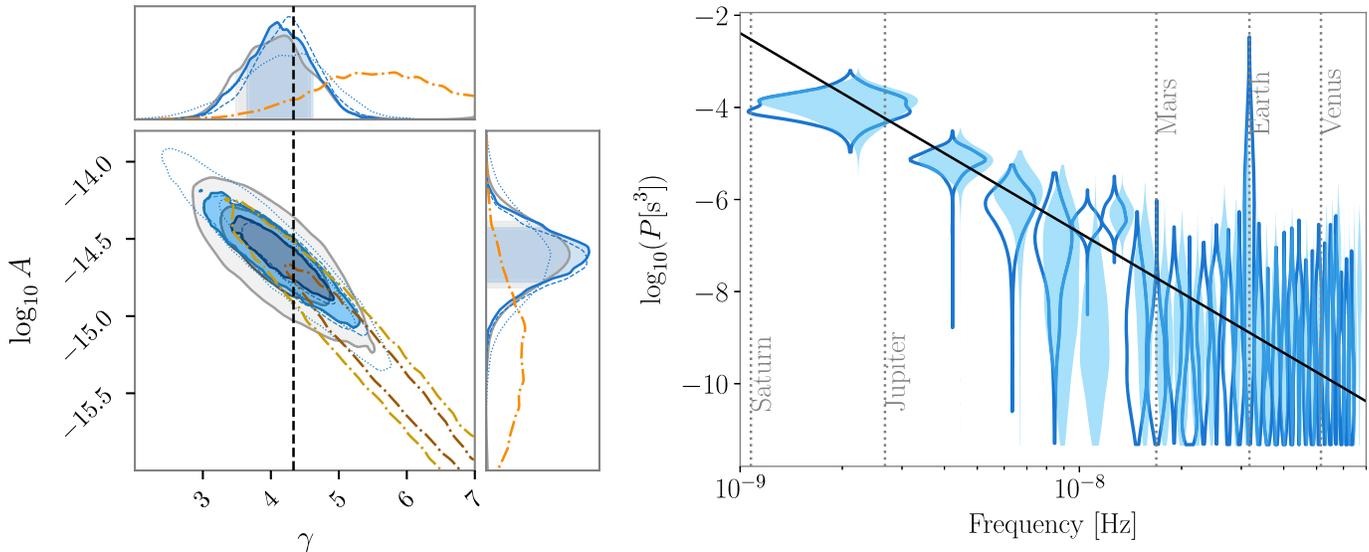
**Figure 1.** Left: measurements of common power-law red-noise parameters and the demonstration of their robustness to assumptions about pulsar-intrinsic noise and the number of fluctuation frequencies $n_c$. The dashed vertical line indicates $\gamma = 13/3$. The solid lines represent the measurement based on $n_c = 30$. Dashed and dotted lines represent $n_c = 20$ and $n_c = 5$. The dashed–dotted lines correspond to the measurement from Arzoumanian et al. (2020). Contours and shaded regions are $1\sigma$ and $2\sigma$ credible levels. Gray lines and regions are based on the assumption of achromatic timing noise in every pulsar, whereas blue ones are based on the assumption of timing noise only in pulsars where it was reported in Goncharov et al. (2021). Right: common red-noise parameter estimation with the free-spectral model. Lines represent the full PPTA data, whereas filled regions represent PPTA DR2 without PSR J0437−4715. The black line is the inferred spectrum assuming a power-law model with $\gamma = 13/3$. Vertical dotted lines represent inverse orbital periods of solar system planets.

comparable to the typical data span. This greatly complicates tests of the noise modeling. Consequently there are a number of caveats for interpreting the CP1 and CP2 results as discussed in the following subsections. We conclude by discussing our search for spatial correlations in the data.

### 4.1. Are the Models of the Intrinsic Noise Correct and Complete?

We have modeled the intrinsic noise to be a power-law process. Intrinsic timing noise for millisecond pulsars is not well studied over the relevant timescales. We know millisecond pulsars (PSRs J0613−0200 and J1824−2452A; Cognard & Backer 2004; McKee et al. 2016) have exhibited glitch events, so small glitches are possibly present in other pulsars. Two of the pulsars in our sample, PSRs J0437−4715 and J2241−5236, have reported evidence for excess nonstationary noise (Goncharov et al. 2021). Large-scale studies of non-millisecond pulsars have demonstrated that the amplitude of their timing noise is approximately determined by the pulsar spin-down rate (Parthasarathy et al. 2019), but there is also clear evidence for discrete changes in spin frequency or frequency derivative (Cordes & Helfand 1980), which may occur at quasi-periodic intervals (Hobbs et al. 2010). It is therefore unlikely that the intrinsic pulsar timing noise is perfectly modeled via a power-law process. We know that the pulse profiles of millisecond pulsars can show secular shape changes (Shannon et al. 2016), which, if unmitigated, result in a nonstationary noise process.

### 4.2. What Assumptions Lead to the Evidence of a Common Process?

The assessment of whether a common red-noise process is present is based on the null hypothesis that all pulsars are affected by independent red timing noise, modeled by a power-

law spectrum described by amplitudes and exponents drawn from a uniform prior. The detection hypothesis is that all pulsars also exhibit a red process with the same power-law spectrum, where amplitudes and exponents are drawn from the delta-function prior. There is a possibility that neither of these models nor their priors are sufficient descriptions of the data, which is often referred to as model misspecification (e.g., Davidson & MacKinnon 1981; Müller 2013). In simulations we demonstrate that noise without a statistically identical spectrum between pulsars can be misinterpreted as the common red process.

We simulated 10 realizations of timing residuals for the 26 PPTA-DR2 pulsars, with a range of realistic white noise levels, and injected power-law timing noise models with amplitudes and spectral indices drawn uniformly across approximately several orders of magnitude ($\log_{10} A$ spanning approximately $-16$ to $-13$ and $\gamma$ spanning 3 to 5). The power spectral densities of the simulated residuals are shown in the left-hand panel of Figure 4. We performed model selection for a model with a common-spectrum process along with intrinsic pulsar timing noise (CP2), against a model with intrinsic timing noise only ($\varnothing$), and obtained $\ln \mathcal{B}_{\varnothing}^{CP2} > 13.5$ in all realizations, implying that our methodology can detect a "common" process if the properties of the noise are broadly similar but far from identical, with the amplitude of the noise varies by three orders of magnitude. Figure 5 shows the recovered common-noise spectrum and the injected timing noise models. We continued to increase the spread in injected noise amplitudes upward from $\approx 3$ orders of magnitude, and found that common noise is disfavored when the spread in amplitude exceeds 5 orders of magnitude. Thus, the simulated data only favors the correct hypothesis when the range of simulated noise parameters starts to resemble the uniform red noise prior assumed in recent analyses.
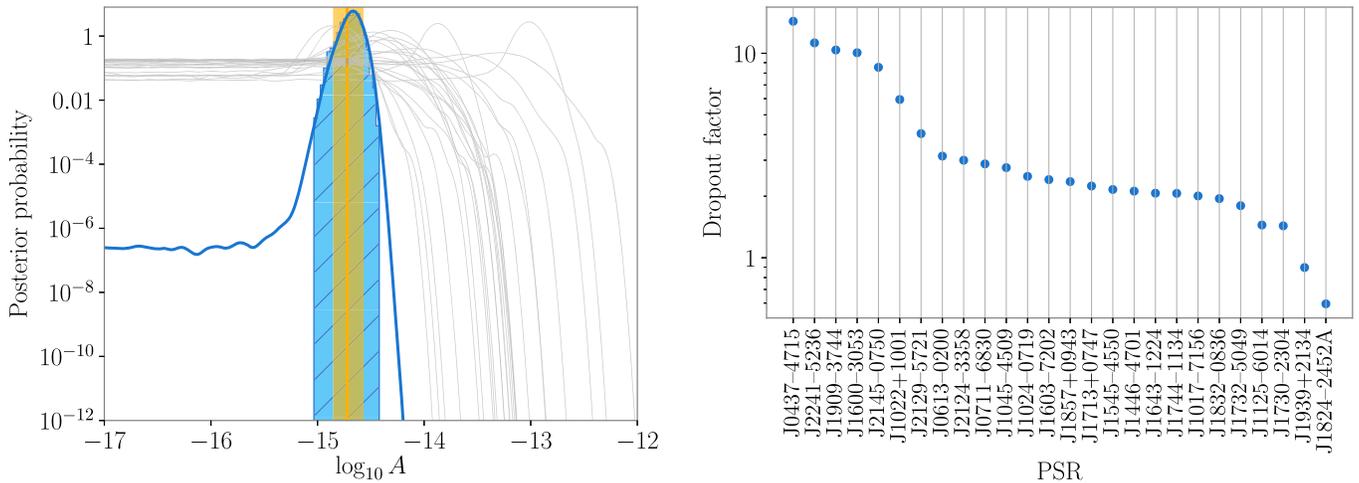
**Figure 2.** Pulsar contributions to the common red noise, assuming a fixed power-law index of $-13/3$ (CP2). Left: posterior distributions for the common red-noise amplitude, $A$. The hatched blue area is the result of a joint analysis of all pulsars with fixed white-noise parameters. The thick blue line shows the distribution obtained from a factorized-likelihood approach. Thin gray lines show contributions from individual pulsars to the factorized posterior. The yellow vertical line and the shaded region represent the median and $1\sigma$ levels of the NANOGrav measurement. Right: dropout factors for PPTA DR2 pulsars. We interpret the dropout factors to represent the consistency of noise in a given pulsar with CP2, as discussed in Section 4.3.

It is physically likely that intrinsic pulsar timing noise has similar, but not identical properties between different pulsars (Shannon & Cordes 2010). As such, a second null hypothesis (not tested by the current analysis) would be that each pulsar has independent red noise, but the properties of that noise cluster in a similar range. This could be examined by assuming that the amplitude and spectral index of the noise terms for the pulsars are not identical, but are drawn from a distribution. The nonuniform noise hypothesis is distinct both from the signal hypothesis with the delta-function prior and from the noise hypothesis with the uniform prior. For example, the new noise prior could be modeled as a Gaussian distribution of width $\mu_A$ and $\mu_\gamma$, and variance $\sigma_A^2$ and $\sigma_\gamma^2$. If the variance is inconsistent with zero it would suggest that the noise processes were not common but just similar. If it is consistent with zero then values could be used to constrain the properties of a common process.

As a gravitational-wave background signal will affect all pulsars, it will be apparent not only as a common spectral process, but also as a "noise floor". The noise level of a given pulsar should not be below this floor apart from statistical fluctuations, including instances where two noise processes cancel each other out. We updated the simulations shown in Figure 4 by including timing noise with identical power-law spectral densities ($\log_{10} A = -15.51$, $\gamma = 5.5$) for 25 pulsars, and a lower, shallower-spectrum timing noise ($\log_{10} A = -14.38$, $\gamma = 1.5$) in the pulsar with the lowest white noise levels. The power spectral densities corresponding to these simulations are shown in the right-hand panel of Figure 4. We performed model selection for CP2 over $\varnothing$, and again found significant support for the common-spectrum noise process ($\ln \mathcal{B}_\varnothing^{\mathrm{CP2}} > 13.1$ across all realizations), which is comparable to the support found in both the NANOGrav and PPTA analyses. As the power spectral density in the lowest frequency channel for the low-noise pulsar is 4 orders of magnitude lower than the common signal, the model selection is not explicitly identifying a noise floor. However, a factorized-likelihood dropout analysis showed that the simulated pulsar with a low noise level was not consistent with the

retrieved common-spectrum process (with a dropout factor $<1$; see the following section).

### 4.3. Do We Understand which Pulsars Are Contributing to the Evidence?

The dropout factor does not explicitly assess individual pulsar contributions to a common-noise process. As shown in Figure 2, unexpected pulsars (such as PSR J2241−5236, which is a relatively new addition to the PPTA sample) have high dropout factors. As the dropout factor is the integral of the product of an individual posterior distribution constraint on the common red process with the posterior distribution of the apparent common-spectrum process in the rest, pulsars with uninformative posterior distributions (i.e., pulsars with relatively high white noise and no evidence of red noise) can still have significant dropout factors. Therefore, the dropout factor represents not a "contribution statistic," but a "consistency statistic": pulsars with high factors have noise that is not inconsistent with the presence of a common process. If we wish to determine which pulsars contribute to the evidence for a red-noise process then we could calculate evidence values in favor of common noise, starting from the pulsar that provides the best (or worst) single-pulsar limit on the gravitational-wave background, and iteratively increasing the number of pulsars, or calculate the change in evidence when removing individual pulsars from the array.

### 4.4. Are We Affected by the Choice of Solar System Ephemeris?

The amplitude of CP2 is higher than the two 95% confidence upper limits previously set by NANOGrav ($A < 1.45 \times 10^{-15}$) and the PPTA ($A < 1.0 \times 10^{-15}$). The PPTA limit was based on the DE421 solar system ephemeris, without marginalizing over errors in the ephemeris and included data up to the beginning of 2015.

In order to carry out an initial exploration, to see if the results in this paper may be affected by errors within the solar system ephemeris, we modeled a subset of the potential errors as parameterized perturbations in the ephemerides (BAYESEPHEM;
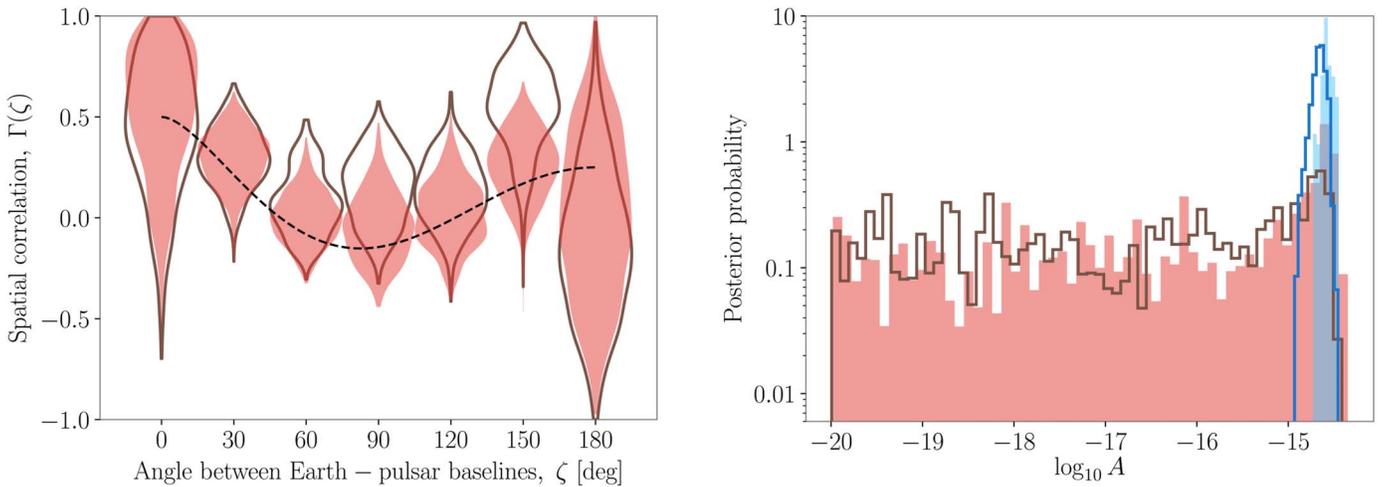
**Figure 3.** Left: inferred inter-pulsar spatial correlations in PPTA DR2 at seven node angles $\zeta$ between Earth-pulsar baselines. The dashed line is the predicted correlation from the gravitational-wave background. Right: power-law amplitude of the Hellings–Downs process without auto-correlation (red) and of the common red noise (blue). In both figures, lines represent the full PTA based on the assumptions of timing noise only in pulsars according to Goncharov et al. (2021), whereas the filled regions represent PPTA DR2 without PSR J0437−4715 and timing noise terms in all pulsars.

Vallisneri et al. 2020). We performed model selection for perturbations in (1) the masses of Mars, Jupiter, and Saturn; (2) the individual Keplerian orbital elements of these planets; and (3) in the rate of rotation about the ecliptic pole. These terms, except for Mars and Saturn, were marginalized over by Arzoumanian et al. (2020). The resulting Bayes factors are negative and hence we conclude that the presence of such errors is not favored by the data. We also performed the same analysis using the JPL DE438, DE430, and DE421 ephemerides. Neither DE436 nor DE438 provide evidence for any errors. Using DE430, we find positive $\ln \mathcal{B} = 1.6$ for an error in one of the Jupiter's orbital elements and $\ln \mathcal{B} = 0.6$ favoring both the error in Saturn's mass and one of the Saturn's orbital elements. Using the DE421 ephemeris, the oldest one that we tested, we only find a positive $\ln \mathcal{B} = 0.3$ for an error in the mass of Saturn.

### 4.5. Searching for Spatial Correlations

Only a correlation analysis will provide incontrovertible evidence of a gravitational-wave background and we currently have no statistical evidence for the presence of spatial correlations. The overlap reduction function in Figure 3 is consistent with the expected correlations from a gravitational-wave background (in particular if PSR J0437−4715 is removed from the analysis). We note that the bins in Figure 3 are interpolated and correlated, which may boost the apparent significance seen by eye. As shown in Figure 2, the noise spectrum in PSR J0437−4715 is consistent with CP2 and hence it provides the highest dropout factor. However, as shown in the right-hand panel of Figure 3, the inclusion of PSR J0437–4715 lowers the posterior probability of spatial correlations at the maximum-posterior value of $A$ of CP2, further diminishing the evidence for a gravitational-wave background. Unfortunately, the only telescope with a long timing baseline for PSR J0437−4715 is Parkes and it is therefore challenging to confirm the noise modeling for this pulsar. In the future it will be possible to compare with observations taken with, for example, MeerKAT as part of the MeerTime project (Bailes et al. 2020), or with the Jansky Very Large Array as part of NANOGrav.

When evidence for Hellings–Downs correlations in pulsar timing array data sets is found, it will be important to examine the hypothesis tested. Simulations, such as ones containing uncorrelated red noise could be used to determine the likelihood of improperly modeled uncommon red noise inducing these spatial correlations.

With the provisos given above we have no evidence that the detected CP1 or CP2 noise process is linked to a gravitational-wave background. However, if the signal is a bona fide astrophysical gravitational-wave background, then the relatively high amplitude would favor high merger-rate densities, short merger timescales, and high normalizations for the black hole—galaxy bulge mass relation (Middleton et al. 2021). The near-future prospects for probing the origin of the signal and the underlying dynamics of the supermassive black hole population are discussed in Pol et al. (2021). Sesana (2013) showed that background amplitudes of $>10^{-15}$ could be caused by the effect of overmassive black holes on black hole–host relations. The detected amplitude is also within observationally constrained limits based on the local supermassive black hole mass function (Zhu et al. 2018).

### 5. Conclusions

Under the assumptions of an analysis of the NANOGrav 12.5 yr data set (Arzoumanian et al. 2020), we have detected with high confidence a common-spectrum time-correlated signal in the timing of the 26 PPTA-DR2 millisecond pulsars. However, as noted above, there are some important caveats that need to be addressed before the signature can be confidently attributed to a physical process common to all the pulsars in the array. We do not confirm or rule out that the common-spectrum process is a spatially correlated stochastic gravitational-wave signal. However, the identified process does not possess monopole or dipole correlations and is not caused by errors in masses and trajectories of Mars, Jupiter, and Saturn, that would have resulted in deterministic timing residuals according to BAYESEPHEM models (Vallisneri et al. 2020). Proposed follow-up work relates to (1) improving our understanding of the properties of the intrinsic timing noise in millisecond pulsars and (2) identifying the optimal model comparisons and
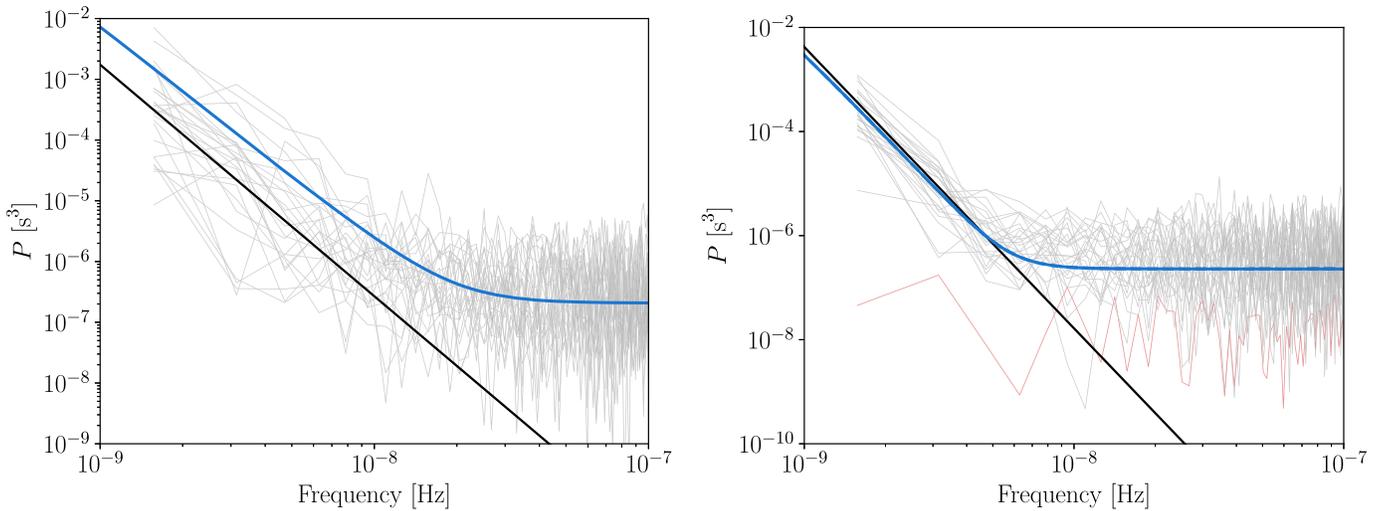
**Figure 4.** Power spectral density of the two sets of simulations that mimic a common red process. The gray lines represent the injected input noise spectra of simulated timing residuals for the 26 pulsars. The spectra were formed using a generalized least squares technique (Coles et al. 2011). The blue lines represent the pulsar-averaged values, and the black lines indicate the recovered common spectrum. Left: pulsars with similar, but not identical, timing noise properties. Right: a simulation where 25 pulsars have identical timing noise properties, and one has a lower and shallower timing noise than the others (red line).
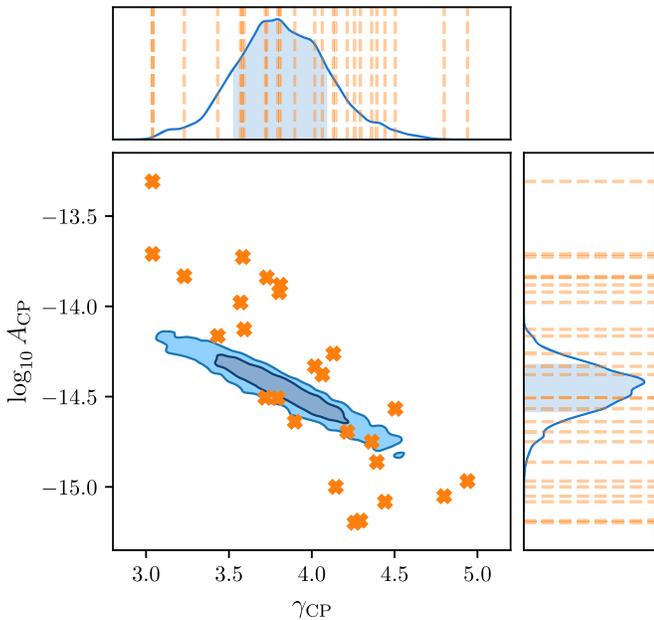


**Figure 5.** Simulated steep timing noise parameters and recovered common red process parameters, corresponding to the left panel in Figure 4. The contours represent posterior probability density for the common process amplitude and spectral index. The amplitudes and spectral indices of the injected timing-noise spectra for the 26 pulsars are indicated by the orange crosses and orange dashed lines.

methodologies that can determine whether the noise detected corresponds to a "noise floor" and is identical in all pulsars.

The PPTA project now has nearly three additional years of data obtained with a higher sensitivity wide-band system (Hobbs et al. 2020) that can be added with the data presented to increase timing baselines. We will also combine the PPTA data sets with observations from other observatories as part of the International Pulsar Timing Array project (Perera et al. 2019), with the latter being ideal to continue this work. Such lengthened and more sensitive data sets will allow us to probe timescales significantly longer than the orbital period of Jupiter and closer to that of Saturn. We will be able to compare noise models obtained from a wide range of telescopes and maximize our chance of determining the nature of the red-noise signals that are present in our data.

These data sets will also enable more sensitive and robust searches for the Hellings–Downs spatial correlations necessary to make a definitive detection of the gravitational-wave background.

### ORCID iDs

Boris Goncharov https://orcid.org/0000-0003-3189-5807
R. M. Shannon https://orcid.org/0000-0002-7285-6348
D. J. Reardon https://orcid.org/0000-0002-2035-4688
A. Zic https://orcid.org/0000-0002-9583-2947
M. Bailes https://orcid.org/0000-0003-3294-3081
S. Dai https://orcid.org/0000-0002-9618-2499
M. Kerr https://orcid.org/0000-0002-0893-4073
M. E. Lower https://orcid.org/0000-0001-9208-0009
R. N. Manchester https://orcid.org/0000-0001-9445-5732
R. Mandow https://orcid.org/0000-0001-5131-522X

H. Middleton ⓘ https://orcid.org/0000-0001-5532-3622
A. Parthasarathy ⓘ https://orcid.org/0000-0002-4140-5616
E. Thrane ⓘ https://orcid.org/0000-0002-4418-3895
N. Thyagarajan ⓘ https://orcid.org/0000-0003-1602-7868
X.-J. Zhu ⓘ https://orcid.org/0000-0001-7049-6468
A. D. Cameron ⓘ https://orcid.org/0000-0002-2037-4216
Y. Feng ⓘ https://orcid.org/0000-0002-0475-7479
R. Luo ⓘ https://orcid.org/0000-0002-4300-121X
C. J. Russell ⓘ https://orcid.org/0000-0002-1942-7296
R. Spiewak ⓘ https://orcid.org/0000-0002-6730-3298
S. Wang ⓘ https://orcid.org/0000-0003-4498-6070
J. B. Wang ⓘ https://orcid.org/0000-0001-9782-1603
L. Zhang ⓘ https://orcid.org/0000-0001-8539-4237

## References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102
Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018, ApJ, 859, 47
Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, ApJL, 905, L34
Bailes, M., Jameson, A., Abbate, F., et al. 2020, PASA, 37, e028
Carlin, B. P., & Chib, S. 1995, J. R. Stat. Soc. Ser. B., 57, 473
Cognard, I., & Backer, D. C. 2004, ApJL, 612, L125
Coles, W., Hobbs, G., Champion, D. J., Manchester, R. N., & Verbiest, J. P. W. 2011, MNRAS, 418, 561
Cordes, J. M., & Helfand, D. J. 1980, ApJ, 239, 640
Davidson, R., & MacKinnon, J. G. 1981, Econometrica, 49, 781
Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94
Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2019, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, Zenodo, doi:10.5281/zenodo.4059815
Finn, L. S., Larson, S. L., & Romano, J. D. 2009, PhRvD, 79, 062003
Goncharov, B., Zhu, X.-J., & Thrane, E. 2020, MNRAS, 497, 3264
Goncharov, B., Reardon, D. J., Shannon, R. M., et al. 2021, MNRAS, 502, 478
Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39
Hinton, S. R. 2016, JOSS, 1, 00045
Hobbs, G., Lyne, A. G., & Kramer, M. 2010, MNRAS, 402, 1027
Hobbs, G., Manchester, R. N., Dunning, A., et al. 2020, PASA, 37, e012
Jenet, F. A., & Romano, J. D. 2015, AmJPh, 83, 635
Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, ApJ, 653, 1571
Keith, M. J., Coles, W., Shannon, R. M., et al. 2013, MNRAS, 429, 2161
Kerr, M., Reardon, D. J., Hobbs, G., et al. 2020, PASA, 37, e020
Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, MNRAS, 453, 2576
Manchester, R. N., Hobbs, G., Bailes, M., et al. 2013, PASA, 30, e017
McKee, J. W., Janssen, G. H., Stappers, B. W., et al. 2016, MNRAS, 461, 2809
Middleton, H., Sesana, A., Chen, S., et al. 2021, MNRAS, 502, L99
Müller, U. K. 2013, Econometrica, 81, 1805
NANOGrav Collaboration, Arzoumanian, Z., Brazier, A., et al. 2015, ApJ, 813, 65
Parthasarathy, A., Shannon, R. M., Johnston, S., et al. 2019, MNRAS, 489, 3810
Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, MNRAS, 490, 4666
Phinney, E. S. 2001, arXiv:astro-ph/0108028
Pol, N. S., Taylor, S. R., Kelley, L. Z., et al. 2021, ApJL, 911, L34
Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. 2014, MNRAS, 442, 56
Reardon, D. J., Shannon, R. M., Cameron, A. D., et al. 2021, MNRAS, in press (doi:10.1093/mnras/stab1990)
Rosado, P. A., Sesana, A., & Gair, J. 2015, MNRAS, 451, 2417
Sampson, L., Cornish, N. J., & McWilliams, S. T. 2015, PhRvD, 91, 084055
Sesana, A. 2013, MNRAS, 433, L1
Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607
Shannon, R. M., Ravi, V., Coles, W. A., et al. 2013, Sci, 342, 334
Shannon, R. M., Ravi, V., Lentati, L. T., et al. 2015, Sci, 349, 1522
Shannon, R. M., Lentati, L. T., Kerr, M., et al. 2016, ApJL, 828, L1
Taylor, S. R., Gair, J. R., & Lentati, L. 2013, PhRvD, 87, 044035
Taylor, S. R., Lentati, L., Babak, S., et al. 2017a, PhRvD, 95, 042002
Taylor, S. R., Simon, J., & Sampson, L. 2017b, PhRvL, 118, 181102
Taylor, S. R., van Haasteren, R., & Sesana, A. 2020, PhRvD, 102, 084039
Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, MNRAS, 455, 4339
Vallisneri, M., Taylor, S. R., Simon, J., et al. 2020, ApJ, 893, 112
van Haasteren, R., Levin, Y., Janssen, G. H., et al. 2011, MNRAS, 414, 3117
Zhu, X.-J., Cui, W., & Thrane, E. 2018, MNRAS, 482, 2588