The Astrophysical Journal Letters, 905:L34 (18pp), 2020 December 20 https://doi.org/10.3847/2041-8213/abd401

The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background

Zaven Arzoumanian[1], Paul T. Baker[2], Harsha Blumer[3,4], Bence Bécsy[5], Adam Brazier[6,7], Paul R. Brook[3,4], Sarah Burke-Spolaor[3,4,8], Shami Chatterjee[6], Siyuan Chen[9,10,11], James M. Cordes[6], Neil J. Cornish[5], Fronefield Crawford[12], H. Thankful Cromartie[6,13], Megan E. DeCesar[14,15,43], Paul B. Demorest[16], Timothy Dolch[17], Justin A. Ellis[18], Elizabeth C. Ferrara[19], William Fiore[3,4], Emmanuel Fonseca[20], Nathan Garver-Daniels[3,4], Peter A. Gentile[3,4], Deborah C. Good[21], Jeffrey S. Hazboun[22,43], A. Miguel Holgado[23,24], Kristina Islo[25], Ross J. Jennings[6], Megan L. Jones[25], Andrew R. Kaiser[3,4], David L. Kaplan[25], Luke Zoltan Kelley[26], Joey Shapiro Key[22], Nima Laal[27], Michael T. Lam[28,29], T. Joseph W. Lazio[30], Duncan R. Lorimer[3,4], Jing Luo[31], Ryan S. Lynch[32], Dustin R. Madison[3,4,43], Maura A. McLaughlin[3,4], Chiara M. F. Mingarelli[33,34], Cherry Ng[35], David J. Nice[14], Timothy T. Pennucci[36,37,43], Nihan S. Pol[3,4,38], Scott M. Ransom[36], Paul S. Ray[39], Brent J. Shapiro-Albert[3,4], Xavier Siemens[25,27], Joseph Simon[30,40,44], Renée Spiewak[41], Ingrid H. Stairs[21], Daniel R. Stinebring[42], Kevin Stovall[16], Jerry P. Sun[27], Joseph K. Swiggum[14,43], Stephen R. Taylor[38], Jacob E. Turner[3,4], Michele Vallisneri[30], Sarah J. Vigeland[25], and Caitlin A. Witt[3,4]

The NANOGrav Collaboration

[1] X-Ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA
[2] Department of Physics and Astronomy, Widener University, One University Place, Chester, PA 19013, USA
[3] Department of Physics and Astronomy, West Virginia University, P.O. Box 6315, Morgantown, WV 26506, USA
[4] Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA
[5] Department of Physics, Montana State University, Bozeman, MT 59717, USA
[6] Cornell Center for Astrophysics and Planetary Science and Department of Astronomy, Cornell University, Ithaca, NY 14853, USA
[7] Cornell Center for Advanced Computing, Cornell University, Ithaca, NY 14853, USA
[8] CIFAR Azrieli Global Scholars Program, CIFAR, Toronto, Canada
[9] Station de Radioastronomie de Nancay, Observatoire de Paris, Universite PSL, CNRS, Universite d'Orleans, F-18330 Nancay, France
[10] FEMTO-ST Institut de recherche, Department of Time and Frequency, UBFC and CNRS, ENSMM, F-25030 Besancon, France
[11] Laboratoire de Physique et Chimie de l'Environment et de l'Espace, LPC2E UMR7328, Universite d'Orleans, CNRS, F-45071 Orleans, France
[12] Department of Physics and Astronomy, Franklin & Marshall College, P.O. Box 3003, Lancaster, PA 17604, USA
[13] Department of Astronomy, University of Virginia, P.O. Box 400325, Charlottesville, VA 22904, USA
[14] Department of Physics, Lafayette College, Easton, PA 18042, USA
[15] George Mason University, Fairfax, VA 22030, resident at U.S. Naval Research Laboratory, Washington, D.C. 20375, USA
[16] National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801, USA
[17] Department of Physics, Hillsdale College, 33 E. College Street, Hillsdale, MI 49242, USA
[18] Infinia ML, 202 Rigsbee Avenue, Durham NC, 27701, USA
[19] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[20] Department of Physics, McGill University, 3600 University St., Montreal, QC H3A 2T8, Canada
[21] Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[22] University of Washington Bothell, 18115 Campus Way NE, Bothell, WA 98011, USA
[23] Department of Astronomy and National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[24] McWilliams Center for Cosmology and Department of Physics, Carnegie Mellon University, Pittsburgh PA, 15213, USA
[25] Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA
[26] Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL 60208, USA
[27] Department of Physics, Oregon State University, Corvallis, OR 97331, USA
[28] School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA
[29] Laboratory for Multiwavelength Astrophysics, Rochester Institute of Technology, Rochester, NY 14623, USA
[30] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[31] Department of Astronomy & Astrophysics, University of Toronto, 50 Saint George Street, Toronto, ON M5S 3H4, Canada
[32] Green Bank Observatory, P.O. Box 2, Green Bank, WV 24944, USA
[33] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY, 10010, USA
[34] Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA
[35] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[36] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[37] Institute of Physics, Eötvös Loránd University, Pázmány P.s. 1/A, H-1117 Budapest, Hungary
[38] Department of Physics and Astronomy, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235, USA
[39] Space Science Division, Naval Research Laboratory, Washington, DC 20375-5352, USA
[40] Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA
[41] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, Victoria 3122, Australia
[42] Department of Physics and Astronomy, Oberlin College, Oberlin, OH 44074, USA

Received 2020 September 9; revised 2020 December 11; accepted 2020 December 16; published 2020 December 24

Abstract

We search for an isotropic stochastic gravitational-wave background (GWB) in the 12.5 yr pulsar-timing data set collected by the North American Nanohertz Observatory for Gravitational Waves. Our analysis finds strong evidence of a

[43] NANOGrav Physics Frontiers Center Postdoctoral Fellow.
[44] Corresponding author: joe.simon@nanograv.org

stochastic process, modeled as a power law, with common amplitude and spectral slope across pulsars. Under our fiducial model, the Bayesian posterior of the amplitude for an $f^{-2/3}$ power-law spectrum, expressed as the characteristic GW strain, has median 1.92×10^{-15} and 5%–95% quantiles of $1.37–2.67 \times 10^{-15}$ at a reference frequency of $f_{\rm yr} = 1\ {\rm yr}^{-1}$; the Bayes factor in favor of the common-spectrum process versus independent red-noise processes in each pulsar exceeds 10,000. However, we find no statistically significant evidence that this process has quadrupolar spatial correlations, which we would consider necessary to claim a GWB detection consistent with general relativity. We find that the process has neither monopolar nor dipolar correlations, which may arise from, for example, reference clock or solar system ephemeris systematics, respectively. The amplitude posterior has significant support above previously reported upper limits; we explain this in terms of the Bayesian priors assumed for intrinsic pulsar red noise. We examine potential implications for the supermassive black hole binary population under the hypothesis that the signal is indeed astrophysical in nature.

Unified Astronomy Thesaurus concepts: Gravitational waves (678); Pulsar timing method (1305); Astronomy data analysis (1858); Millisecond pulsars (1062)

1. Introduction

Pulsar-timing arrays (PTAs; Sazhin 1978; Detweiler 1979; Foster & Backer 1990) seek to detect very-low-frequency (∼1–100 nHz) gravitational waves (GWs) by monitoring the spatially correlated fluctuations induced by the waves on the times of arrival of radio pulses from millisecond pulsars (MSPs). The dominant source of gravitational radiation in this band is expected to be the stochastic background generated by a cosmic population of supermassive black hole binaries (SMBHBs; Sesana et al. 2004; Burke-Spolaor et al. 2019). Other more speculative stochastic GW sources in the nanohertz frequency range include cosmic strings (Siemens et al. 2007; Blanco-Pillado et al. 2018), phase transitions (Caprini et al. 2010; Kobakhidze et al. 2017), and a primordial GW background (GWB) produced by quantum fluctuations of the gravitational field in the early universe, amplified by inflation (Grishchuk 1975; Lasky et al. 2016).

The North American Nanohertz Observatory for Gravitational Waves (NANOGrav; Ransom et al. 2019) has been acquiring pulsar-timing data since 2004. NANOGrav is one of three major PTAs along with the European Pulsar Timing Array (EPTA; Desvignes et al. 2016) and the Parkes Pulsar Timing Array (PPTA; Kerr et al. 2020). Additionally, there are growing PTA efforts in India (Joshi et al. 2018) and China (Lee 2016), as well as some telescope-centered timing programs (Bailes et al. 2016; Ng 2018). In concert, these collaborations support the International Pulsar Timing Array (IPTA; Perera et al. 2019). Over the last decade, PTAs have produced increasingly sensitive data sets, as seen in the steady march of declining upper limits on the stochastic GWB (van Haasteren et al. 2011; Demorest et al. 2013; Shannon et al. 2013; Lentati et al. 2015; Shannon et al. 2015; Arzoumanian et al. 2016, 2018a; Verbiest et al. 2016). It was widely expected that the first inklings of a GWB would manifest in the stagnation of improvement in upper limits, followed by the emergence of a spatially uncorrelated common-spectrum red process in all pulsars, and culminate in the detection of interpulsar spatial correlations with the quadrupolar signature described by Hellings & Downs (1983). In practice, it appears that early indications of a signal may have been obscured by systematic effects due to incomplete knowledge of the assumed position of the solar system barycenter (Vallisneri et al. 2020).

In this article, we present our analysis of NANOGrav's newest "12.5 yr" data set (Alam et al. 2021a, hereafter NG12). We find a strong preference for a stochastic common-spectrum process, modeled as a power law, in the timing behaviors of all pulsars in the data set. Building on the statistical-inference framework put in place during our GW study of the 11 yr data set (Arzoumanian et al. 2018a, hereafter NG11gwb), we report Bayes factors from extensive model comparisons. We find the $\log_{10}$ Bayes factor for a spatially uncorrelated common-spectrum process versus independent red-noise processes in each pulsar to range from 2.7 to 4.5, depending on which solar system ephemeris (SSE) modeling scheme we employ. We model a spatially uncorrelated common-spectrum process to have the same power spectral density across all pulsars in the data set, but with independent realizations in the specific timing behavior of each pulsar. The evidence is only slightly higher for a common-spectrum process with quadrupolar correlations, with a $\log_{10}$ Bayes factor against a spatially uncorrelated common-spectrum process ranging from 0.37 to 0.64, again depending on SSE modeling. Correspondingly, the Bayesian–frequentist hybrid optimal-statistic analysis (Anholm et al. 2009; Demorest et al. 2013; Chamberlin et al. 2015; Vigeland et al. 2018), which measures interpulsar correlated power only, is unable to distinguish between different spatially correlated processes. Thus, lacking definitive evidence of quadrupolar spatial correlations, the analysis of this data set must be considered inconclusive with regard to GW detection.

With an eye toward searches in future, more informative data sets, we perform a suite of statistical tests on the robustness of our findings. Focusing first on the stochastic common-spectrum process, we examine the contribution of each pulsar to the overall Bayes factor with a dropout analysis (Aggarwal et al. 2019; S. Vigeland et al. 2020, in preparation) and find broad support among the pulsars in the data set. Moving on to spatial correlations, we build null background distributions for the correlation statistics by applying random phase shifts and sky scrambles to our data (Cornish & Sampson 2016; Taylor et al. 2017a) and find that the no-correlations hypothesis is rejected only mildly, with p values ∼5% (i.e., 2σ).

The posterior on the amplitude of the common-spectrum process, $A_{\rm CP}$, modeled with an $f^{-2/3}$ power-law spectrum, has a median of 1.9×10^{-15}, with 5%–95% quantiles of $1.4–2.7 \times 10^{-15}$ at a reference frequency of $f_{\rm yr} = 1{\rm yr}^{-1}$, based on a log-uniform prior and using the latest JPL SSE (DE438, Folkner & Park 2018), which we take as our fiducial model in this paper. This refined version of the SSE incorporates data from the NASA orbiter Juno[45] and claims a Jupiter orbit accuracy a factor of 4 better than previous SSEs, which is promising given that our previous analysis showed that errors in Jupiter's orbit dominated the SSE-induced GWB systematics (Vallisneri et al. 2020).

[45] https://www.missionjuno.swri.edu

The fact that the median value of $A_{\rm CP}$ is higher than the 95% upper limit reported for the 11 yr data set, $A_{\rm GWB} < 1.45 \times 10^{-15}$ (NG11gwb), requires explanation. While many factors contribute to this discrepancy, simulations show that the standard PTA data model (and most crucially, the uniform priors on the amplitude of pulsar-intrinsic red-noise processes) can often yield Bayesian upper limits lower than the true GWB level by shifting GWB power to pulsar red noise (Hazboun et al. 2020a). Once all factors are taken into account, the data sets can be reconciled. However, this accounting suggests that the astrophysical interpretation of past Bayesian upper limits from PTAs may have been overstated. Indeed, it is worth noting that while the source of the common-spectrum process in this data set remains unconfirmed, the posterior on $A_{\rm CP}$ is compatible with many models for the GWB that had previously been deemed in tension with PTA analyses.

This paper is laid out as follows: Section 2 describes the 12.5 yr data set. Our data model is presented in Section 3. In Section 4, we report on our search for a common-spectrum process in the data set and present the results from our extensive exploration for interpulsar correlations. Section 5 contains a suite of statistical checks on the significance of our detection metrics. In Section 6, we discuss the amplitude of the recovered process, addressing both the discrepancies with previous published upper limits and the potential implications for the SMBHB population, and we conclude with our expectations for future searches.

2. The 12.5 yr Data Set

The NANOGrav 12.5 yr data set has been released using two separate and independent analyses. The narrowband analysis, consisting of the time-of-arrival (TOA) data and pulsar-timing models presented in NG12, is very similar in its form and construction to our previous data sets in which many TOAs were calculated within narrow radio-frequency bands for data collected simultaneously across a wide bandwidth. A separate "wideband" analysis (Alam et al. 2021b) was also performed in which a single TOA is extracted from broadband observations. Both versions of the data set are publicly available online.[46] The data set consists of observations of 47 MSPs made between 2004 July and 2017 June. This is the fourth public NANOGrav data set and adds two MSPs and 1.5 yr of observations to the previously released 11 yr data set (NG11). Only pulsars with a timing baseline greater than 3 yr are used in our GW analyses (Arzoumanian et al. 2016, hereafter NG9gwb), and thus all results in this paper are based on the 45 pulsars that meet that criteria. This is a significant increase from the analyses in NG11gwb, which used 34 pulsars, and the analyses in NG9gwb, which used 18. Additionally, it is crucial to note that the 12.5 yr data set is more than just an extension of the 11 yr data set—changes to the data-processing pipeline, discussed below, have improved the entire span of the data. In the following section, we briefly summarize the instruments, observations, and data reduction process for the 12.5 yr data set. A more detailed discussion of the data set can be found in NG12.

2.1. Observations

We used the 305 m Arecibo Observatory (Arecibo or AO) and the 100 m Green Bank Telescope (GBT) to observe the pulsars. Arecibo observed all sources that lie within its decl. range ($0° < \delta < +39°$), while GBT observed those sources that lie outside of Arecibo's decl. range, plus PSRs J1713+0747 and B1937+21. Most sources were observed approximately once per month. Six pulsars were observed weekly as part of a high-cadence observing campaign, which began at the GBT in 2013 and at AO in 2015 with the goal of improving our sensitivity to individual GW sources (Burt et al. 2011; Christy et al. 2014): PSRs J0030+0451, J1640+2224, J1713+0747, J1909−3744, J2043+1711, and J2317+1439.

Early observations were recorded using the ASP and GASP systems at Arecibo and GBT, respectively, which sampled bandwidths of 64 MHz (Demorest 2007). Between 2010 and 2012, we transitioned to wideband systems (PUPPI at Arecibo and GUPPI at GBT) that can process up to 800 MHz bandwidths (DuPlain et al. 2008; Ford et al. 2010). At most observing epochs, the pulsars were observed with two different wideband receivers covering different frequency ranges in order to achieve good sensitivity in the measurement of pulse dispersion due to the interstellar medium. At Arecibo, the pulsars were observed using the 1.4 GHz receiver plus either the 430 MHz receiver or 2.1 GHz receiver, depending on the pulsar's spectral index and timing characteristics. (Early observations of one pulsar also used the 327 MHz receiver.) At GBT, the monthly observations used the 820 MHz and 1.4 GHz receivers. However, these two separate frequency ranges were not observed simultaneously; instead, the observations were separated by a few days. The weekly observations at GBT used only the 1.4 GHz receiver.

2.2. Processing and Time-of-arrival Data

Most of the procedures used to reduce the data, generate the TOAs, and clean the data set were similar to those used to generate previous NANOGrav data sets (NG9, NG11); however, several new steps were added. We improved the data reduction pipeline by removing low-amplitude artifact images from the profile data that are caused by small mismatches in the gains and timing of the interleaved analog-to-digital converters in the backends. We also excised radio-frequency interference (RFI) from the calibration files as well as the data files.

We used the same procedures as in NG9 and NG11 to generate the TOAs from the profile data. As we have done in previous data sets, we cleaned the TOAs by removing RFI, low signal-to-noise TOAs (NG9), and outliers (NG11). Compared to previous data sets, we reorganized and systematized the TOA cleaning and timing-model parameter selection processes to improve consistency of processing across all pulsars. We also performed a new test where observing epochs were removed one by one to determine whether removing a particular epoch significantly changed the timing model. This is essentially an outlier analysis for observing epochs rather than individual TOAs.

2.3. Timing Models and Noise Analysis

For each pulsar, the cleaned TOAs were fit to a timing model that described the pulsar's spin period and spin period derivative, sky location, proper motion, and parallax. For binary pulsars, the timing model also included five Keplerian binary parameters and additional post-Keplerian parameters if they improved the timing fit as determined by an F test. We modeled variations in the pulse dispersion as a piecewise constant through the inclusion of DMX parameters (NG9;

[46] http://data.nanograv.org

Jones et al. 2017). The timing-model fits were primarily performed using the TEMPO timing software, and the software packages TEMPO2 and PINT were used to check for consistency. The timing-model fits were done using the TT(BIPM2017) timescale and the JPL SSE model DE436 (Folkner & Park 2016). The latest JPL SSE (DE438; Folkner & Park 2018), which we take as our fiducial model for the analyses in this paper, was not available when TOA processing was being done. However, this does not affect the results presented later, as the corresponding changes in the timing parameters are well within their linear range, which is marginalized away in the analysis (NG9, NG9gwb).

We modeled noise in the pulsars' residuals with three white-noise components plus a red-noise component. The white-noise components are EQUAD, which adds white noise in quadrature; ECORR, which describes white noise that is correlated within the same observing epoch but uncorrelated between different observing epochs; and EFAC, which scales the total template-fitting TOA uncertainty after the inclusion of the previous two white-noise terms. For all of these components, we used separate parameters for every combination of pulsar, backend, and receiver.

Many processes can produce red noise in pulsar residuals. The stochastic GWB appears in the residuals as red noise; however, it appears specifically correlated between different pulsars (Hellings & Downs 1983). Other astrophysical sources of red noise include spin noise, pulse profile changes, and imperfectly modeled dispersion-measure variations (Cordes 2013; Jones et al. 2017; Lam et al. 2017). These red-noise sources are unique to a given pulsar. There are also potential terrestrial sources of red noise, including clock errors and ephemeris errors (Tiburzi et al. 2016), which are correlated differently than the GWB. We model the intrinsic red noise of each pulsar as a power law, similar to the GWB (see Section 3.1).

The changes to the data-processing procedure described above significantly improved the quality of the data. In order to quantify the effect of these changes, we produced an "11 yr slice" data set by truncating the 12.5 yr data set at the MJD corresponding to the last observation in the 11 yr data set and compared the results of a full noise analysis of this data set to those for the 11 yr data set. As discussed in NG12, we found a reduction in the amount of white noise in the 11 yr slice compared to the 11 yr data set. However, we also found that the red noise changed for many pulsars. Specifically, there is a slight preference for a steeper spectral index across most of the pulsars, indicating that for some pulsars, the reduction in white noise produced an increased sensitivity to low-frequency red-noise processes, like the GWB.

3. Data Model

The statistical framework for the characterization of noise processes and GW signals in pulsar-timing data is well documented (see e.g., NG9gwb; NG11gwb). In this section, we give a concise description of our probabilistic model of the 12.5 yr data set, focusing on the differences from earlier studies. The model attempts to represent every known deterministic and stochastic source of timing residuals that could be interpreted as GWs: it extends the individual timing models of the pulsars (discussed in Section 2.3) by adding common-spectrum processes with specific correlation structures between pulsars. In Section 3.1, we define our spectral models of time-correlated (red) processes, which include pulsar-intrinsic red noise and the GWB; in Section 3.2, we list the combinations of time-correlated processes included in our Bayesian model-comparison trials; and in Section 3.3, we discuss our prescriptions for the SSE. Our Bayesian and frequentist techniques of choice will be described alongside our results in Sections 4 and 5, with more technical details in Appendices B and C.

3.1. Models of Time-correlated Processes

The principal results of this paper are referred to a fiducial power-law spectrum of the characteristic GW strain:

$$h_c(f) = A_{\rm GWB}\left(\frac{f}{f_{\rm yr}}\right)^{\alpha}, \tag{1}$$

with $\alpha = -2/3$ for a population of inspiraling SMBHBs in circular orbits whose evolution is dominated by GW emission (Phinney 2001). We performed our analysis in terms of the timing-residual cross-power spectral density,

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

where $\gamma = 3 - 2\alpha$ (so the fiducial SMBHB $\alpha = -2/3$ corresponds to $\gamma = 13/3$) and where Γ_{ab} is the overlap reduction function (ORF), which describes average correlations between pulsars a and b in the array as a function of the angle between them. For an isotropic GWB, the ORF is given by Hellings & Downs (1983), and we refer to it casually as "quadrupolar" or "HD" correlations.

Other spatially correlated effects present with different ORFs. Systematic errors in the SSE have a dipolar ORF, $\Gamma_{ab} = \cos\zeta_{ab}$, where ζ_{ab} represents the angle between pulsars a and b, while errors in the timescale (the "clock") have a monopolar ORF, $\Gamma_{ab} = 1$. Pulsar-intrinsic red noise is also modeled as a power law; however, in that case, there is no ORF. The $A_{\rm GWB}$ in Equation (2) is replaced with $A_{\rm red}$, and γ with $\gamma_{\rm red}$. There is a separate ($A_{\rm red}$, $\gamma_{\rm red}$) pair for each pulsar in the array.

As in NG9gwb and NG11gwb, we implemented stationary Gaussian processes with a power-law spectrum in rank-reduced fashion by approximating them as a sum over a sine–cosine Fourier basis with frequencies k/T and prior (weight) covariance $S_{ab}(k/T)/T$, where T is the span between the minimum and maximum TOAs in the array (van Haasteren & Vallisneri 2014). We use the same basis vectors to model all red noise in the array, both pulsar-intrinsic noise and global signals, like the GWB. Using a common set of vectors helps the sampling and reduces the likelihood computation time. In previous work, the number of basis vectors was chosen to be large enough (with $k = 1, \ldots, 30$) that inference results (specifically the Bayesian upper limit) for a common-spectrum signal became insensitive to adding more components. However, doing so has the disadvantage of potentially coupling white noise to the highest-frequency components of the red-noise process, thus biasing the recovery of the putative GWB, which is strongest in the lowest frequency bins.

For this paper, we revisit the issue and set the number of frequency components used to model common-spectrum signals to five, on the basis of theoretical arguments backed by a preliminary analysis of the data set. We begin with the

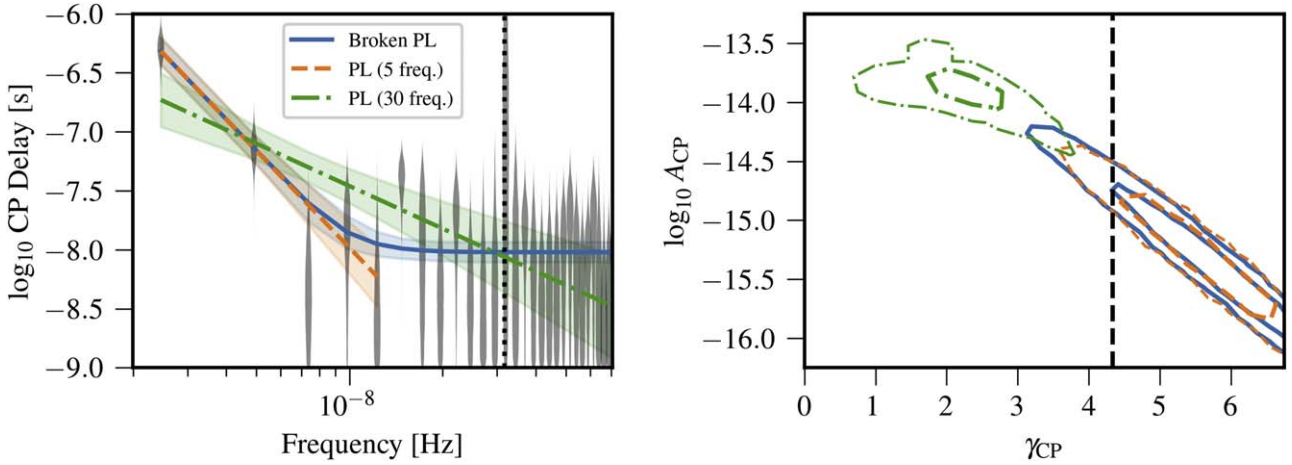

Figure 1. Posteriors for a common-spectrum process in NG12, as recovered with four models: free spectrum (gray violin plots in left panel), broken power law (solid blue lines and contours), 5-frequency power law (dashed orange lines and contours), and 30-frequency power law (dotted–dashed green lines and contours). In the left panel, the violin plots show marginalized posteriors of the equivalent amplitude of the sine–cosine Fourier pair (i.e., $\sqrt{S(f)/T}$, in units of seconds) at the frequencies on the horizontal axis; the lines show the mean reconstructed power laws in the left panel, and the 1σ (thicker) and 2σ posterior contours for the amplitude and spectral slope in the right panel. In the left panel, the shaded regions trace $\pm 1\sigma$ ranges for the common-spectrum process power as a function of frequency, as implied by the Bayesian posteriors for the power-law parameters. The dotted vertical line in the left panel sits at $f_{\rm yr} = 1{\rm yr}^{-1}$, where PTA sensitivity is reduced by the fitting of timing-model parameters; the corresponding free-spectrum amplitude posterior is unconstrained. The dashed vertical line in the right panel sits at $\gamma = 13/3$, the expected value for a GWB produced by a population of inspiraling SMBHBs. For both the broken power-law and 5-frequency power-law models, the amplitude ($A_{\rm CP}$) posterior shown on the right is extrapolated from the lowest frequencies to the reference frequency $f_{\rm yr}$. We observe that the slope and amplitude of the 30-frequency power law are driven by higher-frequency noise, whereas the 5-frequency power law recovers the low-frequency GWB-like slope of the free spectrum and broken power law.

former. By computing a strain-spectrum sensitivity curve for the 12.5 yr data set using the HASASIA tool (Hazboun et al. 2019) and obtaining the signal-to-noise ratio (S/N) of a $\gamma = 13/3$ power-law GWB, we observed that the five lowest frequency bins contribute 99.98% of the S/N, with the majority coming from the first bin. We also injected a $\gamma = 13/3$ power-law GWB into the 11 yr data set (NG11), and measured the response of each frequency using a 30-frequency free-spectrum model, in which we allowed the variance of each sine–cosine pair in the red-noise Fourier basis to vary independently. We observed that the lowest few frequencies are the first to respond as we raised the GWB amplitude from undetectable to detectable levels (see Figure 14 in Appendix A). The details of this injection analysis are described in Appendix A.

Moving on to empirical arguments, in Figure 1 we plot the power-spectrum estimates for a spatially uncorrelated common-spectrum process in the 12.5 yr data set, as computed for the following: a free-spectrum model (gray violin plots); variable-γ power-law models (Equation (2) with $A_{\rm GWB} = A_{\rm CP}$ and $\Gamma_{ab} = \delta_{ab}$) with 5 and 30 frequency components (dashed lines, showing maximum a posteriori values, as well as $1\sigma/2\sigma$ posterior contours); and a broken power-law model (solid lines), given by

$$S(f) = \frac{A_{\rm CP}^2}{12\pi^2}\left(\frac{f}{f_{\rm yr}}\right)^{-\gamma}\left(1+\left(\frac{f}{f_{\rm bend}}\right)^{1/\kappa}\right)^{\kappa(\gamma-\delta)} f_{\rm yr}^{-3}, \qquad (3)$$

where γ and δ are the slopes at frequencies lower and higher than $f_{\rm bend}$, respectively, and κ controls the smoothness of the transition. In this paper, we set $\delta = 0$ to appropriately capture the white noise coupled at higher frequencies and $\kappa = 0.1$, which is small enough to contain the transition between slopes to within an individual frequency bin. Both the free spectrum and the broken power law capture a steep red process at the lowest frequencies, in accordance with expectations for a GWB, which is accompanied by a flatter "forest" at higher frequencies. The 30-frequency power law is impacted by power at high frequencies (where we do not expect any detectable contributions from a GWB) and adopts a low spectral index that does not capture the full power in the lowest frequencies. By contrast, the five-frequency power law agrees with the free spectrum and broken power law in recovering a steep-spectral process.

The problem of pulsar-intrinsic excess noise leaking into the common-spectrum process at high frequencies has already been discussed for the 9 and 11 yr NANOGrav data sets (Aggarwal et al. 2019, 2020; Hazboun et al. 2020b), and we are addressing it through the creation of individually adapted noise models for each pulsar (J. Simon et al. 2020, in preparation). For this paper, we find a simpler solution by limiting all common-spectrum models to the five lowest frequencies. By contrast, we used 30 frequency components for all rank-reduced power-law models of pulsar-intrinsic red noise,[47] which is consistent with what is used in individual pulsar noise analyses and in the creation of the data set.

3.2. Models of Spatially Correlated Processes

We analyzed the 12.5 yr data set using a hierarchy of data models, which are compared in Bayesian fashion by evaluating the ratios of their evidence. All models include the same basic block for each pulsar, consisting of measurement noise, timing-model errors, pulsar-intrinsic white noise, and pulsar-intrinsic red noise described by a 30-frequency variable-γ power law, but they differ by the presence of one or two red-noise processes that appear in all pulsars with the same spectrum. As in previous work (NG9gwb; NG11gwb), we fixed all pulsar-intrinsic white-noise parameters to their maximum in the

[47] The Fourier basis is still built on frequencies k/T, where T is the maximum time span between TOAs in the array, and the same basis vectors are still used for all red-noise models.

Table 1
Data Models

NG11gwb Labels	1	2A	2B	2D	3A	(New)	3B	3D
Spatial Correlations		Single Common-spectrum Process				Two Common-spectrum Processes		
Uncorrelated		✓				✓		
Dipole			✓				✓	
Monopole				✓				✓
HD					✓	✓	✓	✓
Pulsar-intrinsic red-noise	✓	✓	✓	✓	✓	✓	✓	✓

Note. The data models analyzed in this paper are organized by the presence of spatially correlated common-spectrum noise processes. Model names are added for a direct comparison to the naming scheme employed in NG11gwb.

posterior probability distribution recovered from single-pulsar noise studies for computational efficiency.

The models are listed in Table 1, which also reports their labels as used in NG11gwb. The most basic variant (model 1 in NG11gwb) includes measurement noise and pulsar-intrinsic processes alone.

The next group of four models includes a single common-spectrum red-noise process. The first among them (model 2A of NG11gwb) features a GWB-like red-noise process with common spectrum, but without HD correlations. Because we expect the correlations to be much harder to detect than the diagonal S_{aa} terms in Equation (2), due to the values of the HD ORF (Γ_{ab}) being less than or equal to 0.5, and because the corresponding likelihood, which does not include any correlations, is very computationally efficient, this model has been the workhorse of PTA searches. However, the positive identification of a GWB will require evidence of a common-spectrum process with HD correlations, which also belongs to this group (model 3A of NG11gwb). The group is rounded out by common-spectrum processes with dipolar and monopolar spatial correlations, which may represent SSE and clock anomalies. For a convincing GWB detection, we expect the data to favor HD correlations strongly over dipolar, monopolar, or no spatial correlations.

The last group includes an additional common-spectrum red-noise process on top of the GWB-like common-spectrum HD-correlated process. The second process is taken to have either no spatial correlations, dipolar correlations, or monopolar correlations.

3.3. Solar System Ephemeris

In the course of the GWB analysis of NANOGrav's 11 yr data set (NG11gwb), we determined that GW statistics were surprisingly sensitive to the choice of SSE, and we developed a statistical treatment of SSE uncertainties (BAYESEPHEM; Vallisneri et al. 2020), designed to harmonize GW results for SSEs ranging from JPL's DE421 (published in 2009 and based on data up to 2007) to DE436 (published in 2016, and based on data up to 2015).

This was a rather conservative choice: it would be reasonable to expect that more recent SSEs, based on larger data sets and on more sophisticated data reduction, would be more accurate —an expectation backed by the (somewhat fragmentary) error estimates offered by SSE compilers. However, our analysis showed that errors in Jupiter's orbit (which create an apparent motion of the solar system barycenter and therefore a spurious Rømer delay) dominate the GWB systematics and that Jupiter's orbit has been adjusted across DE421–DE436 by amounts ($\lesssim$50 km) comparable to or larger than the stated uncertainties. Thus, we decided to err on the side of caution, with the understanding that the Bayesian marginalization over SSE uncertainties would subtract power from the putative GWB process, as confirmed by simulations (Vallisneri et al. 2020).

Luckily, these circumstances have since changed. Jupiter's orbit is being refined with data from the NASA orbiter Juno: the latest JPL SSE (DE438, Folkner & Park 2018) fits the range and VLBI measurements from six perijoves and claims orbit accuracy a factor of 4 better than previous SSEs (i.e., $\lesssim$10 km). In addition, the longer time span of the 12.5 yr data set (NG12) reduces the degeneracy between a GWB and Jupiter's orbit (Vallisneri et al. 2020). Accordingly, we adopt DE438 as the fiducial SSE for the results reported in this paper. For completeness and verification, we also report statistics obtained with BAYESEPHEM, adopting the same treatment of NG11gwb; and with the SSE INPOP19a (Fienga et al. 2019), which incorporates range data from nine Juno perijoves.

The DE438 and INPOP19a Jupiter orbit estimates are not entirely compatible, because the underlying data sets do not overlap completely and are weighted differently; nevertheless, the orbits differ in ways that affect GWB results only slightly, which further increases our confidence in DE438. In our analysis, we used DE438 and INPOP19a without uncertainty corrections: while it is technically straightforward to constrain BAYESEPHEM using the orbital-element covariance matrices provided by the SSE authors, the resulting orbital perturbations are so small that GW results are barely affected (Vallisneri et al. 2020).

4. Gravitational-wave Background Estimates

Our Bayesian analysis of the 12.5 yr data set shows definitive evidence for the presence of a time-correlated stochastic process with a common amplitude $A_{\rm CP}$ and a common spectral index $\gamma_{\rm CP}$ across all pulsars. Given this finding, we do not quote an upper limit on a GWB amplitude as in NG9gwb and NG11gwb, but rather report the median value and 90% credible interval of $A_{\rm CP}$, as well as the $\log_{10}$ Bayes factor for a common-spectrum process versus pulsar-intrinsic red noise only. Further details of our Bayesian methodology can be found in Appendix B. In addition, we characterize the evidence for HD correlations, which we take as the crucial marker of GWB detection, by obtaining the Bayes factors between the models of Table 1.

Our results are presented in Section 4.1 and summarized in Figures 2 and 3. In Sections 4.2 and 4.3, we explore the evidence for spatial correlations further, by way of the optimal statistic (Anholm et al. 2009; Demorest et al. 2013; Chamberlin et al. 2015) and of a novel Bayesian technique that isolates the cross-correlations in the Gaussian-process likelihood. The statistical significance of our results for both the common-spectrum process and HD correlations is examined in Section 5.

4.1. Bayesian Analysis

Figure 2 shows marginalized $A_{\rm CP}$ posteriors obtained from the 12.5 yr data using a model that includes pulsar-intrinsic red

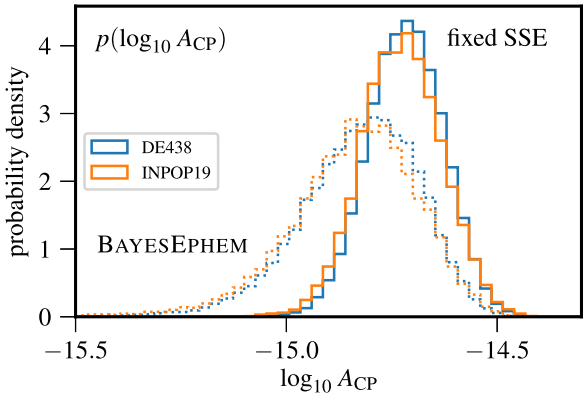

Figure 2. Bayesian posteriors for the ($f_{\rm yr} = 1\ {\rm yr}^{-1}$) amplitude $A_{\rm CP}$ of a common-spectrum process, modeled as a $\gamma = 13/3$ power law using only the lowest five-component frequencies. The posteriors are computed for the NANOGrav 12.5 yr data set using individual ephemerides (solid lines) and BAYESEPHEM (dotted). Unlike similar analyses in NG11gwb and Vallisneri et al. (2020), these posteriors, even those using BAYESEPHEM, imply a strong preference for a common-spectrum process. Results are consistent for both recent SSEs (DE438 and INPOP19a) updated with Jupiter data from the Juno mission. SSE corrections remain partially entangled with $A_{\rm CP}$. Thus, when BAYESEPHEM is applied, the distributions broaden toward lower amplitudes, shifting the peak of the distribution by ∼20%.

noise plus a spatially uncorrelated common-spectrum process with a fixed spectral index $\gamma_{\rm CP} = 13/3$. Following the discussion of Section 3.1, the common-spectrum process is represented by five sine–cosine pairs. The sine–cosine pairs are modeled to have the same power spectral density, but the values of the coefficients are independent across pulsars. By contrast, in the spatially correlated models, the coefficients are constrained to have the appropriate correlations according to the ORFs. Under fixed ephemeris DE438, the $A_{\rm CP}$ posterior has a median value of 1.92×10^{-15} with 5%–95% quantiles at 1.37–2.67×10^{-15}; the INPOP19a posterior is very close—a reassuring finding, given that past versions of the JPL and INPOP SSEs led to discrepant results (NG11gwb).

If we allow for BAYESEPHEM corrections to DE438, the $A_{\rm CP}$ posterior shifts lower, with median value of 1.53×10^{-15} and 5%–95% quantiles at 0.79–2.38×10^{-15}; the posterior for INPOP19a with BAYESEPHEM corrections is again very close. It is well understood that BAYESEPHEM will absorb power from a common-spectrum process (Roebber 2019; Vallisneri et al. 2020), but we note that this coupling weakens with increasing data set time span: it is weaker here than in the 11 yr analysis and would be even weaker with 15 yr of data (Vallisneri et al. 2020).

These peaked, compact $A_{\rm CP}$ posteriors are accompanied by large Bayes factors in favor of a spatially uncorrelated common-spectrum process versus pulsar-intrinsic pulsar red noise alone: $\log_{10}$ Bayes factor = 4.5 for DE438 and 2.7 with BAYESEPHEM. Next, we assess the evidence for spatial correlations by computing Bayes factors between the models in Table 1. Our results are summarized in Table 2 and more visually in Figure 3. There is little evidence for the addition of HD correlations ($\log_{10}$ Bayes factor = 0.64 with DE438, 0.37 with BAYESEPHEM), and the HD-correlated $A_{\rm CP}$ posteriors are very similar to those of Figure 2. By contrast, monopolar and dipolar correlations are moderately disfavored ($\log_{10}$ Bayes factor = −2.3 and −2.4, respectively, with DE438). The monopole is disfavored less under BAYESEPHEM, which may be explained by the BAYESEPHEM-reduced amplitude of the processes.

The evidence for a second common-spectrum process on top of an HD-correlated process is inconclusive. Furthermore, the amplitude posteriors for additional monopolar and dipolar processes display no clear peaks, while the posterior for an additional spatially uncorrelated process shows that power is drawn away from the HD-correlated process (which is understandable given the scant evidence for HD correlations).

We completed the same analyses with a common-spectrum model where $\gamma_{\rm CP}$ was allowed to vary. As seen in Figure 1, the posteriors on $\gamma_{\rm CP}$, while consistent with $13/3$ (≈4.33), are very broad. Under fixed ephemeris DE438, the $\gamma_{\rm CP}$ posterior from a spatially uncorrelated process has a median value of 5.52 with 5%–95% quantiles at 3.76–6.78. The amplitude posterior is larger in this case, but that is due to the inherent degeneracy between $A_{\rm CP}$ and γ. The evidence for spatial correlations in a varied-$\gamma_{\rm CP}$ model is almost identical to that reported in Table 2.

Altogether, the smaller Bayes factors in the discrimination of spatial correlations are fully expected, given that spatial correlations are encoded by the cross-terms in the interpulsar covariance matrix, which are subdominant with respect to the self-terms that drive the detection of a common-spectrum process. Nevertheless, if a GWB is truly present, the Bayes factors will continue to increase as data sets grow in time span and number of pulsars. Indeed, the trends on display here are broadly similar to the results of NG11gwb, but they have become more marked.

4.2. Optimal Statistic

The optimal statistic (Anholm et al. 2009; Demorest et al. 2013; Chamberlin et al. 2015) is a frequentist estimator of the amplitude of an HD-correlated process, built as a sum of correlations among pulsar pairs, weighted by the assumed pulsar-intrinsic and interpulsar noise covariances. It is a useful complement to Bayesian techniques, specifically for the characterization of spatial correlations. The statistic $\hat{A}^2$ is defined by Equation (7) of NG11gwb, and it is related to the GWB amplitude by $\langle \hat{A}^2 \rangle = A^2_{\rm GWB}$, where the mean is taken over an ensemble of GWB realizations of the same $A_{\rm GWB}$. The statistical significance of an observed $\hat{A}^2$ value is quantified by the corresponding S/N (see Equation (8) of NG11gwb).

Table 3 and Figure 4 summarize the optimal-statistic analysis of the 12.5 yr data set. As in NG11gwb, we computed two variants of the statistic: a fixed-noise version obtained by fixing the pulsar red-noise parameters to their maximum a posteriori values in Bayesian runs that include a spatially uncorrelated common-spectrum process; and a noise-marginalized version (Vigeland et al. 2018), which has proven more accurate when pulsars have intrinsic red noise, and which is sampled over 10,000 red-noise parameter vectors drawn from those same posteriors. For each variant, we computed versions of the statistic tailored to HD, monopolar, and dipolar spatial corrections.

We recovered similarly low S/N for all three correlation patterns, indicating that the optimal statistic cannot distinguish among them. Nevertheless, these results are markedly different from those of NG11gwb, which found no trace of correlations. The highest S/N is found for the monopolar process, which may seem to be in conflict with the Bayes factors of Table 2; however, Figure 4 shows that the corresponding amplitude estimate $\hat{A}^2$ is more than a factor of 2 lower than implied by the $A_{\rm CP}$ posterior, shown there by the dashed curve. A compatible

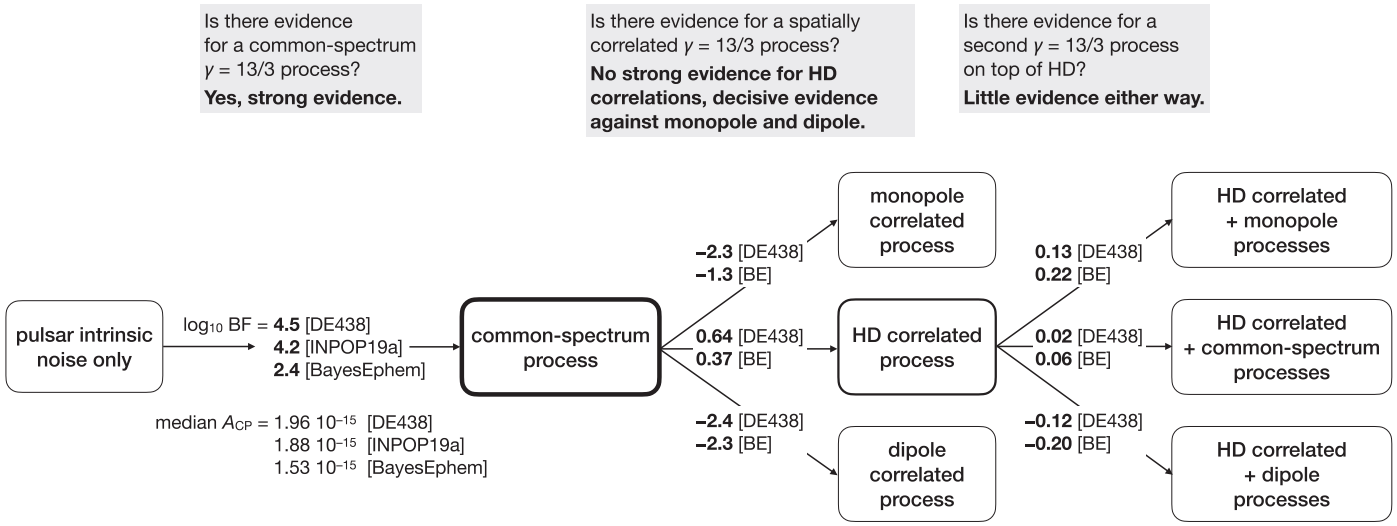

Figure 3. A visual representation of Bayesian model comparisons on the 12.5 yr data set. Each box represents a model from Table 1; arrows are annotated with the $\log_{10}$ Bayes factor between the two models that they connect, computed for both fixed and BAYESEPHEM-corrected SSE. Moving from left to right, we find strong evidence for a common-spectrum process, weak evidence for its HD correlations, moderately negative evidence for monopolar or dipolar correlations, and approximately even odds for a second common-spectrum process. The $\log_{10}$ Bayes factor between any two models can be approximated by summing the values along a path that connects them.

Table 2
Bayesian Model-comparison Scores

Ephemeris	Uncorr. Process versus Noise Only	Dipole	Mono. versus Uncorrelated Process	HD	HD+dip.	HD+mono. versus HD-correlated Process	HD+uncorr.
DE438	4.5(9)	−2.4(2)	−2.3(2)	0.64(1)	−0.116(4)	0.126(4)	0.0164(1)
BAYESEPHEM	2.4(2)	−2.3(2)	−1.3(1)	0.371(5)	−0.199(5)	0.217(6)	0.0621(4)

Note. The $\log_{10}$ Bayes factors between pairs of models from Table 1 are also visualized in Figure 3. All common-spectrum power-law processes are modeled with a fixed spectral index $\gamma = 13/3$ and with the lowest five frequency components. The digit in the parentheses gives the uncertainty on the last quoted digit.

Table 3
Optimal Statistic $\hat{A}^2$ and Corresponding S/N

Correlation	Fixed Noise		Noise Marginalized	
	$\hat{A}^2$	S/N	Mean $\hat{A}^2$	Mean S/N
HD	4×10^{-30}	2.8	$2(1) \times 10^{-30}$	1.3(8)
Monopole	9×10^{-31}	3.4	$8(3) \times 10^{-31}$	2.6(8)
Dipole	9×10^{-31}	2.4	$5(3) \times 10^{-31}$	1.2(8)

Note. The optimal statistic, $\hat{A}^2$, and corresponding S/N are computed from the 12.5 yr data set for an HD-, monopolar-, and dipolar-correlated common process modeled as a power law with fixed spectral index, $\gamma = 13/3$, using the five lowest frequency components. We show fixed intrinsic red-noise and noise-marginalized values. All are computed with fixed ephemeris DE438.

amplitude estimate is found only for the HD process. In other words, the optimal-statistic analysis is consistent with the Bayesian analysis. They agree on the presence of an HD-correlated process at the common amplitude indicated by the Bayesian analysis, and both find it strongly unlikely that there are monopolar or dipolar processes of equal amplitude. These optimal-statistic results are robust with respect to changing γ within the range recovered in Figure 1.

Figure 5 shows the angular distribution of cross-correlated power for both NG11 and NG12, as obtained by grouping pulsar pairs into angular-separation bins (with each bin hosting a similar number of pairs). The error bars show the standard deviations of angular separations and cross-correlated power within each bin. The dashed and dotted lines show the values expected theoretically from HD- and monopolar-correlated processes with amplitudes set from the measured $\hat{A}^2$ (the first column of Table 3). While errors are smaller for NG12 than for NG11, neither correlation pattern is visually apparent.

4.3. Bayesian Measures of Spatial Correlation

Inspired by the optimal statistic, we have developed two novel Bayesian schemes to assess spatial correlations. We report here on their application to the 12.5 yr data.

First, we performed Bayesian inference on a model where the uncorrelated common-spectrum process is augmented with a second HD-correlated process with autocorrelation coefficients set to zero. In other words, we decouple the amplitudes of the auto- and cross-correlation terms. The uncorrelated common-spectrum process regularizes the overall covariance matrix, which would not otherwise be positive definite with this new "off diagonal only" GWB. Figure 6 shows marginalized amplitude posteriors for the diagonal and off-diagonal processes, which appear consistent. It is however evident that cross-correlations carry much weaker information: as a matter of fact, the $\log_{10}$ Bayes factor in favor of the additional process (computed à la Savage–Dickey; see Dickey 1971) is 0.10 ± 0.01 with fixed DE438 and -0.03 ± 0.01 under BAYESEPHEM. These factors are smaller than the HD versus uncorrelated values of Table 2, arguably because the off-

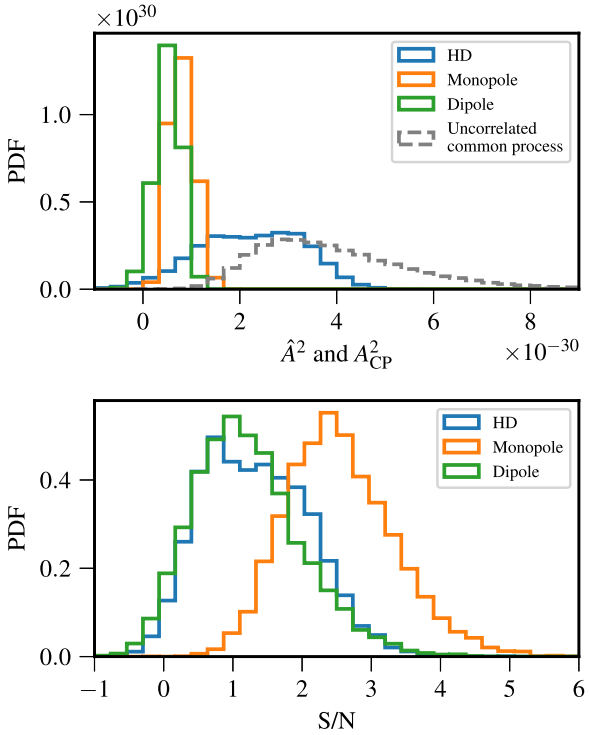

Figure 4. Distributions of the optimal statistic and S/N for HD (blue), monopole (orange), and dipole (green) spatial correlations, as induced by the posterior probability distributions of pulsar-intrinsic red-noise parameters in a Bayesian inference run that includes a spatially uncorrelated common-spectrum process. The means of each distribution are the noise-marginalized $\hat{A}^2$ given in Table 3. The top panel also shows the posterior of an uncorrelated common red process $A^2_{\rm CP}$ (dashed gray) from Figure 2 for comparison. All three cross-correlation patterns are identified in the data with modest significance, but it is only for an HD-correlated process that the amplitude estimate is compatible with the posteriors of Figure 2.

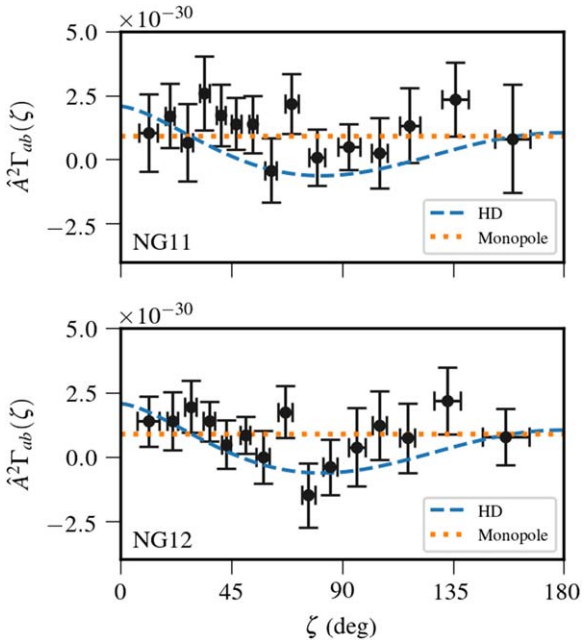

Figure 5. Average angular distribution of cross-correlated power, as estimated with the optimal statistic on the 11 yr data set (top) and 12.5 yr data set (bottom). The number of pulsar pairs in each binned point is held constant for each data set. Due to the increase in pulsars in the 12.5 yr data set, the number of pairs per bin increases accordingly. Pulsar-intrinsic red-noise amplitudes are set to their maximum posterior values from the Bayesian analysis, while the SSE is fixed to DE438. The dashed blue and dotted orange lines show the cross-correlated power predicted for HD and monopolar correlations with amplitudes $\hat{A}^2 = 4 \times 10^{-30}$ and 9×10^{-31}, respectively.

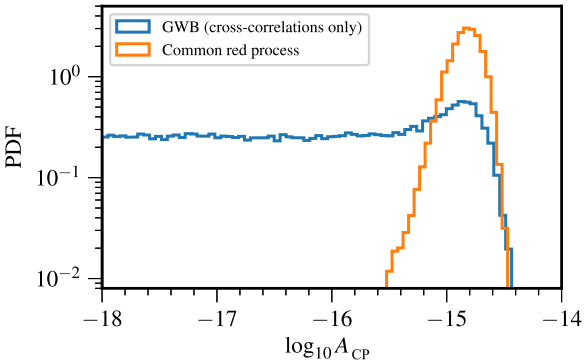

Figure 6. Bayesian amplitude posteriors in a model that includes a common-spectrum process and an off-diagonal HD-correlated process where all autocorrelation terms are set to zero (see main text of Section 4.3). The posteriors shown here are marginalized with respect to each other. The inference run includes BAYESEPHEM.

diagonal portion of the model is given the additional burden of selecting the appropriate amplitude.

Second, we performed Bayesian inference on a common-spectrum model that includes a parameterized ORF: specifically, interpulsar correlations are obtained by the spline interpolation of seven nodes spread across angular separations; node values are estimated as independent parameters with uniform priors in [−1, 1] (Taylor et al. 2013). Figure 7 shows the marginalized posteriors of the angular correlations and bears a direct comparison with Figure 5. The posteriors, although not very informative, are consistent with the HD ORF, which is overplotted in the figure. However, they are inconsistent with the monopolar ORF, also overplotted in the figure. This behavior is similar to the evidence reported in Table 2.

5. Statistical Significance

As described above, the 12.5 yr data set offers strong evidence for a spatially uncorrelated common-spectrum process across pulsars in the data set, but it favors only slightly the interpretation of this process as a GWB by way of HD interpulsar correlations. In this section, we test the robustness of the first statement by examining the contribution of each pulsar to the overall Bayes factor, and we characterize the statistical significance of the second by building virtual null distributions for the HD detection statistics. We expect that studies of both kinds will be important to establishing confidence in future detection claims.

5.1. Characterizing the Evidence for a Common-spectrum Process across the PTA

Under a model that includes a noise-like process of the common spectrum across all pulsars without interpulsar correlations, and in the absence of other physical effects linking observations across pulsars (such as ephemeris corrections), the PTA likelihood factorizes into individual pulsar terms:

$$p(\{d_j\}_N|\{\boldsymbol{\theta}_j\}_N, A_{\rm CP}) = \prod_{j=1}^{N} p(d_j|\boldsymbol{\theta}_j, A_{\rm CP}), \quad (4)$$

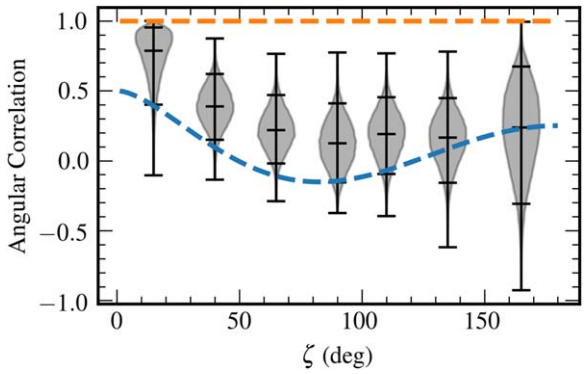

Figure 7. Bayesian reconstruction of interpulsar spatial correlations, parameterized as a seven-node spline. Violin plots show marginalized posteriors for node correlations, with medians, 5% and 95% percentiles, and extreme values. The dashed blue line shows the HD ORF expected for a GWB, while the dashed horizontal orange line shows the expected interpulsar correlation signature for a monopole systematic error, e.g., drifts in clock standards.

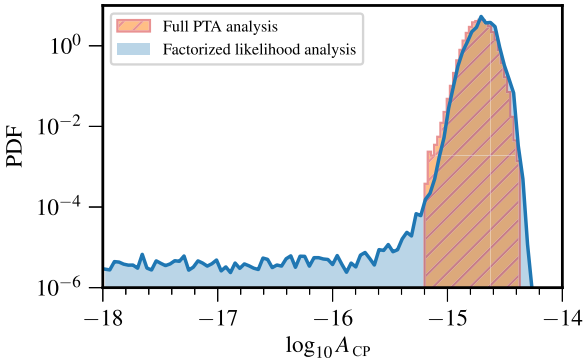

Figure 8. Marginalized $A_{\rm CP}$ posterior of a common-spectrum process modeled with a fixed $\gamma = 13/3$ power law with five component frequencies and no interpulsar correlations, as evaluated with full-PTA sampling and with the factorized-likelihood approach of Section 5.1. We fixed the ephemeris to DE438 (without corrections) and varied the white-noise hyperparameters for the factorized likelihood, but not in the full-PTA run. Note the logarithmic vertical scale, which emphasizes the very-low-density tail of the distribution; full-PTA sampling has trouble accessing that region because low $A_{\rm CP}$ requires the fine-tuning of relatively high $A_{\rm red}$ in most pulsars.

where d_j and $\boldsymbol{\theta}_j$ denote the data set and the intrinsic-noise parameters for each pulsar j, and where $A_{\rm CP}$ denotes the amplitude of the common-spectrum process.

Equation (4) suggests a trivially parallel approach to estimating the $A_{\rm CP}$ posterior: we performed independent inference runs for each pulsar, sampling timing-model parameters, pulsar-intrinsic white-noise parameters, pulsar-intrinsic red-noise parameters, as well as $A_{\rm CP}$. We adopted DE438 (without corrections) as the SSE, and we set log-uniform priors for all red-process amplitudes (as seen in Table 5). We then obtained $p(A_{\rm CP}|\{d_j\}_N)$ by multiplying the individual $p(A_{\rm CP}|d_j)$ posteriors (as represented, e.g., by kernel density estimators), while correcting for the duplication of the prior $p(A_{\rm CP})$.

As shown in Figure 8, the resulting posterior matches the analysis of Section 4, while sampling very low $A_{\rm CP}$ values more accurately. We can then evaluate the $p_{\rm all}({\rm CP})/p_{\rm all}({\rm no\ CP})$ Bayes factor in the Savage–Dickey approximation (see Dickey 1971), obtaining a value of ∼65,000, or a $\log_{10}$ Bayes factor of ∼4.8, which is broadly consistent with the transdimensional sampling estimates reported in Table 2. The agreement of the two distributions in Figure 8 validates the approximation of fixing pulsar-intrinsic white-noise hyperparameters in the full-PTA analysis, which we accepted for the sake of sampling efficiency.

In a dropout analysis (Aggarwal et al. 2019; S. Vigeland et al. 2020, in preparation) we perform inference on the joint PTA data set but introduce a binary indicator parameter for each pulsar that can turn off the common-spectrum process term in the likelihood of its data. These indicators are sampled in Monte Carlo fashion with all other parameters. The dropout factor (the number of "on" samples divided by "off" samples for a pulsar) quantifies the support offered by each pulsar to the common-signal hypothesis.

In this paper, we allow only a single pulsar to drop out at any time in the exploration of the posterior. We performed such dropout runs with fixed pulsar-intrinsic white-noise parameters and fixed ephemeris DE438; the resulting dropout factors are displayed by the blue dots of Figure 9, sorted by decreasing value. Of the 45 pulsars used in this analysis, roughly 10 have values significantly larger than 1 and (by implication) contribute most of the evidence toward the recovered common-spectrum process, 3 (notably PSR J1713+0747) disfavor that hypothesis, and prefer to "drop out," while the rest remain agnostic.

The dropout factor for each pulsar k is linked to the posterior predictive likelihood for the single-pulsar data set d_k, integrated over the $A_{\rm CP}$ posterior from all other pulsars (Wang et al. 2019):

$$\begin{aligned}{\rm ppl}_k({\rm CP}) = \int [&p(d_k|\boldsymbol{\theta}_k, A_{\rm CP}) \\ &\times p(A_{\rm CP}|\{d_{j\neq k}\}) \times p(\boldsymbol{\theta}_k)\, dA_{\rm CP}]d\boldsymbol{\theta}_k. \end{aligned} \quad (5)$$

If the likelihood factorizes per Equation (4), then the dropout factor is

$${\rm dropout}_k = \frac{p_{\rm all}({\rm CP})}{p_k({\rm no\ CP})p_{j\neq k}({\rm CP})} = \frac{{\rm ppl}_k({\rm CP})}{p_k({\rm no\ CP})}, \quad (6)$$

where $p_{\rm all}({\rm CP})$ and $p_{j\neq k}({\rm CP})$ denote the Bayesian evidence for the common-spectrum model from all pulsars together and from all pulsars excluding k, respectively; and where $p_k({\rm no\ CP})$ is the evidence for the intrinsic-noise-only model in the data from pulsar k.

The posterior predictive likelihood quantifies model support by Bayesian cross-validation: namely, the $A_{\rm CP}$ posterior obtained from $n-1$ pulsars is used to compute the likelihood of the data measured for the excluded pulsar, which acts as an out-of-sample testing data set (Wang et al. 2019). In other words, single-pulsar data sets with dropout factor larger than 1 can be predicted successfully from the $A_{\rm CP}$ posterior from all other pulsars, lending credence to the common-spectrum process model as a whole. Small dropout factors indicate problematic single-pulsar data sets or deficiencies in the global model.

Equation (6) can be recast as

$$\begin{aligned}{\rm dropout}_k = &\frac{p_k({\rm CP})}{p_k({\rm no\ CP})} \\ &\times \int \frac{p(A_{\rm CP}|\{d_{j\neq k}\})\, p(A_{\rm CP}|d_k)}{p(A_{\rm CP})}\, {\rm d}A_{\rm CP}, \end{aligned} \quad (7)$$

which allows the numerical evaluation of dropout factors from factorized likelihoods, where the Bayes factor can be computed à la Savage–Dickey from the single-pulsar analysis of each

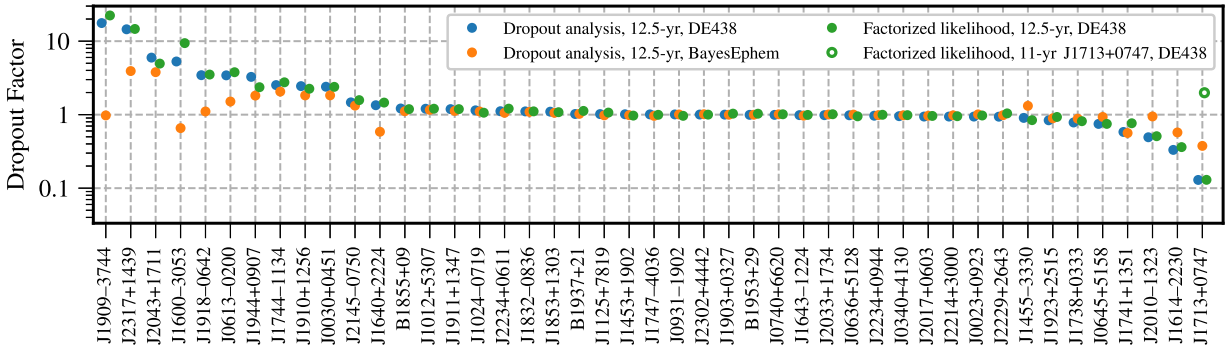

Figure 9. Characterizing the evidence from each pulsar in favor of a common-spectrum, no-correlations, stochastic process modeled as a $\gamma = 13/3$ power law. Direct dropout factors (see Equation (6)) from fixed pulsar-intrinsic white-noise fixed DE438 runs are shown as blue points; they match the estimates from variable white-noise, fixed DE438 factorized likelihoods indicated by green points. The orange points show dropout factors when we include BAYESEPHEM corrections. Most of the evidence arises from the 10 pulsars on the left, while PSRs J2010−1323, J1614−2230, and J1713+0747 remain skeptical. All of these effects are diminished by BAYESEPHEM, except for PSR J1713+0747. However, a factorized-likelihood analysis using the 11 yr version of PSR J1713+0747 shows modest evidence for the common process, as indicated by the hollow green point. This suggests that an unmodeled noise process in the 12.5 yr version of PSR J1713+0747 is preventing the pulsar from showing evidence for the common-spectrum process.

pulsar. The resulting dropout$_k$ estimates are shown as the green dots in Figure 9, and they agree closely with the direct dropout estimates.

Unlike the factorized-likelihood approximation, the dropout analysis remains possible when model parameters that correlate the likelihoods are included, such as BAYESEPHEM correction coefficients. Dropout factors for that case are shown as orange dots in Figure 9, and they can still be interpreted as indicators of the positive or negative evidence contributed by each pulsar toward the common-spectrum process hypothesis. Introducing BAYESEPHEM yields reduced factors for the first 10 pulsars, consistent with the partial absorption of GW-like residuals into ephemeris corrections (Vallisneri et al. 2020). Two of the contrarian pulsars also revert to neutral factors, but PSR J1713+0747 does not.

Altogether, the dropout analysis suggests that the strong evidence for a common-spectrum process originates from more than just a few outliers of NANOGrav pulsars. In Table 4, we summarize the timing properties of the 10 pulsars with dropout factors greater than 2. As expected, most of the evidence for the common-spectrum process comes from pulsars with longer observing baselines. We also note that of the 13 pulsars that have been observed for more than 12 years, 6 have dropout factors greater than 2, and only 1 has a dropout factor significantly less than 1 (PSR J1713+0747). Data sets for three pulsars remain somewhat inconsistent with the consensus. If this trend continues as more data are collected, it will be necessary to explain their behavior either as an expected statistical fluctuation or as the result of pulsar-specific modeling or measurement issues. Work is ongoing to develop advanced noise models specific to each pulsar (J. Simon et al. 2020, in preparation), which will provide a first quantitative assessment.

In the case of PSR J1713+0747, an unmodeled noise process may indeed be to blame. A factorized-likelihood analysis using the version of PSR J1713+0747 in the NANOGrav 11 yr data set (NG11) does show weak evidence for the common process, with a dropout factor of 2.0, indicated by a hollow green circle in Figure 9. This suggests that some issue with the timing or noise model used to describe the 12.5 yr version of PSR J1713+0747 is causing its anomalously low dropout factor. This is likely due in some part to the "second" chromatic timing event (Lam et al. 2018). An extensive study of PSR J1713+0747's noise property's response to the "first" chromatic timing event showed that it took a few years of additional data for the red-noise properties of the pulsar to return to "normal" (Hazboun et al. 2020b). If this is the primary cause of PSR J1713+0747's behavior in the 12.5 yr data set, then future data sets should show a return to previously measured intrinsic red-noise values. In which case, the pulsar would then contribute to any future detection claims.

Table 4
Timing Properties of Pulsars with High Dropout Factors

Pulsar	Dropout Factor (DE438)	Obs Time (yr)	Timing rms[a] (μs)
J1909−3744	17.6	12.7	0.061
J2317+1439	14.5	12.5	0.252
J2043+1711	6.0	6.0	0.151
J1600−3053	5.3	9.6	0.245
J1918−0612	3.4	12.7	0.299
J0613−0200	3.4	12.3	0.178
J1944+0907	3.3	9.3	0.365
J1744+1134	2.5	12.9	0.307
J1910+1256	2.4	8.3	0.187
J0030+0451	2.4	12.4	0.200

Notes. The 10 pulsars that show the strongest evidence for a common-spectrum process include many pulsars with long observational baselines and low timing rms, as expected.
[a] Weighted rms of epoch-averaged postfit timing residuals, excluding red-noise contributions. See Table 3 of NG12.

5.2. *Characterizing the Statistical Significance of Hellings–Downs Correlations*

Formally, it is the posterior odds ratio itself that relays the data's support for each model. What it does not tell you is how often noise processes alone could manifest an odds ratio as large as the data give. While arbitrary rules of thumb have been developed to interpret odds ratios (e.g., Kass & Raftery 1995; Jeffreys 1998), this interpretation is highly problem-specific. However, most analysts would agree that ratios $\sim$1 are inconclusive, while very large or small ratios point to a strong preference for either model. In classical hypothesis testing, one computes a detection statistic from the data suspected to contain a signal, then compares the value of the statistic with its background distribution, computed over a population of data

Table 5
Prior Distributions Used in All Analyses Performed in This Paper

Parameter	Description	Prior	Comments
		White Noise	
E_k	EFAC per backend/receiver system	Uniform [0, 10]	single-pulsar analysis only
Q_k (s)	EQUAD per backend/receiver system	log-uniform [−8.5, −5]	single-pulsar analysis only
J_k (s)	ECORR per backend/receiver system	log-uniform [−8.5, −5]	single-pulsar analysis only
		Red Noise	
$A_{\rm red}$	log-Uniform [−20, −11]	one parameter per pulsar	
$\gamma_{\rm red}$	red-noise power-law spectral index	Uniform [0, 7]	one parameter per pulsar
		Common Process, Free Spectrum	
ρ_i (s^2)	power-spectrum coefficients at $f = i/T$	uniform in $\rho_i^{1/2}$ [10^{-18}, 10^{-8}][a]	one parameter per frequency
		Common Process, Broken-power-law Spectrum	
$A_{\rm CP}$	broken power-law amplitude	log-uniform [−18, −14] ($\gamma_{\rm CP} = 13/3$)	one parameter for PTA
		log-uniform [−18, −11] ($\gamma_{\rm CP}$ varied)	one parameter for PTA
$\gamma_{\rm CP}$	broken-power-law low-freq. spectral index	delta function ($\gamma_{\rm common} = 13/3$)	fixed
		uniform [0,7]	one parameter per PTA
δ	broken-power-law high-freq. spectral index	delta function ($\delta = 0$)	fixed
$f_{\rm bend}$ (Hz)	broken-power-law bend frequency	log-uniform [−8.7,−7]	one parameter for PTA
		Common Process, Power-law Spectrum	
$A_{\rm CP}$	common-process strain amplitude	log-uniform [−18, −14] ($\gamma_{\rm CP} = 13/3$)	one parameter for PTA
		log-uniform [−18, −11]($\gamma_{\rm CP}$ varied)	one parameter for PTA
$\gamma_{\rm CP}$	common-process power-law spectral index	delta function ($\gamma_{\rm CP} = 13/3$)	fixed
		uniform [0,7]	one parameter for PTA
		BAYESEPHEM	
$z_{\rm drift}$ (rad yr^{-1})	drift rate of Earth's orbit about the ecliptic z-axis	uniform [-10^{-9}, 10^{-9}]	one parameter for PTA
$\Delta M_{\rm jupiter}$ ($M_\odot$)	perturbation to Jupiter's mass	$\mathcal{N}(0, 1.55 \times 10^{-11})$	one parameter for PTA
$\Delta M_{\rm saturn}$ ($M_\odot$)	perturbation to Saturn's mass	$\mathcal{N}(0, 8.17 \times 10^{-12})$	one parameter for PTA
$\Delta M_{\rm uranus}$ ($M_\odot$)	perturbation to Uranus' mass	$\mathcal{N}(0, 5.72 \times 10^{-11})$	one parameter for PTA
$\Delta M_{\rm neptune}$ ($M_\odot$)	perturbation to Neptune's mass	$\mathcal{N}(0, 7.96 \times 10^{-11})$	one parameter for PTA
PCA$_i$	ith PCA component of Jupiter's orbit	uniform [−0.05, 0.05]	six parameters for PTA

sets known to host no signal and thus representing the null hypothesis. The percentile of the observed detection statistic within the background distribution is known as the p value; it quantifies how incompatible the data are with the null hypothesis (but not the probability that the hypothesis of interest is true).

The problem for GW detectors is that it is not possible to construct the background distribution by physically turning off sensitivity to GWs. However, one can operate on the data. For the coincident detection of transient GW signals with ground-based observatories, the null model is realized by applying relative time shifts to the time series of detection statistics from multiple detectors, thus removing the very possibility of coincidence. Similar techniques can be applied to the detection of HD correlations in PTA data sets.

Several methods have been developed to perform a frequentist study of the null hypothesis distribution in PTAs (Cornish & Sampson 2016; Taylor et al. 2017a); the relevant null hypothesis is that of a red process with identical spectral properties in all pulsars, but without any GW-induced interpulsar correlations (our so-called common red process). By performing repeated trials of spatial-correlation template scrambles ("sky scrambles") and Fourier basis phase offsets ("phase shifts"), we can effectively null any spatial correlations in the true data set and construct a distribution of our detection statistic (whether frequentist S/N or Bayesian odds ratio) under the null hypothesis. It is with these null distributions that we obtain the p value of our measured statistic.

In a phase-shift analysis, random phase shifts are inserted in the Fourier basis components that describe the GWB process in each pulsar, thus breaking any interpulsar correlations that may be present in the data (Taylor et al. 2017a). Detection statistics are then computed using both frequentist (i.e., the noise-marginalized mean-S/N optimal statistic) and Bayesian (i.e., the Bayes factor for an HD-correlated model versus a common-spectrum but spatially uncorrelated model) analyses from 1000 and 300 realizations (respectively) of the phase shifts. The resulting distributions are shown in Figures 10 and 11. The p values (in this case, the fraction of background samples with statistic higher than observed for the undisturbed model) are 0.091 and 0.013.

In a sky-scramble analysis, the positions of the pulsars used to compute the expected HD correlations are randomized (Cornish & Sampson 2016; Taylor et al. 2017a), under the requirement that the scrambled ORFs have minimal similarity to the true function.[48] Again, we compute both frequentist and Bayesian HD detection statistics over large sets of realizations: the resulting background distributions are shown in Figures 10

[48] Specifically, we measure the match statistic $\bar{M}$ between the ORFs Γ_{ab} and Γ'_{ab} (Taylor et al. 2017a):

$$\bar{M} = \frac{\sum_{a,b\neq a}\Gamma_{ab}\Gamma'_{ab}}{\sqrt{\left(\sum_{a,b\neq a}\Gamma_{ab}\Gamma_{ab}\right)\left(\sum_{a,b\neq a}\Gamma'_{ab}\Gamma'_{ab}\right)}}, \quad (8)$$

where a and b index the array pulsars and require that $\bar{M} < 0.1$.

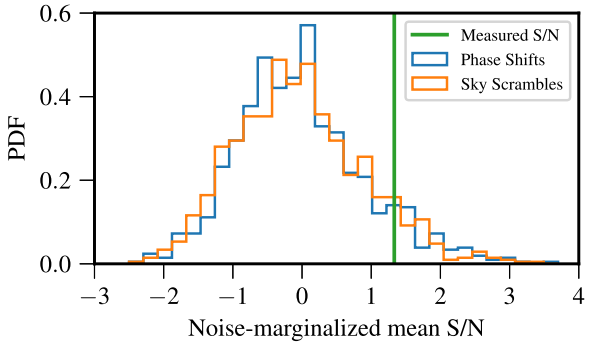

Figure 10. Distribution of the noise-marginalized optimal-statistic mean S/N for 1000 phase shifts (blue curve) and 1000 sky scrambles (orange curve). The vertical green line marks the mean S/N measured in the unperturbed model. Higher mean values of the S/N are obtained in 91 phase shifts ($p = 0.091$) and 82 sky scrambles ($p = 0.082$).

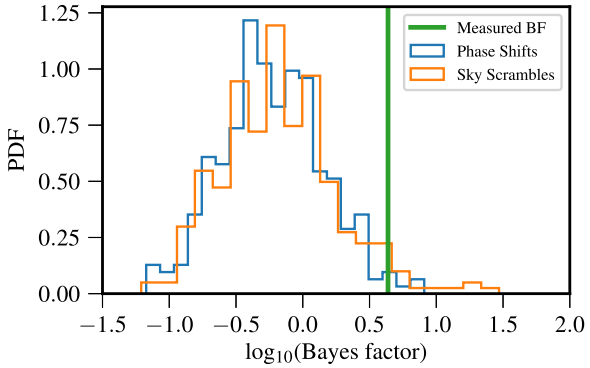

Figure 11. Distribution of the correlated vs. uncorrelated common-process Bayes factor for 300 phase shifts (blue curve) and 300 sky scrambles (orange curve). The vertical green line marks the Bayes factor computed in the unperturbed model. Higher Bayes ratios are obtained in 4 phase shifts ($p = 0.013$) and 13 sky scrambles ($p = 0.043$). The small numbers indicated that statistical error may be large in the p-value estimates.

and 11. The optimal-statistic p value agrees closely with its phase-shift counterpart; the Bayes factor p value is higher, but the small-number error is likely to be significant.

All of these p values hover around 5%, which is much higher than the 3σ ("evidence") and 5σ ("discovery") standards of particle physics, corresponding to $p = 0.001$ and 3×10^{-7}, respectively. Nevertheless, progressively smaller p values for future data sets would indicate that compelling evidence is accumulating.

6. Discussion

As reported in Section 4.1, the $A_{\rm CP}$ posterior has significant support above the upper limits reported in our GWB searches in the 11 yr and 9 yr data sets (NG9gwb; NG11gwb); in fact, almost the entire posterior sits above the most stringent upper limit in the literature ($A_{\rm GWB} < 1 \times 10^{-15}$; Shannon et al. 2015). Without a reanalysis of the data presented in Shannon et al. (2015), which is beyond the scope of this work, we cannot fully explain the discrepancy between the results presented in this paper and the upper limit quoted in Shannon et al. (2015). A revised analysis of the PPTA data is planned as a part of an upcoming IPTA publication using the DR2 combined data set (Perera et al. 2019); preliminary results show broad consistency with this work. However, we note that the Shannon et al. constraint relies on four pulsars, whereas at least 10 pulsars in the NG12 data support a common-spectrum process (see Figure 9); furthermore, the Shannon et al. analysis adopts the DE421 SSE, which, even with NANOGrav data, yields a lower upper limit than later SSEs (Arzoumanian et al. 2018a; Vallisneri et al. 2020).

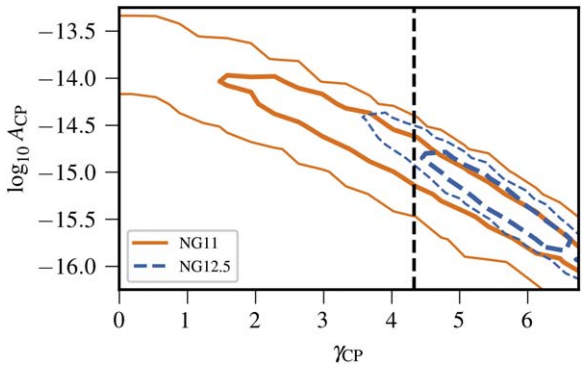

Figure 12. Common-spectrum process parameter posteriors for the NG12 (dashed curves) and NG11gwb (solid) data sets, as estimated with a five-frequency power-law model under DE438. For each data set, the two curves trace 1σ and 2σ contours, which appear entirely consistent. The dashed vertical line marks $\gamma = 13/3$, as expected for GWB from SMBHBs.

In Section 6.1, we discuss in detail the discrepancy between the published NG11gwb results and those reported in this paper and find an explanation in the choice of Bayesian prior for the amplitude $A_{\rm red}$ of pulsar-intrinsic red-noise processes (Hazboun et al. 2020a). While we focus our discussion solely on NANOGrav's previous GWB analyses, we expect the conclusions to apply broadly to all pulsar-timing data sets and analyses. While the GWB attribution of the common-spectrum process remains inconclusive, in Section 6.2 we consider the broad astrophysical implications of a GWB at the levels encompassed by the $A_{\rm CP}$ posterior. In Section 6.3, we describe the next steps for NANOGrav GWB searches as well as our expectations for the growth of spatial correlations in future data sets.

6.1. Comparison of 11 yr and 12.5 yr Results

We recognize that the common-spectrum amplitude estimated from the 12.5 yr data set ($1.4–2.7 \times 10^{-15}$) may seem surprising when compared to the Bayesian upper limits quoted from analyses of earlier data (1.45×10^{-15} in NG11gwb and 1.5×10^{-15} in NG9gwb). First, we note that applying the fiducial analysis of this paper (common-spectrum uncorrelated process under DE438) to the 11 yr data set results in $A_{\rm CP}$–$\gamma_{\rm CP}$ posteriors that are entirely consistent with those reported here, as shown in Figure 12.

The remaining dissonance between the earlier upper limits and the findings of this paper is explained by examining the structure of our analysis. The strength of the Bayesian approach to PTA searches is that it allows for simultaneous modeling of multiple time-correlated processes present in the data. Within the construction of our analysis, amplitude estimates for one such process are sensitive to the priors assumed for the others, especially when the process of interest is still below the threshold of positive detection.

Looking at the 11 yr upper limit specifically (which was quoted as 1.34×10^{-15} for a spatially uncorrelated common-spectrum process in NG11gwb), we note that introducing BAYESEPHEM corrections with unconstrained priors on Jupiter's orbital perturbation parameters would have necessarily

absorbed power from a common-spectrum process, if such a process was present. Correspondingly, the 11 yr upper limit rises to 1.94×10^{-15} if we take DE438 as the fiducial SSE, without corrections (Vallisneri et al. 2020).

Even more important, the Bayesian upper limits in NG9gwb and NG11gwb were computed by placing a uniform prior on the amplitude of pulsar-intrinsic red noise, which amounts to assuming that loud intrinsic noise is typical among PTA pulsars, rather than the exception, as suggested by the estimates in this paper. Doing so is conservative with respect to detecting a GWB, but it has the effect of depressing upper limits. As discussed in Hazboun et al. (2020a), simulations show that injecting a common-spectrum stochastic signal in synthetic data sets leads to 95% upper limits lower than $A_{\rm GWB}^{\rm inj}$ in 50% of data realizations, if the intrinsic red noise is given a uniform amplitude prior.

Reweighting the 11 yr upper limit with a log-uniform prior on intrinsic-noise amplitudes yields 2.4×10^{-15} under DE438 and 2.1×10^{-15} with BAYESEPHEM. Both values are more consistent with the findings of this paper. The differences in data reduction and in the treatment of white noise between 11 yr and 12.5 yr data sets (discussed in Section 2.3) seem to account for the remaining distance, but those differences are very challenging to evaluate formally, so we do not address them further here.

Altogether, this discussion suggests that past Bayesian upper limits from PTAs may have been overinterpreted in astrophysical terms. Those limits were indeed correct within the Bayesian logic, but they were necessarily affected by our uncertain assumptions. If future data sets bring about a confident GWB detection, our astrophysical conclusions will finally rest on a much stronger basis.

6.2. Astrophysical Implications

The first hint of a signal from our analysis of NG12 is indeed tantalizing. However, without definite evidence for HD correlations in the recovered common-spectrum process, there is little we can say about the physical origin of this signal. Models that give rise to a GWB in the nanohertz frequency range (∼1–100 nHz) through either primordial GWs from inflation (Grishchuk 1975; Lasky et al. 2016), bursts from networks of cosmic strings (Siemens et al. 2007; Blanco-Pillado et al. 2018), or the mergers of SMBHBs (Rajagopal & Romani 1995; Phinney 2001; Jaffe & Backer 2003; Wyithe & Loeb 2003) have been proposed. Black hole mergers are likely the most-studied source, though what fraction (if any) of galaxy mergers is able to produce coalescing SMBHBs is virtually unconstrained. If the common-spectrum process is due to SMBHBs, it would be the first definitive demonstration that SMBHBs are able to form, reach subparsec separations, and eventually coalesce due to GW emission.

While the recovered amplitude for the common-spectrum process in this data set is larger than the upper limit on a stochastic GWB quoted in NG11gwb, the qualitative astrophysical conclusions reported there apply to this data set as well (see Section 5 of NG11gwb). We note also that the amplitude posteriors found here can accommodate many GWB models and assumptions (such as the Kormendy & Ho measurement of the $M_{\rm BH}$–$M_{\rm bulge}$ relationship) that had previously been in tension with PTA upper limits.

The cosmic history of SMBHB mergers is encoded in the shape and amplitude of the GWB strain spectrum they produce (Sesana 2013; McWilliams et al. 2014; Ravi et al. 2014; Sampson et al. 2015; Middleton et al. 2016; Chen et al. 2017, 2019; Kelley et al. 2017; Taylor et al. 2017b; Mingarelli 2019). At the lower end of the nanohertz band, signs of the binary-hardening mechanism may still be present, and we refer the reader to Section 5 of NG11gwb and references therein for further details. The overall amplitude of the GWB spectrum is determined not only by the number of binaries able to reach the relevant orbital frequencies, but also their distribution of masses (Simon & Burke-Spolaor 2016). The GWB amplitude is relatively insensitive to the redshift distribution of sources (Phinney 2001) except at the highest frequencies, which are affected even more by the local number density and eccentricity distribution of sources (Sesana 2013; Kelley et al. 2017). Additionally, the amplitude recovered in this paper, if assumed to be primarily due to a GWB, may imply that the black hole mass function is underestimated, specifically when extrapolated from observations of the local supermassive black hole population (Zhu et al. 2019).

Last, beyond the marginal evidence for HD correlations, we find a broad posterior for the spectral slope γ of the common-spectrum process when we allow γ to vary. Therefore, the emerging signal could also be attributed to one of the other cosmological sources capable of producing a nanohertz GWB. The predicted spectral index for these is only slightly different from the SMBHB value of $13/3$ (≈4.33): it is 5 for a primordial GWB (Grishchuk 2005) and $16/3$ (≈5.33) for cosmic strings (Ölmez et al. 2010). Data sets with longer time spans and more pulsars will allow for precise parameter estimation in addition to providing confidence toward or against GWB detection.

6.3. Expectations for the Future

The analysis of NANOGrav pulsar-timing data presented in this paper is the first PTA search to show definite evidence for a common-spectrum stochastic signal across an array of pulsars. However, evidence for the tell-tale quadrupolar HD correlations is currently lacking, and there are other potential contributors to a common-spectrum process. A majority of the pulsars with long observational baselines show the strongest evidence for a common-spectrum process; this subset of pulsars could be starting to show similar spin noise with a consistent spectral index. However, it is unlikely that strong spin noise would appear at a similar amplitude in all MSPs (Lam et al. 2017). Additionally, the per-pulsar evidence is significantly reduced when we apply BAYESEPHEM, as expected; there remain other solar system effects for which we do not directly account, such as planetary Shapiro delay (Hobbs & Edwards 2012), that could contribute to the common-spectrum process. Finally, there are other sources of systematic noise that we may have uncovered (Tiburzi et al. 2016) and further potential for sources yet to be diagnosed, all of which would require further study to isolate. Thus, attributing the signal uncovered in this work to an astrophysical GWB will necessitate verification with independent pipelines on larger (and/or independent) data sets.

One avenue to validate the processing of timing observations will be the analysis of the "wideband" version of NANOGrav's 12.5 yr data set, which is produced by a significantly different reduction pipeline (Alam et al. 2021b). A preliminary analysis of wideband data using the techniques of this paper shows results consistent with those detailed here. Additionally, our

treatment and understanding of pulsar-intrinsic noise will be enhanced soon with the adoption of advanced noise models tailored to each pulsar (J. Simon et al. 2020, in preparation), which include more powerful descriptions of dispersion-measure oscillations among other enhancements.

In the medium term, NANOGrav is compiling its next data set, which adds multiple years of observations and many new pulsars to NG12, some of which will have baselines long enough to be incorporated in GW searches. If we assume optimistically that the common-spectrum signal identified here is indeed astrophysical, the optimal-statistic S/N should then grow by a factor of a few (Pol et al. 2020).

Finally, data from the other PTA collaborations will play an important role: the second IPTA data release (Perera et al. 2019) includes the 9 yr NANOGrav data set alongside EPTA and PPTA timing observations. The analysis of this joint data set is ongoing, and early results are again consistent with those discussed here. Thus, future data sets will be strong arbiters of the astrophysical interpretation of our findings.

NANOGrav's pursuit of a stochastic GWB detection has hardly been linear. In NG11gwb, we reanalyzed the 9 yr data set using BAYESEPHEM and updated the results reported in NG9gwb to reflect our new understanding of ephemeris errors. In this work, we reweighted the 11 yr analysis to account for the emerging physical picture of PTA data quality. While we cannot foresee how we will revise this 12.5 yr analysis in light of the 15 yr data set, the ouroboric nature of hierarchical Bayesian inference will undoubtedly require some refinements. The LIGO–Virgo discovery of high-frequency, transient GWs from stellar black hole binaries appeared meteorically, with incontrovertible statistical significance. By contrast, the PTA discovery of very-low-frequency GWs from SMBHBs will emerge from the gradual and not always monotonic accumulation of evidence and arguments. Still, our GW vista on the unseen universe continues to get brighter.

This work has been carried out by the NANOGrav collaboration, which is part of the International Pulsar Timing Array. We thank the members of the IPTA Steering Committee whose comments helped improve and clarify the manuscript. We also thank the anonymous reviewers for useful suggestions and comments, which improved the quality of the manuscript. The NANOGrav project receives support from National Science Foundation (NSF) Physics Frontiers Center award number 1430284. The Arecibo Observatory is a facility of the NSF operated under cooperative agreement (#AST-1744119) by the University of Central Florida (UCF) in alliance with Universidad Ana G. Méndez (UAGM) and Yang Enterprises (YEI), Inc. The Green Bank Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. A majority of the computational work was performed on the Nemo cluster at UWM supported by NSF grant No. 0923409. This work made use of the Super Computing System (Spruce Knob) at WVU, which are funded in part by the National Science Foundation EPSCoR Research Infrastructure Improvement Cooperative Agreement #1003907, the state of West Virginia (WVEPSCoR via the Higher Education Policy Commission) and WVU. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The Flatiron Institute is supported by the Simons Foundation. Pulsar research at UBC is supported by an NSERC Discovery Grant and by the Canadian Institute for Advanced Research. J.S. and M.V. acknowledge support from the JPL RTD program. S.B.S. acknowledges support for this work from NSF grants #1458952 and #1815664. S.B.S. is a CIFAR Azrieli Global Scholar in the Gravity and the Extreme Universe program. T.T. P. acknowledges support from the MTA-ELTE Extragalactic Astrophysics Research Group, funded by the Hungarian Academy of Sciences (Magyar Tudományos Akadémia), which was used during the development of this research.

Facilities: Arecibo, GBT.

Software: `ENTERPRISE` (Ellis et al. 2019), `enterprise_extensions` (Taylor et al. 2018), `hasasia` (Hazboun et al. 2019), `libstempo` (Vallisneri 2020), `matplotlib` (Hunter 2007), `PTMCMC` (Ellis & van Haasteren 2017), TEMPO, TEMPO2 (Nice et al. 2015), PINT (Luo et al. 2019).

Author Contributions

An alphabetical-order author list was used for this paper in recognition of the fact that a large, decade-timescale project such as NANOGrav is necessarily the result of the work of many people. All authors contributed to the activities of the NANOGrav collaboration leading to the work presented here and reviewed the manuscript, text, and figures prior to the paper's submission. Additional specific contributions to this paper are as follows. Z.A., H.B., P.R.B., H.T.C., M.E.D., P.B. D., T.D., J.A.E., R.D.F., E.C.F., E.F., N.G.-D., P.A.G., D.C.G., M.L.J., M.T.L., D.R.L., R.S.L., J.L., M.A.M., C.N., D.J.N., T. T.P., N.S.P., S.M.R., K.S., I.H.S., R.S., J.K.S., R.S., and S.J.V. developed the 12.5 yr data set through a combination of observations, arrival time calculations, data checks and refinements, and timing-model development and analysis; additional specific contributions to the data set are summarized in NG12. J.S. coordinated the writing of the paper and led the search. P.T.B., P.R.B., S.C., J.A.E., J.S.H., A.M.H., K.I., A.R. K., N.L., N.S.P., J.S., K.S., J.P.S., S.R.T., J.E.T., S.J.V., and C. A.W. performed different analyses associated with this work, including exploratory analyses on preliminary versions of the data set. S.C., J.S.H., N.L., J.S., J.P.S., X.S., and S.R.T. developed and tested new noise models and created a detailed noise portrait of the 12.5 yr data set. N.S.P. ran the injection analysis in the 11 yr data set. J.S.H. performed the noise parameter comparison between the 11 yr data set and the 11 yr "slice" of the 12.5 yr data set and produced the hasasia calculations. S.R.T. developed and executed new Bayesian schemes to assess spatial correlations. J.S.H., S.R.T., M.V., and S.J.V. developed and performed new tests of the statistical significance of the common-spectrum process. N.J.C., X.S., and M.V. provided feedback on searches and new analysis techniques. L.Z.K., C.M.F.M., and J.S. developed the astrophysical interpretation. J.S.H., N.S.P., J.S., S.R.T., M.V., and S.J.V. prepared the figures and tables. J.S.H., L.Z.K., C.M.F. M., N.S.P., J.S., S.R.T., M.V., and S.J.V. wrote the paper and collected the bibliography.

Appendix A
Injection Analysis of the NANOGrav 11 yr Data Set

To test the response of our real data sets to the presence of a stochastic GWB, we inject a range of GWB amplitudes directly

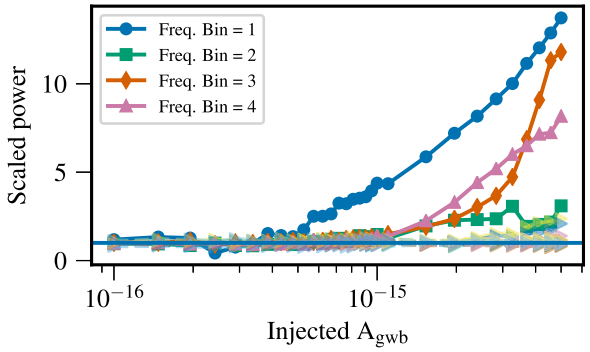

Figure 13. Response of each frequency from a common free-spectrum model to the presence of an injected GWB into the 11 yr data set (NG11) as a function of the injected GWB amplitude. The x-axis shows the injected GWB amplitude, while the y-axis shows the mean ratio across four realizations of the GWB of the average power in each frequency bin scaled to the mean power in that bin at an injected amplitude of $A = 10^{-16}$. The lowest frequency bin responds to the GWB at much smaller injected amplitudes than the other bins, while the lowest four frequency bins have the strongest response to the presence of the injected GWB at larger amplitudes.

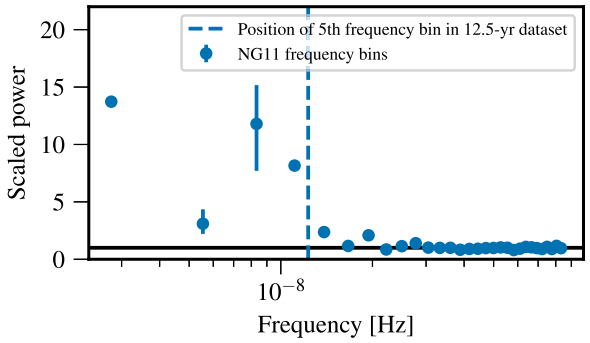

Figure 14. Response of a common free-spectrum model's red-noise Fourier-domain components to a GWB injected in the 11 yr data set (NG11). We plot the component frequency along the horizontal axis and the ratios of the mean estimated component power between injection amplitudes $A_{CP} = 5 \times 10^{-15}$ and $A_{CP} = 10^{-16}$ along the vertical. Clearly, the response to an increasing GWB amplitude is limited to the first few bins. See Appendix A for more details.

into the 11 yr data set (NG11). We use the 11 yr data set rather than the current 12.5 yr data set because it does not contain any significant common-spectrum processes and so the GWB injection is able to be cleanly recovered. While retaining the TOAs and their corresponding errors from NG11, we inject a stochastic GWB (Chamberlin et al. 2015) using functionality in the LIBSTEMPO software package. Using a power-law model with a spectral index of $\alpha = -2/3$ (i.e., $\gamma = 13/3$), we create 10 data set realizations for each characteristic strain amplitude in the range $10^{-16} \leqslant A_{GWB} \leqslant 5 \times 10^{-15}$. We analyze all realizations with our full detection pipeline. While the complete results of this analysis will be reported in an upcoming publication, here we concentrate on the spectral response of NG11 to the presence of the stochastic GWB.

As stated in Section 3.1, we calculate the power in each frequency bin using the free-spectrum model (see Section 3.2) without including HD correlations or BAYESEPHEM. In Figure 14, we show the ratio of power recovered by each frequency bin between an injection of $A = 5 \times 10^{-15}$ and $A = 1 \times 10^{-16}$. As we can see, the lowest four frequency bins are the most responsive to the presence of a power law GWB in the data set.

We also examine the evolution of the power in each frequency bin as a function of the injected amplitude. Figure 13 shows the evolution of the power in each frequency bin, which is scaled to the power in that bin at an injected amplitude of $A = 10^{-16}$. Due to its power-law nature, the GWB affects the lowest frequency bin at amplitudes much smaller than that for the higher-frequency bins. We see again that the lowest four frequency bins are the ones that are most reactive to the presence of a GWB in the data set. This result provides further confirmation that using the five lowest frequencies is sufficient to recover a GWB in the 12.5 yr data set (Section 3.1).

Appendix B
Bayesian Methods

We used Markov Chain Monte Carlo (MCMC) methods to stochastically sample the joint posterior of our model parameter spaces and use Monte Carlo integration to deduce marginalized distributions, where $\int f(\theta)p(\theta|d)d\theta \approx \langle f(\theta_i)\rangle$ for the integral of an arbitrary function $f(\theta)$ over the posterior $p(\theta|d)$ of which the samples $\{\theta_i\}$ are randomly drawn. Where necessary, we estimated the uncertainty on the marginalized posterior value to be the Monte Carlo sampling error of the location $\hat{\theta}_x$ of the xth quantile:

$$\frac{\sqrt{x(1-x)/N}}{p(\theta = \hat{\theta}_x|d)}, \tag{B1}$$

where N is the number of (quasi-)independent samples in our MCMC chain (Wilcox 2012).

As described in NG11gwb, we employ two techniques for model selection based on the relationship between the competing models. For nested models that compare the additional presence of a signal to that of noise alone, we used the Savage–Dickey approximation (Dickey 1971). This requires adequate sampling coverage of low-amplitude posterior regions in order to compute the Savage–Dickey density ratio, which corresponds to the prior to posterior density at zero amplitude: Bayes factor $= p(A = 0)/p(A = 0|d)$. In practice, this means that the method is only useful for moderate model odds contrasts, and while this was used extensively in NG11gwb, the strength of the recovered signal in this paper exceeds the reliability of the Savage–Dickey approximation without additional sampling strategies to explore the low-amplitude posterior region. For disjoint models, models that are not easily distinguished parametrically, and indeed all model selection in this paper, we used the product-space method (Carlin & Chib 1995; Godsill 2001; Hee et al. 2015; Taylor et al. 2020). This recasts model selection as a parameter estimation problem, introducing a model-indexing variable that is sampled along with the parameters of the competing models and which controls which model likelihood is active at each MCMC iteration. The ratio of samples spent in each bin of the model-indexing variable returns the posterior odds ratio between models. The efficiency of model transitions is controlled by our prior model probabilities, which we usually set to be equal. However, one can improve the odds ratio computation by performing a pilot run, whose odds ratio estimate can be used to reweight the models in a follow-up run. This will ensure more equitable chain visitation to each model, after which the model index posterior is reweighted back to the true model contrast.

Appendix C
Software

We used the software packages `enterprise` (Ellis et al. 2019) and `enterprise_extensions` (Taylor et al. 2018) to perform the Bayesian and frequentist searches. These packages implement the signal models, likelihood, and priors. Our Bayesian priors for all parameters are described in Table 5. We used the software package `PTMCMCSampler` (Ellis & van Haasteren 2017) to perform the MCMC for the Bayesian searches. We primarily used adaptive Metropolis and differential evolution jump proposals. For some analyses, we used draws from empirical distributions to sample the pulsars' red-noise parameters, with the empirical distributions constructed from posteriors obtained from previous Bayesian analyses. These draws significantly decreased the number of samples needed for the pulsars' red-noise parameters to burn in. This technique was first used to analyze the 11 yr data set for GWs from individual SMBHBs, and a detailed description can be found in Appendix B of Aggarwal et al. (2019).

ORCID iDs

Paul T. Baker https://orcid.org/0000-0003-2745-753X
Harsha Blumer https://orcid.org/0000-0003-4046-884X
Bence Bécsy https://orcid.org/0000-0003-0909-5563
Adam Brazier https://orcid.org/0000-0001-6341-717
Paul R. Brook https://orcid.org/0000-0003-3053-6538
Sarah Burke-Spolaor https://orcid.org/0000-0003-4052-7838
Shami Chatterjee https://orcid.org/0000-0002-2878-1502
Siyuan Chen https://orcid.org/0000-0002-3118-5963
James M. Cordes https://orcid.org/0000-0002-4049-1882
Neil J. Cornish https://orcid.org/0000-0002-7435-0869
Fronefield Crawford https://orcid.org/0000-0002-2578-0360
H. Thankful Cromartie https://orcid.org/0000-0002-6039-692X
Megan E. DeCesar https://orcid.org/0000-0002-2185-1790
Paul B. Demorest https://orcid.org/0000-0002-6664-965X
Timothy Dolch https://orcid.org/0000-0001-8885-6388
Elizabeth C. Ferrara https://orcid.org/0000-0001-7828-7708
William Fiore https://orcid.org/0000-0001-5645-5336
Emmanuel Fonseca https://orcid.org/0000-0001-8384-5049
Nathan Garver-Daniels https://orcid.org/0000-0001-6166-9646
Peter A. Gentile https://orcid.org/0000-0001-8158-683X
Deborah C. Good https://orcid.org/0000-0003-1884-348X
Jeffrey S. Hazboun https://orcid.org/0000-0003-2742-3321
A. Miguel Holgado https://orcid.org/0000-0003-4143-8132
Ross J. Jennings https://orcid.org/0000-0003-1082-2342
Megan L. Jones https://orcid.org/0000-0001-6607-3710
Andrew R. Kaiser https://orcid.org/0000-0002-3654-980X
David L. Kaplan https://orcid.org/0000-0001-6295-2881
Luke Zoltan Kelley https://orcid.org/0000-0002-6625-6450
Joey Shapiro Key https://orcid.org/0000-0003-0123-7600
Michael T. Lam https://orcid.org/0000-0003-0721-651X
Duncan R. Lorimer https://orcid.org/0000-0003-1301-966X
Jing Luo https://orcid.org/0000-0001-5373-5914
Ryan S. Lynch https://orcid.org/0000-0001-5229-7430
Dustin R. Madison https://orcid.org/0000-0003-2285-0404
Maura A. McLaughlin https://orcid.org/0000-0001-7697-7422
Chiara M. F. Mingarelli https://orcid.org/0000-0002-4307-1322
Cherry Ng https://orcid.org/0000-0002-3616-5160
David J. Nice https://orcid.org/0000-0002-6709-2566
Timothy T. Pennucci https://orcid.org/0000-0001-5465-2889
Nihan S. Pol https://orcid.org/0000-0002-8826-1285
Scott M. Ransom https://orcid.org/0000-0001-5799-9714
Paul S. Ray https://orcid.org/0000-0002-5297-5278
Brent J. Shapiro-Albert https://orcid.org/0000-0002-7283-1124
Xavier Siemens https://orcid.org/0000-0002-7778-2990
Joseph Simon https://orcid.org/0000-0003-1407-6607
Renée Spiewak https://orcid.org/0000-0002-6730-3298
Ingrid H. Stairs https://orcid.org/0000-0001-9784-8670
Daniel R. Stinebring https://orcid.org/0000-0002-1797-3277
Kevin Stovall https://orcid.org/0000-0002-7261-594X
Joseph K. Swiggum https://orcid.org/0000-0002-1075-3837
Stephen R. Taylor https://orcid.org/0000-0003-0264-1453
Jacob E. Turner https://orcid.org/0000-0002-2451-7288
Michele Vallisneri https://orcid.org/0000-0002-4162-0033
Sarah J. Vigeland https://orcid.org/0000-0003-4700-9072
Caitlin A. Witt https://orcid.org/0000-0002-6020-9274

References

Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2019, ApJ, 880, 116
Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2020, ApJ, 889, 38
Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021a, ApJS, 252, 4
Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021b, ApJS, 252, 5
Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, PhRvD, 79, 084030
Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018a, ApJ, 859, 47
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2015, ApJ, 813, 65
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2016, ApJ, 821, 13
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2018b, ApJS, 235, 37
Bailes, M., Barr, E., Bhat, N. D. R., et al. 2016, PoS, 277, 011
Blanco-Pillado, J. J., Olum, K. D., & Siemens, X. 2018, PhLB, 778, 392
Burke-Spolaor, S., Taylor, S. R., Charisi, M., et al. 2019, A&ARv, 27, 5
Burt, B. J., Lommen, A. N., & Finn, L. S. 2011, ApJ, 730, 17
Caprini, C., Durrer, R., & Siemens, X. 2010, PhRvD, 82, 063511
Carlin, B. P., & Chib, S. 1995, Journal of the Royal Statistical Society. Series B (Methodological), 57, 473
Chamberlin, S. J., Creighton, J. D. E., Siemens, X., et al. 2015, PhRvD, 91, 044048
Chen, S., Middleton, H., Sesana, A., Del Pozzo, W., & Vecchio, A. 2017, MNRAS, 468, 404
Chen, S., Sesana, A., & Conselice, C. J. 2019, MNRAS, 488, 401
Christy, B., Anella, R., Lommen, A., et al. 2014, ApJ, 794, 163
Cordes, J. M. 2013, CQGra, 30, 224002
Cornish, N. J., & Sampson, L. 2016, PhRvD, 93, 104047
Demorest, P. B. 2007, PhD thesis, Univ. California
Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94
Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341
Detweiler, S. 1979, ApJ, 234, 1100
Dickey, J. M. 1971, The Annals of Mathematical Statistics, 42, 204
DuPlain, R., Ransom, S., Demorest, P., et al. 2008, Proc. SPIE, 7019, 70191D
Ellis, J., & van Haasteren, R. 2017, jellis18/PTMCMCSampler: Official Release, v1.0.0, zenodo, doi:10.5281/zenodo.1037579
Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T. 2019, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, v2.0, Astrophysics Source Code Library, ascl:1912.015
Fienga, A., Deram, P., Viswanathan, V., et al. 2019, Notes Scientifiques et Techniques de l'Institut de Mecanique Celeste (Paris: Observatoire de Paris)
Folkner, W. M., & Park, R. S. 2016, JPL Planetary and Lunar Ephemeris DE436, Tech. Rep., Jet Propulsion Laboratory, Pasadena, CA

Folkner, W. M., & Park, R. S. 2018, Planetary Ephemeris DE438 for Juno, Tech. Rep. IOM 392R-18-004, Jet Propulsion Laboratory, Pasadena, CA
Ford, J. M., Demorest, P., & Ransom, S. 2010, Proc. SPIE, 7740, 77400A
Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300
Godsill, S. J. 2001, Journal of Computational and Graphical Statistics, 10, 230
Grishchuk, L. P. 1975, JETP, 40, 409
Grishchuk, L. P. 2005, PhyU, 48, 1235
Hazboun, J., Romano, J., & Smith, T. 2019, J. Open Source Softw, 4, 1775
Hazboun, J. S., Simon, J., Siemens, X., & Romano, J. D. 2020a, ApJL, 905, L6
Hazboun, J. S., Simon, J., Taylor, S. R., et al. 2020b, ApJ, 890, 108
Hee, S., Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015, MNRAS, 455, 2461
Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39
Hobbs, G., & Edwards, R. 2012, Tempo2: Pulsar Timing Package, Astrophysics Source Code Library, ascl:1210.015
Hunter, J. D. 2007, CSE, 9, 90
Jaffe, A. H., & Backer, D. C. 2003, ApJ, 583, 616
Jeffreys, H. 1998, The Theory of Probability (Oxford: Oxford Univ. Press)
Jones, M. L., McLaughlin, M. A., Lam, M. T., et al. 2017, ApJ, 841, 125
Joshi, B. C., Arumugasamy, P., Bagchi, M., et al. 2018, JApA, 39, 51
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773
Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2017, MNRAS, 471, 4508
Kerr, M., Reardon, D. J., Hobbs, G., et al. 2020, PASA, 37, e020
Kobakhidze, A., Lagger, C., Manning, A., & Yue, J. 2017, EPJC, 77
Lam, M. T., Cordes, J. M., Chatterjee, S., et al. 2017, ApJ, 834, 35
Lam, M. T., Ellis, J. A., Grillo, G., et al. 2018, ApJ, 861, 132
Lasky, P. D., Mingarelli, C. M. F., Smith, T. L., et al. 2016, PhRvX, 6, 011035
Lee, K. J. 2016, in ASP Conf. Ser. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain & D. Li (San Francisco, CA: ASP), 19
Lentati, L., Taylor, S. R., Mingarelli, C. M. F., et al. 2015, MNRAS, 453, 2576
Luo, J., Ransom, S., Demorest, P., et al. 2019, Astrophysics Source Code Library, record, ascl:1902.007
McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2014, ApJ, 789, 156
Middleton, H., Del Pozzo, W., Farr, W. M., Sesana, A., & Vecchio, A. 2016, MNRAS, 455, L72
Mingarelli, C. M. F. 2019, NatAs, 3, 8
Nice, D., Demorest, P., Stairs, I., et al. 2015, Astrophysics Source Code Library, record, ascl:1509.002
Ng, C. 2018, in IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years, ed. P. Weltevrede et al. (Cambridge: Cambridge Univ. Press), 179
Ölmez, S., Mandic, V., & Siemens, X. 2010, PhRvD, 81, 104028
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. 2020, arXiv:2010.11950
Rajagopal, M., & Romani, R. W. 1995, ApJ, 446, 543
Ransom, S., Brazier, A., Chatterjee, S., et al. 2019, BAAS, 51, 195
Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. 2014, MNRAS, 442, 56
Roebber, E. 2019, ApJ, 876, 55
Sampson, L., Cornish, N. J., & McWilliams, S. T. 2015, PhRvD, 91, 084055
Sazhin, M. V. 1978, SvA, 22, 36
Sesana, A. 2013, CQGra, 30, 224014
Sesana, A., Haardt, F., Madau, P., & Volonteri, M. 2004, ApJ, 611, 623
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
Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, CQGra, 30, 224015
Siemens, X., Mandic, V., & Creighton, J. 2007, PhRvL, 98, 111101
Simon, J., & Burke-Spolaor, S. 2016, ApJ, 826, 11
Taylor, S. R., Baker, P. T., Hazboun, J. S., Simon, J. J., & Vigeland, S. J. 2018, Enterprise Extensions, v1.0, https://github.com/nanograv/enterprise_extensions
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. 2020, libstempo: Python wrapper for Tempo2, v2.3.4, Astrophysics Source Code Library, ascl:2002.017
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
van Haasteren, R., & Vallisneri, M. 2014, PhRvD, 90, 104012
Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016, MNRAS, 458, 1267
Vigeland, S. J., Islo, K., Taylor, S. R., & Ellis, J. A. 2018, PhRvD, 98, 044003
Wang, H., Taylor, S. R., & Vallisneri, M. 2019, MNRAS, 487, 3644
Wilcox, R. 2012, Introduction to Robust Estimation and Hypothesis Testing, Statistical Modeling and Decision Science (Amsterdam: Elsevier)
Wyithe, J. S. B., & Loeb, A. 2003, ApJ, 590, 691
Zhu, X.-J., Cui, W., & Thrane, E. 2019, MNRAS, 482, 2588