

## Astrophysics Milestones For Pulsar Timing Array Gravitational Wave Detection

NIHAN S. POL,<sup>1, 2, 3</sup> STEPHEN R. TAYLOR,<sup>1</sup> LUKE ZOLTAN KELLEY,<sup>4</sup> SARAH J. VIGELAND,<sup>5</sup> JOSEPH SIMON,<sup>6, 7</sup>  
SIYUAN CHEN,<sup>8, 9, 10</sup> ZAVEN ARZOUMANIAN,<sup>11</sup> PAUL T. BAKER,<sup>12</sup> BENECÉ BÉCSY,<sup>13</sup> ADAM BRAZIER,<sup>14</sup>  
PAUL R. BROOK,<sup>2, 3</sup> SARAH BURKE-SPOLAOR,<sup>2, 3, 15</sup> SHAMI CHATTERJEE,<sup>14</sup> JAMES M. CORDS,<sup>14</sup> NEIL J. CORNISH,<sup>13</sup>  
FRONEFIELD CRAWFORD,<sup>16</sup> H. THANKFUL CROMARTIE,<sup>17</sup> MEGAN E. DECESAR,<sup>18, 19, \*</sup> PAUL B. DEMOREST,<sup>20</sup>  
TIMOTHY DOLCH,<sup>21, 22</sup> ELIZABETH C. FERRARA,<sup>23</sup> WILLIAM FIORE,<sup>2, 3</sup> EMMANUEL FONSECA,<sup>24</sup>  
NATHAN GARVER-DANIELS,<sup>2, 3</sup> DEBORAH C. GOOD,<sup>25</sup> JEFFREY S. HAZBOUN,<sup>26, \*</sup> ROSS J. JENNINGS,<sup>14</sup>  
MEGAN L. JONES,<sup>5</sup> ANDREW R. KAISER,<sup>2, 3</sup> DAVID L. KAPLAN,<sup>5</sup> JOEY SHAPIRO KEY,<sup>26</sup> MICHAEL T. LAM,<sup>27, 28</sup>  
T. JOSEPH W. LAZIO,<sup>6</sup> JING LUO,<sup>29</sup> RYAN S. LYNCH,<sup>30</sup> DUSTIN R. MADISON,<sup>2, 3, \*</sup> ALEXANDER MCEWEN,<sup>5</sup>  
MAURA A. MCLAUGHLIN,<sup>2, 3</sup> CHIARA M. F. MINGARELLI,<sup>31, 32</sup> CHERRY NG,<sup>33</sup> DAVID J. NICE,<sup>18</sup>  
TIMOTHY T. PENNUCCI,<sup>34, 35, \*</sup> SCOTT M. RANSOM,<sup>34</sup> PAUL S. RAY,<sup>36</sup> BRENT J. SHAPIRO-ALBERT,<sup>2, 3</sup>  
XAVIER SIEMENS,<sup>37, 5</sup> INGRID H. STAIRS,<sup>25</sup> DANIEL R. STINEBRING,<sup>38</sup> JOSEPH K. SWIGGUM,<sup>18, \*</sup> MICHELE VALLISNERI,<sup>6</sup>  
HALEY WAHL,<sup>2, 3</sup> CAITLIN A. WITT,<sup>2, 3</sup>

THE NANOGrav COLLABORATION

<sup>1</sup>Department of Physics and Astronomy, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235, USA

<sup>2</sup>Department of Physics and Astronomy, West Virginia University, P.O. Box 6315, Morgantown, WV 26506, USA

<sup>3</sup>Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA

<sup>4</sup>Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, Evanston, IL 60208

<sup>5</sup>Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

<sup>6</sup>Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

<sup>7</sup>Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA

<sup>8</sup>Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, LPC2E UMR7328, Université d'Orléans, CNRS, 45071 Orléans, France

<sup>9</sup>Station de Radioastronomie de Nancay, Observatoire de Paris, Université PSL, CNRS, Université d'Orléans, 18330 Nancay, France

<sup>10</sup>FEMTO-ST Institut de recherche, Department of Time and Frequency, UBFC and CNRS, ENSMM, 25030 Besancon, France

<sup>11</sup>X-Ray Astrophysics Laboratory, NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA

<sup>12</sup>Department of Physics and Astronomy, Widener University, One University Place, Chester, PA 19013, USA

<sup>13</sup>Department of Physics, Montana State University, Bozeman, MT 59717, USA

<sup>14</sup>Cornell Center for Astrophysics and Planetary Science and Department of Astronomy, Cornell University, Ithaca, NY 14853, USA

<sup>15</sup>CIFAR Azrieli Global Scholars program, CIFAR, Toronto, Canada

<sup>16</sup>Department of Physics and Astronomy, Franklin & Marshall College, P.O. Box 3003, Lancaster, PA 17604, USA

<sup>17</sup>University of Virginia, Department of Astronomy, P.O. Box 400325, Charlottesville, VA 22904, USA

<sup>18</sup>Department of Physics, Lafayette College, Easton, PA 18042, USA

<sup>19</sup>AAAS, STPF, ORISE Fellow hosted by the U.S. Department of Energy

<sup>20</sup>National Radio Astronomy Observatory, 1003 Lopezville Rd., Socorro, NM 87801, USA

<sup>21</sup>Department of Physics, Hillsdale College, 33 E. College Street, Hillsdale, MI 49242, USA

<sup>22</sup>Eureka Scientific, Inc. 2452 Delmer Street, Suite 100, Oakland, CA 94602-3017

<sup>23</sup>NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

<sup>24</sup>Department of Physics, McGill University, 3600 University St., Montreal, QC H3A 2T8, Canada

<sup>25</sup>Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada

<sup>26</sup>University of Washington Bothell, 18115 Campus Way NE, Bothell, WA 98011, USA

<sup>27</sup>School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA

<sup>28</sup>Laboratory for Multiwavelength Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA

<sup>29</sup>Department of Astronomy & Astrophysics, University of Toronto, 50 Saint George Street, Toronto, ON M5S 3H4, Canada

<sup>30</sup>Green Bank Observatory, P.O. Box 2, Green Bank, WV 24944, USA

<sup>31</sup>*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, New York, 10010, USA*

<sup>32</sup>*Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA*

<sup>33</sup>*Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George St., Toronto, ON M5S 3H4, Canada*

<sup>34</sup>*National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA*

<sup>35</sup>*Institute of Physics, Eötvös Loránd University, Pázmány P. s. 1/A, 1117 Budapest, Hungary*

<sup>36</sup>*Space Science Division, Naval Research Laboratory, Washington, DC 20375-5352, USA*

<sup>37</sup>*Department of Physics, Oregon State University, Corvallis, OR 97331, USA*

<sup>38</sup>*Department of Physics and Astronomy, Oberlin College, Oberlin, OH 44074, USA*

## ABSTRACT

The NANOGrav Collaboration reported strong Bayesian evidence for a common-spectrum stochastic process in its 12.5-yr pulsar timing array dataset, with median characteristic strain amplitude at periods of a year of  $A_{\text{yr}} = 1.92^{+0.75}_{-0.55} \times 10^{-15}$ . However, evidence for the quadrupolar Hellings & Downs interpulsar correlations, which are characteristic of gravitational wave signals, was not yet significant. We emulate and extend the NANOGrav dataset, injecting a wide range of stochastic gravitational wave background (GWB) signals that encompass a variety of amplitudes and spectral shapes, and quantify three key milestones: (I) Given the amplitude measured in the 12.5 yr analysis and assuming this signal is a GWB, we expect to accumulate robust evidence of an interpulsar-correlated GWB signal with 15–17 yrs of data, i.e., an additional 2–5 yrs from the 12.5 yr dataset; (II) At the initial detection, we expect a fractional uncertainty of 40% on the power-law strain spectrum slope, which is sufficient to distinguish a GWB of supermassive black-hole binary origin from some models predicting more exotic origins; (III) Similarly, the measured GWB amplitude will have an uncertainty of 44% upon initial detection, allowing us to arbitrate between some population models of supermassive black-hole binaries. In addition, power-law models are distinguishable from those having low-frequency spectral turnovers once 20 yrs of data are reached. Even though our study is based on the NANOGrav data, we also derive relations that allow for a generalization to other pulsar-timing array datasets. Most notably, by combining the data of individual arrays into the International Pulsar Timing Array, all of these milestones can be reached significantly earlier.

*Keywords:* Gravitational waves – Methods: data analysis – Pulsars: general

## 1. INTRODUCTION

Supermassive black hole binaries (SMBHBs) are expected to form following the mergers of massive galaxies during the buildup of hierarchical structure in  $\Lambda$ CDM (Burke-Spolaor et al. 2019, and references therein). No observationally-confirmed SMBHBs at sub-parsec separations are known, but a growing number of strong candidates have been identified (Graham et al. 2015; Charisi et al. 2016; Liu et al. 2019; Chen et al. 2020). If SMBHBs are indeed able to form, and also to reach  $\sim$ milli-parsec separations, then they will become the sources of the strongest gravitational waves (GWs) in the Universe. The stochastic superposition of GWs produced by SMBHBs across cosmic time is expected to produce an aggregate background signal with an approximate characteristic-strain spectrum of the form  $h_c(f) = A_{\text{GWB}}(f/f_{\text{yr}})^\alpha$  with  $\alpha = -2/3$  for a continuous population of circular and purely GW-driven

binary systems (Rajagopal & Romani 1995; Jaffe & Backer 2003; Sesana et al. 2008). At low frequencies ( $f \lesssim 10^{-8}$  Hz), interactions between SMBHBs and their ambient galactic environments can turn over or flatten the spectrum ( $\alpha > -2/3$ ) (Sampson et al. 2015; Taylor et al. 2017; Kelley et al. 2017b). At high frequencies ( $f \gtrsim 10^{-8}$  Hz) rapid binary hardening can lead to a low occupation of spectral bins that steepens the spectrum ( $\alpha < -2/3$ ) and may introduce spikes from individual sources (Sesana et al. 2008). Despite these effects, the stochastic gravitational wave background (GWB) from SMBHBs is expected to be the dominant source of GWs in the nanohertz frequency regime.

Pulsar Timing Arrays (Sazhin 1978; Detweiler 1979; Foster & Backer 1990) (PTAs) measure the times-of-arrival (TOAs) of radio pulses from millisecond pulsars as a means of measuring the local space-time curvature, and thus signs of passing GWs, analogous to how the LIGO-Virgo Collaboration (Harry & LIGO Scientific Collaboration 2010; Accadia et al. 2012) uses pairs of perpendicular laser arms. As PTAs accrue larger

\* NANOGrav Physics Frontiers Center Postdoctoral Fellow

datasets over time, the increasing number of recorded pulses leads to improved sensitivity, especially as new pulsars are added to the arrays (Siemens et al. 2013; Taylor et al. 2016). The NANOGrav 12.5 yr dataset (Alam et al. 2021) shows conclusive evidence (with Bayesian odds of  $\sim 10^4 : 1$ ) that a common-spectrum low-frequency stochastic process is present across a large fraction of the 45 millisecond pulsars included in the analysis (Arzoumanian et al. 2020). Its properties are consistent with expectations of the strain-spectrum from a stochastic GWB, but it does not yet show statistically-significant quadrupolar interpulsar correlations, which are widely considered the definitive evidence for the GWB (Hellings & Downs 1983; Tiburzi et al. 2016). The latter are the so-called ‘Hellings & Downs’ (HD) correlations of GWB-induced timing delays between pairs of pulsars, which are a function only of the angular separation between the pulsars on the sky. Future datasets, such as the in-preparation NANOGrav 15 yr dataset, in addition to data from the European PTA (EPTA; Desvignes et al. 2016), Australian Parkes PTA (PPTA; Kerr et al. 2020) and the International Pulsar Timing Array (IPTA; Verbiest et al. 2016; Perera et al. 2019), will reveal the nature of the signal. These data will either show spatial correlations consistent with a GWB, or other correlation signatures that suggest unmodeled noise processes affecting the PTA data.

In this paper, we forecast the evolution of GWB inference into the future, under the assumption that the signal is produced by cosmological SMBHBs. There may be alternate sources of the GWB, and there have been many recent suggestions of such interpretations for the NANOGrav 12.5 yr results (see, e.g., Vaskonen & Veermäe 2021; De Luca et al. 2021; Ellis & Lewicki 2021; Blasi et al. 2021). The scaling results that we derive here are generalizable to such other sources of the GWB. Similar analyses performed in the past were focused only on the time to detection of the GWB (e.g., Taylor et al. 2016; Rosado et al. 2015; Vigeland & Siemens 2016; Kelley et al. 2017c), and while this remains important (especially given the results in Arzoumanian et al. (2020)) we consider here for first time the evolution of the GWB parameter measurement uncertainties. The evolution of these properties is of crucial importance in determining the source and underlying physics that produces the GWB. For example, while a GWB from SMBHBs produces a spectral index of  $\alpha = -2/3$ , a GWB produced by primordial GWs can have a spectral index,  $\alpha = -2$  or  $-1$  (Grishchuk 2005; Lasky et al. 2016), while a GWB produced by some models of cosmic strings have  $\alpha = -7/6$  (Ölmez et al. 2010). A distinction between these source models can only be made once the spec-

tral index measurement uncertainty is small enough to exclude one (or more) of the predicted spectral indices. Similarly, different models of a given source can produce different amplitudes of the GWB, and thus knowing when we can distinguish between them is dependent on the evolution of the measurement uncertainties on the GWB amplitude.

However, for a background produced by inspiraling SMBHBs, the canonical  $\alpha = -2/3$  spectral index GW strain power-law emerges when assuming a continuous distribution of circular binary sources evolving purely due to GW emission (Phinney 2001). Following galaxy mergers, two SMBHBs can only become bound and reach the small separations ( $\lesssim 10^{-2}$  pc) required to produce detectable GWs through environmental interactions, such as dynamical friction and stellar “slingshot” scattering (Begelman et al. 1980). These interactions make binaries harden faster than they would due purely to GW emission, and thus the resulting GW strain can be lower than in the  $-2/3$  power-law, particular at low frequencies ( $f \ll 1 \text{ yr}^{-1}$ ). At higher frequencies, finite number effects (Sesana et al. 2008) can also cause significant deviations from a pure power-law. We explore the effects of these variations in the SMBHB GWB spectrum by injecting and recovering “realistic” spectra that have been produced through full mock populations of SMBH binaries.

The paper is structured as follows. In Sec. 2 we describe the methods used to simulate and analyze the PTA datasets. In Sec. 3, we show that the Bayesian model odds ratio can be related to the frequentist signal-to-noise ratio through a simple analytic expression for PTAs, while in Sec. 4 we derive a new statistic, the total signal-to-noise ratio, that encapsulates the information in both the auto- and cross-correlation components of the PTA data. In Sec. 5 we describe the evolution of the GWB detection significance, parameter uncertainties, and spectral characterization of the GWB, and define three key milestones that PTAs should achieve along the way. Finally, in Sec. 6, we show how the timeline to achieving these milestones described can be accelerated through combining the data from individual PTAs into the IPTA.

## 2. METHODS

### 2.1. Simulation framework

We have developed a framework for realistic pulsar-timing data simulation that employs the NANOGrav 12.5 yr dataset (Alam et al. 2021) as its foundation, deriving observational timestamps and TOA uncertainties from all 45 pulsars therein, as well as useful metadata such as observing radio frequency and telescope.

We account for pulse-phase jitter noise, as well as the usual TOA uncertainties from radiometer noise, using the maximum likelihood pulsar noise parameters measured by NANOGrav in the 12.5 yr data preparation and analysis. We epoch-average the TOA uncertainties and timestamps which reduces the effective number of TOAs by almost an order of magnitude. We also inject intrinsic red noise in each pulsar according to its measured values in the NANOGrav 12.5 yr dataset, with the caveat that we filter out a GWB-like  $\alpha = -2/3$  red process from each pulsar by modeling it alongside the intrinsic pulsar noise. This is important in isolating the true intrinsic red noise in each pulsar, which may otherwise be conflated with a common process in single-pulsar noise estimation. These simulations are the first to inject red noise in each pulsar that reflects the measured NANOGrav common red noise being appropriately filtered out.

In order to forecast data accumulation beyond the existing 12.5 yr baseline, we derive distributions of the observational cadences and TOA uncertainties from the final year’s worth of data in each pulsar. We then adopt a status-quo approach of drawing new observation times and uncertainties from these distributions until we reach a 20 year array baseline. We retain the 45 pulsars from the 12.5 yr analysis and do not add any new pulsars as a function of time. This latter choice implies that the temporal growth rate of the detection statistics that we derive in this work will be relatively conservative, since an addition of pulsars to the PTA is known to increase the detection significance of the GWB (Siemens et al. 2013).

Finally, we inject ten different realizations of noise and isotropic GWB signals (with spectral index  $\alpha = -2/3$ ) for each of eight different amplitudes in the range  $10^{-16} \leq A_{\text{GWB}} \leq 5 \times 10^{-15}$  into our synthesized 20-year baseline array. We then study how the signal evolves in our data by analyzing increasingly longer slices of each simulation, from 10 years up to the final 20 years. Put together, our PTA simulation technique is the most sophisticated that has ever been presented in the literature.

## 2.2. Bayesian Data Analysis Methods

The methods used for modeling the injected GWB are the same as those used in the NANOGrav 11 yr (Arzoumanian et al. 2018) and 12.5 yr (Arzoumanian et al. 2020) analyses. Thus, we only briefly summarize the probabilistic framework for estimating the GWB parameters.

In the standard detection pipeline, the GWB is modeled as a power-law across frequency for the character-

istic strain,

$$h_c(f) = A_{\text{GWB}} \left( \frac{f}{1 \text{ yr}^{-1}} \right)^\alpha, \quad (1)$$

where  $A_{\text{GWB}}$  is the amplitude of the GWB,  $f$  is the frequency, and  $\alpha = -2/3$  is the spectral index expected for a population of inspiraling SMBHBs whose inspiral is dominated by GWs. This GWB spectrum can be expressed in terms of the timing-residual cross-spectral density between pulsars  $a$  and  $b$  as,

$$S_{ab}(f) = \Gamma_{ab} \frac{A_{\text{GWB}}^2}{12\pi^2} \left( \frac{f}{1 \text{ yr}^{-1}} \right)^{-\gamma} \text{yr}^3, \quad (2)$$

where  $\gamma \equiv 3 - 2\alpha$ , and  $\Gamma_{ab}$  is the overlap reduction function that encodes the interpulsar spatial correlations, such as the HD spatial correlations (Hellings & Downs 1983) expected for an isotropic GWB.

In addition to the GWB, we also need to model the intrinsic noise for each pulsar in the PTA. For each pulsar, we model uncorrelated instrumental Gaussian noise in the multiplicative EFAC and quadrature-additive EQUAD parameters. Noise that is correlated across observing frequencies but uncorrelated across TOAs is modeled using the ECORR parameter. These parameters are collectively referred to as “white noise”. We include an additional power-law process per pulsar to model any intrinsic low-frequency noise (i.e. “spin red noise”) in the given pulsar. The aforementioned white noise parameters that correct the reported TOA uncertainties were folded into the simulated uncertainties in the preparation of our datasets, so we do not separately model these white noise parameters in our analysis. The red noise for the pulsars is allowed to vary when we search for the GWB.

The GWB parameter constraints are derived by modeling the GWB as a time-correlated process that is statistically common (i.e. with a common spectrum) but uncorrelated among all pulsars in the PTA. This ignores the influence of interpulsar correlations on GWB parameter constraints, which should be minimal as these are dominated by autocorrelation information in the pulsars. Constraints on the GWB spectral index,  $\alpha$ , are reported as marginalized over all parameters, including  $A_{\text{GWB}}$ , while constraints on  $A_{\text{GWB}}$ , are reported as marginalized over all noise parameters but conditioned on the fiducial spectral index  $\alpha = -2/3$ . The inclusion of HD correlations in the search process is computationally expensive, so we only use it to discriminate whether the observed common red process is from a GWB or other spatially correlated signals like terrestrial clock errors (monopolar spatial correlations) or Solar System ephemeris errors (dipolar spatial correlations). These

spatial correlations are encoded in the overlap reduction function in Eq. 2.

To calculate the Bayesian detection significance, we use a product-space sampling method that computes the probabilistic preference for one signal model over another (Arzoumanian et al. 2018; Taylor et al. 2020, and references therein). We construct a super-model consisting of the union of multiple GWB signal models, where an indexing variable determines which model is “active” and used to calculate the corresponding likelihood. The ratio of the posterior probabilities of the model indices can then be used to estimate the Bayesian odds for the preference of one model over the other. Thus, to calculate the Bayesian odds for a HD-correlated process in the dataset,  $\mathcal{O}_{\text{HD}}$ , we create a super-model with the same GWB parameters in both models, while the index parameter toggles the presence of HD correlations.

In addition to power-law strain spectrum models, we also measure the GWB strain spectrum agnostically with an independent amplitude at each frequency. Our priors correspond to independent log-uniform constraints on the characteristic strain amplitude for each Fourier-basis frequency, given by  $k/T$ , where  $k = 1, 2, \dots, 30$  and  $T$  is the time-span between the first and last TOA in the given dataset. Unless explicitly specified, all of the results presented in this paper have been derived using 30 frequencies to model the GWB (and intrinsic red noise) and employing the recent JPL solar system ephemeris, DE438 (Folkner & Park 2018). However, we also explore the effect of using a different number of frequencies and BayesEphem (Vallisneri et al. 2020) in modeling the GWB.

### 2.3. Frequentist Data Analysis Methods

In addition to the Bayesian pipeline, we also use a frequentist approach to model the GWB (Arzoumanian et al. 2016, 2018). This uses an optimal statistic,  $\hat{A}^2$ , to measure the amplitude of the GWB, corresponding to the sum of the correlations between pulsar pairs weighted by the intrinsic pulsar noise, and conditioned on the assumed spectral index of the GWB. The optimal statistic can also be used to calculate the signal-to-noise ratio,  $\rho_{\text{template}}$ , for a given interpulsar correlation template, e.g., monopolar or dipolar for clock and ephemeris errors respectively, and HD (or quadrupolar) for an astrophysical GWB. To calculate the template S/N, the optimal statistic first calculates the amount of cross-correlated power between different pulsar pairs in the PTA. This pairwise cross-correlated power, suitably binned in angular separation space, can be used as a visual test for the presence of quadrupolar correlations in

the data, or as intermediate products for further study of the spatial correlations.

In this work, we use the noise-marginalized optimal statistic (Vigeland et al. 2018), which is more robust when pulsars have intrinsic red noise. We use posterior samples from Bayesian MCMC analyses to marginalize over the pulsar intrinsic red noise and calculate corresponding distributions for  $\hat{A}^2$  and  $\rho_{\text{template}}$ .

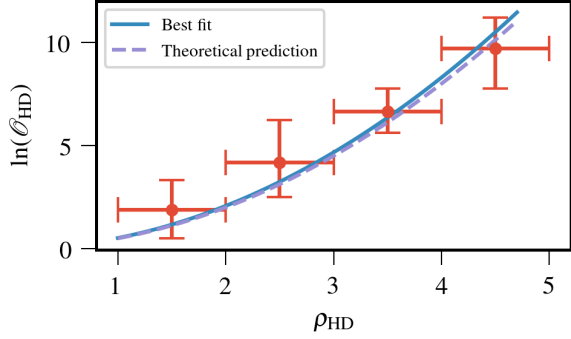
These Bayesian and frequentist methods reflect production-level analyses that are employed in PTA detection and parameter estimation analyses. Thus, our results are a crucial validation of the efficacy of the pipelines that will be performing the important business of SGWB detection in the next several years.

### 3. BRIDGING BAYESIAN ODDS AND FREQUENTIST SIGNAL-TO-NOISE RATIOS

Under certain circumstances it is possible to relate Bayesian model selection to frequentist hypothesis testing. When data is informative such that the likelihood is strongly peaked, the Bayesian evidence can be computed under the *Laplace approximation*, such that  $Z_{\mathcal{H}} \equiv \int d\theta p(d|\theta, \mathcal{H})p(\theta|\mathcal{H}) \approx p(\theta_{\text{ML}}|\mathcal{H})\Delta V_{\mathcal{H}}/V_{\mathcal{H}}$ , where  $p(d|\theta_{\text{ML}}, \mathcal{H})$  maximizes the likelihood with parameters  $\theta$  given data  $d$  under model  $\mathcal{H}$  (Romano & Cornish 2017).  $\Delta V_{\mathcal{H}}/V_{\mathcal{H}}$  measures the compactness of the parameter space volume occupied by the likelihood with respect to the total prior volume, incorporating the Bayesian notion of model parsimony. Taking the ratio of Bayesian evidences between two models, labeled 1 and 2, and assuming equal prior odds, allows the Bayesian odds ratio,  $\mathcal{O}_{12}$ , to be written as  $\ln \mathcal{O}_{12} \approx \ln \Lambda_{\text{ML}}(d) + \ln [(\Delta V_1/V_1)/(\Delta V_2/V_2)]$ , where  $\Lambda_{\text{ML}}(d)$  is the maximum likelihood ratio. The relevant maximum likelihood statistic for PTA GWB detection is the aforementioned *optimal statistic*: a noise-weighted two-point correlation statistic between all unique pulsar pairs, which compares a model with HD interpulsar correlations versus one with no correlations between pulsars (Anholm et al. 2009; Demorest et al. 2013; Chamberlain et al. 2015). The signal-to-noise ratio (S/N,  $\rho$ ) of such GWB-induced correlations can be written as  $\ln \Lambda_{\text{ML}} = \rho^2/2$ . While the likelihood may be marginally more compact under the model with HD correlations, there is no difference in parameter dimensionality; hence we ignore the likelihood compactness terms. The relationship between the Bayesian odds ratio in favor of HD correlations and the frequentist S/N of such correlations can then be written as,

$$\ln \mathcal{O}_{\text{HD}} \approx \rho_{\text{HD}}^2/2. \quad (3)$$

In a bid to assess the validity of this relationship for PTA GWB searches, we used our simulated datasets to



**Figure 1.** The Bayesian odds ratio preferring a GWB over a spatially uncorrelated process ( $\mathcal{O}_{\text{HD}}$ ) is plotted against the cross-correlation signal-to-noise ratio (S/N,  $\rho$ ). These quantities were calculated using the injected simulated datasets across all realizations and injected amplitudes. The theoretical prediction (dashed blue line) for the relation between these two measures of confidence agrees with the empirical best-fit model (solid blue line).

calculate both of these statistics across all realizations, injected amplitudes and time slices. These values, and the theoretical relationship in Eq. 3, are shown in Fig. 1. We also allowed the pre-factor of  $1/2$  to vary in an empirical fit, finding a value of  $0.52 \pm 0.01$ , which agrees well with Eq. 3. This is the first time that this relationship has been validated for PTA GWB detection.

#### 4. A STATISTIC TO QUANTIFY PTA MILESTONES

There are many factors that influence the detectability of a GWB signal in a given PTA configuration. Some of these are related to the signal itself, i.e. the amplitude of the characteristic strain spectrum at  $f = 1/\text{yr}$ ,  $A_{\text{GWB}}$ , as well as the frequency-dependent shape of the spectrum. The quality of our detector is parametrized through the overall observational timing baseline of the array, the number of pulsars, and the timing quality of each pulsar (given by its radio-frequency dependent noise characteristics). We need a statistic that accounts for all of these factors, acting as a fiducial scaling parameter in terms of which we can track the evolution of parameter uncertainties and signal detectability. The optimal statistic is not appropriate here, since it only considers cross-correlations and thus underestimates the information content in the full signal (especially the auto-correlations) that dominate parameter estimation.

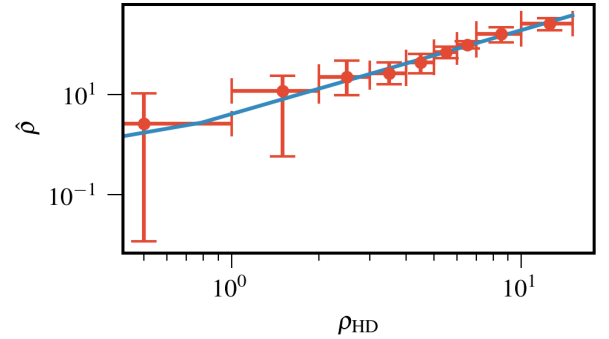
The full signal likelihood (modeling auto- and cross-correlations) is a sufficient statistic for this purpose. Hence we form a total signal S/N,  $\hat{\rho}$ , through the log-likelihood ratio of the GWB+noise versus noise-only models, where noise is treated as uncorrelated between

pulsars. Therefore,

$$\hat{\rho} = \sqrt{2 \ln \Lambda} \equiv \sqrt{2 \left[ \ln p(d|\vec{\theta}_{\text{signal}}, \vec{\theta}_{\text{noise}}) - \ln p(d|\vec{\theta}_{\text{noise}}) \right]}. \quad (4)$$

This differs from HD cross-correlation S/N,  $\rho_{\text{HD}}$ , (or an equivalent cross-correlation statistic) in considering all distinct pairings of the pulsars, including autocorrelations.

We would like to note that this new statistic is not meant to be a replacement for the cross-correlation S/N. Since an astrophysical GWB is characterized by the unique HD cross-correlation signature, the corresponding cross-correlation S/N,  $\rho_{\text{HD}}$ , will be the primary arbiter for confirming a signal as an astrophysical GWB.



**Figure 2.** The evolution of the total S/N,  $\hat{\rho}$ , as a function of the HD cross-correlation S/N,  $\rho_{\text{HD}}$ . As we can see, HD template S/N values of 1, 3, and 7 correspond to  $\hat{\rho}$  values of 5, 25 and 105 respectively.

As PTAs move closer to detection, the power in the GWB at low frequencies is expected to dominate the intrinsic white and red noise in some of the pulsars in the PTA. In this “intermediate signal regime”, the total S/N,  $\hat{\rho}$ , is expected to be much larger than the corresponding HD S/N ( $\rho_{\text{HD}}$ , Romano et al. 2020). We can directly connect them by calculating the two statistics on the same simulated injected datasets. The binned values for the two statistics calculated this way are shown in Fig. 2, and related by the empirical scaling relation,

$$\hat{\rho} = 25 \left( \frac{\rho_{\text{HD}}}{3} \right)^{1.7}. \quad (5)$$

Using this scaling relation,  $\rho_{\text{HD}}$  values of  $[1, 3, 7]$  map to  $\hat{\rho}$  values of approximately  $[5, 25, 105]$  respectively. Since the power in the auto-correlations is significantly larger than in cross-correlations, the corresponding S/N values for the total SN,  $\hat{\rho}$ , are higher than those for HD cross-correlation,  $\rho_{\text{HD}}$  (Romano et al. 2020).

From existing scaling laws for  $\rho_{\text{HD}}$  (Siemens et al. 2013), we can generalize  $\hat{\rho}$  to any PTA configuration. For a PTA in the weak and intermediate signal regime (Siemens et al. 2013), the total S/N can be written as,

$$\hat{\rho} \propto \left[ M c \frac{A_{\text{GWB}}^2}{\sigma^2} T^\gamma \right]^{1.7} \quad (6a)$$

$$\hat{\rho} \propto \left[ M \left( \frac{A_{\text{GWB}}}{\sigma \sqrt{c}} \right)^{1/\gamma} T^{1/2} \right]^{1.7}, \quad (6b)$$

respectively, where  $M$  is the number of pulsars in the array,  $c = 1/\Delta t$  is the inverse of the observational cadence,  $A_{\text{GWB}}$  is the amplitude of the GWB,  $\sigma$  is the white noise timing RMS,  $T$  is the timing baseline of the PTA, and  $\gamma$  is as previously defined ( $\gamma \equiv 3 - 2\alpha = 13/3$  for a population of SMBHBs).

## 5. PTA MILESTONES

### 5.1. Milestone I: Detection

The upper panel of Fig. 3 shows the evolution of the median  $\hat{\rho}$  as a function of data baseline and injected GWB amplitude, the middle panels show the cross-correlated power in different angular separation bins for a GWB with an amplitude of  $A_{\text{GWB}} = 2 \times 10^{-15}$ , for three different data baselines (12 yr, 15 yr, and 20 yr), and the lower panel shows the evolution of the S/N for HD, monopole, and dipole cross-correlations as a function of data baseline for a GWB with amplitude between  $A_{\text{GWB}} = [1 - 2] \times 10^{-15}$ . With an increase in the length of the PTA dataset,  $\hat{\rho}$  for the GWB increases, as well as the S/N for HD correlations,  $\rho_{\text{HD}}$ . The bottom panel of Fig. 3 shows that, starting at the 15 yr slice, the median HD correlations are preferred over other types of spatial correlation for a GWB with strength  $A_{\text{GWB}} = [1 - 2] \times 10^{-15}$ . Only with HD correlations clearly favored over monopolar or dipolar signals (which happens near the 18 yr baseline in our simulations) can a signal be confidently attributed to GWBs instead of uncharacterized noise sources.

To test the influence of the number of frequencies used to model the GWB, we repeat the analysis in the lower panel of Fig. 3 for a GWB modeled with only the lowest 5 frequencies. The choice of these lowest 5 frequencies is designed to avoid GWB significance and parameter estimation being contaminated by poorly-modeled excess noise in the dataset that dominates at higher frequencies (Arzoumanian et al. 2020). We find that the initial detection of a GWB is not inhibited or delayed by using only the lowest 5 frequency bins, which only reduces the S/N for HD correlations by a few percent. We also explored the influence that accounting for systematics in Solar-System ephemeris modeling has on

our GWB detection significance. We do so by including BAYESEPHM (Vallisneri et al. 2020) in our model, which marginalizes over uncertainties in Jupiter orbital elements, gas giant masses, and the celestial reference frame. We find that BAYESEPHM reduces detection significance by only  $\sim 4$  units of total S/N,  $\hat{\rho}$ , consistent with earlier results reported in Vallisneri et al. (2020). Thus, the presence of solar system ephemeris uncertainties is unlikely to inhibit the detection of a GWB by PTAs.

As shown in Fig. 3 as a black data point, detection statistics for the NANOGrav 12.5 yr dataset are consistent with our simulations, and we expect the significance of a GWB to increase in the real dataset at least at the rate shown here, if not faster due to the addition of more pulsars to the array (Siemens et al. 2013) and new advanced noise mitigate schemes that should allow the GWB spectrum to be modeled to higher frequencies.

### 5.2. Milestone II: Source Of The GWB

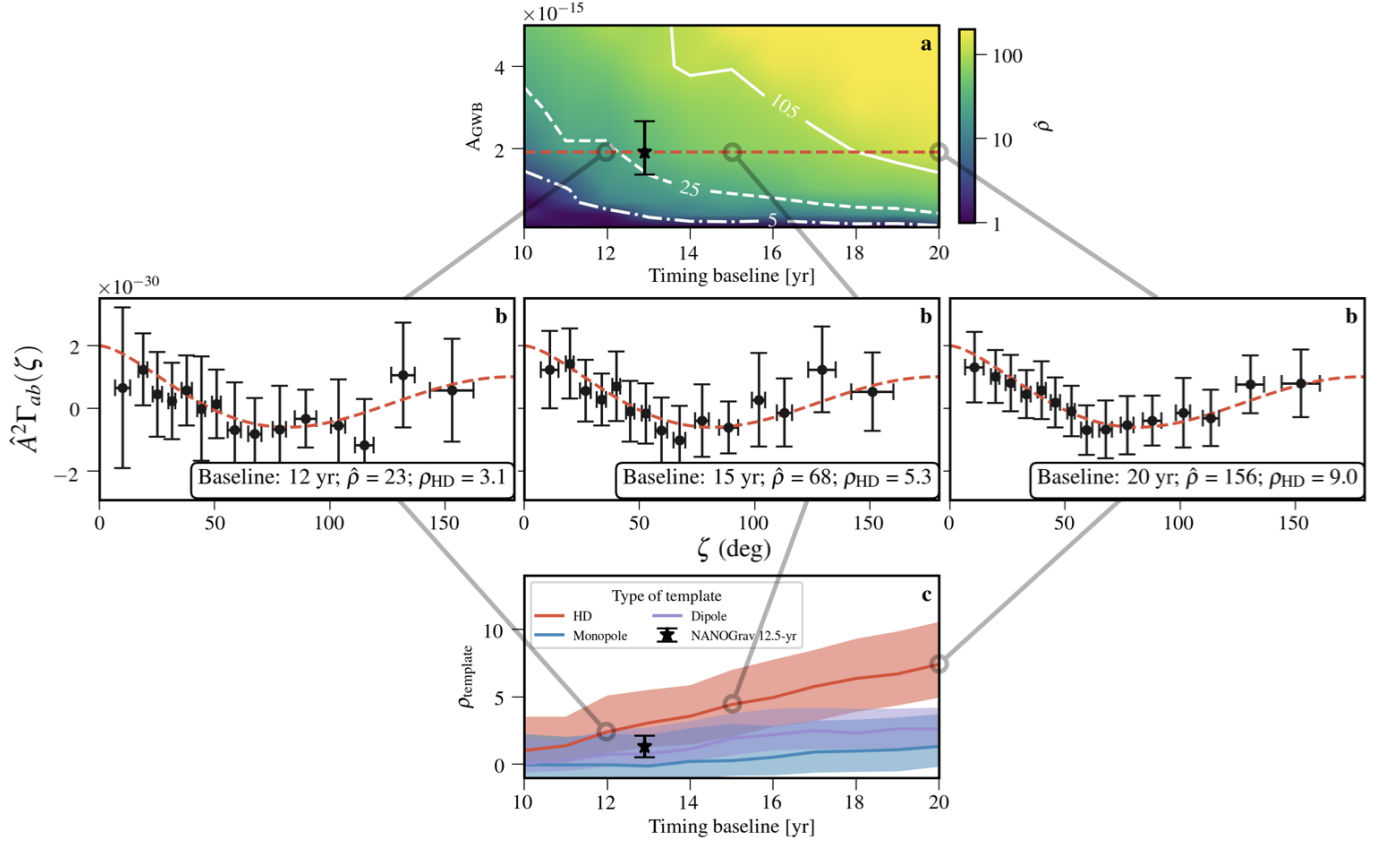
Fig. 4 shows the evolution of the fractional parameter measurement uncertainty,  $\Delta X/X$ , as a function of total S/N, where  $X$  is the median measured value of the GWB amplitude ( $A_{\text{GWB}}$ ) and spectral index ( $\alpha$ ), and  $\Delta X$  is the corresponding 95% credible interval parameter uncertainty. These uncertainties are well fit by the relations,

$$\frac{\Delta A_{\text{GWB}}}{A_{\text{GWB}}} = 44\% \times \left( \frac{\hat{\rho}}{25} \right)^{-2/5}, \quad (7a)$$

$$\frac{\Delta \alpha}{\alpha} = 40\% \times \left( \frac{\hat{\rho}}{25} \right)^{-1/2}. \quad (7b)$$

The choice of using the lowest 5 frequencies to model the GWB does not affect the fractional parameter uncertainties until the signal has total S/N  $\hat{\rho} \gtrsim 60$ , whereupon using 30 frequencies improves their recovery. At such high detection significance, the GWB begins to dominate more than just the lowest 5 frequency bins, accruing greater significance and more informative constraints from higher frequencies. We also find that the inclusion of BAYESEPHM has minimal effect on the fractional uncertainty of the measured GWB parameters.

The measured spectral index of the GWB will be the primary arbiter of the source of the GWB. A spectral index of  $\alpha = -2/3$ , corresponding to our injected signal, is expected for a purely GW-driven population of inspiraling SMBHB systems, while other more exotic sources of the GWB are predicted to have different spectral indices. At initial detection ( $\hat{\rho} = 25$  or  $\rho_{\text{HD}} = 3$ ), the spectral index measurement should have a fractional uncertainty of 40% (Eq. 7b). In our suite of simulations, this precision is already sufficient to disfavor (at 95% credibility)



**Figure 3.** *Top:* The evolution of the median total S/N,  $\hat{\rho}$ , over ten realizations as a function of time. The dash-dot, dashed and solid contours represent total S/N values of 5, 25, and 105 corresponding to HD S/N,  $\rho_{\text{HD}}$ , of 1, 3 and 7 respectively. The black data point represents the amplitude of the GWB from the NANOGrav 12.5 yr analysis. *Middle:* The median binned cross-correlated power over ten realizations for increasing S/N from left to right. The uncertainties in this panel represent the spread of the cross-correlated power over the ten realizations. The theoretical HD curve for a power-law GWB with an amplitude of  $A_{\text{GWB}} = 2 \times 10^{-15}$  is shown by the red dashed line. *Bottom:* The cross-correlation S/N ratio,  $\rho_{\text{template}}$ , for different types of spatial correlations with an injected power-law GWB amplitude of  $A_{\text{GWB}} = [1 - 2] \times 10^{-15}$ . The shaded regions represent the median 95% credible intervals across ten realizations of the PTA dataset. The black data point represents the significance of the GWB from the NANOGrav 12.5 yr analysis.

models of primordial GWs with matter ( $\alpha = -2$ , Grishchuk 2005) and radiation-dominated ( $\alpha = -1$ , Lasky et al. 2016) equations of state.<sup>1</sup> This precision is also sufficient to disfavor some models of the GWB produced by cosmic strings, such as those from kinks and cusps in the string loops which are predicted to have a spectral index,  $\alpha = -7/6$  (Ölmez et al. 2010). As the significance of the GWB signal grows, the fractional uncertainty on the measured spectral index will continue to decrease and allow us to test other sources of the GWB. As we show later, an increase in the timing baseline of the dataset allows more comprehensive spectral model-

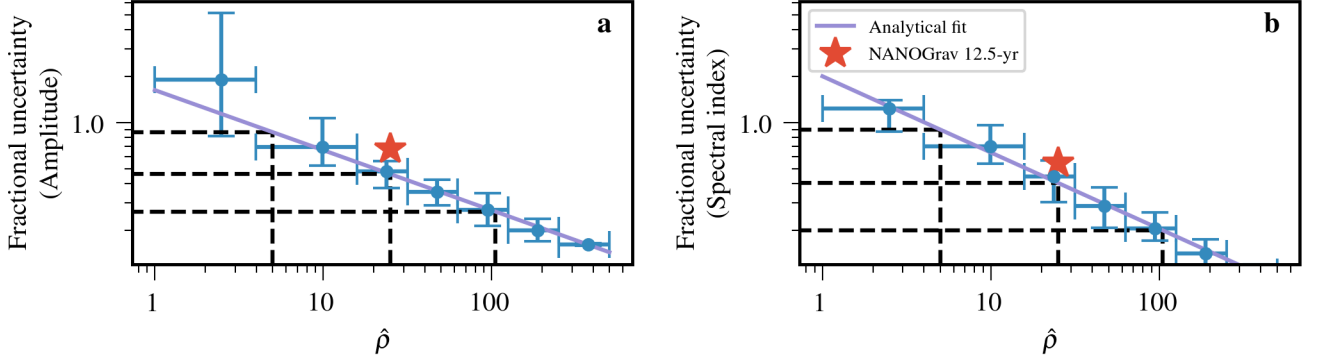
ing of the GWB, thereby reducing the need to rely on power-law fits.

### 5.3. Milestone III: Properties Of The GWB

Once there is sufficient evidence that the GWB is most likely due to a population of merging SMBHBs, we can use our spectral characterization to probe the astrophysics of the underlying SMBHB population. For example, the amplitude of a GWB produced by SMBHBs is set primarily by the mass distribution of SMBHBs, often parametrized relative to their host galaxies (e.g., in the M- $\sigma$  relation; Gebhardt et al. 2000; Ferrarese & Merritt 2000), and the efficiency by which they reach sub-parsec separations. The recovered amplitude can thus be used to distinguish between different population models of SMBHBs.

Here we consider three such models (McWilliams et al. 2014; Simon & Burke-Spolaor 2016; Sesana

<sup>1</sup> But note that such a broadband primordial GWB signal with similar amplitude to the common-spectrum process found in the NANOGrav 12.5 yr analysis could be constrained at higher frequencies by ground-based GW interferometers and Big Bang nucleosynthesis (e.g., Kuroyanagi et al. 2021).



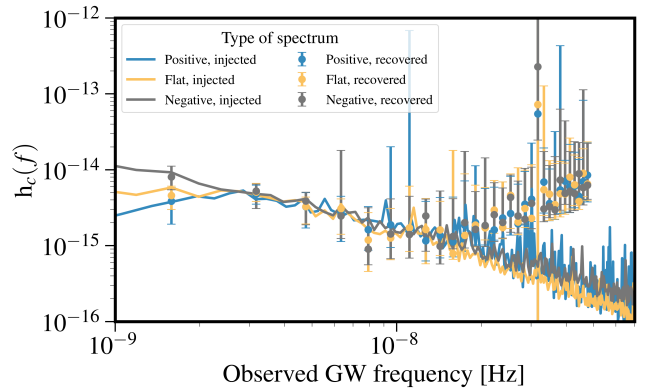
**Figure 4.** The evolution of the median fractional uncertainty on the measured amplitude,  $A$ , and spectral index,  $\alpha$ , as a function of total S/N,  $\hat{\rho}$ , is shown in panels (a) and (b) respectively. The evolution of the fractional uncertainty for the amplitude and spectral index with the S/N is best fit by a power-law as shown (parametrized in Eq. 7). For the amplitude and spectral index, the fractional uncertainties corresponding to  $\hat{\rho}$  values of 5, 25 and 105 are 84%, 44% and 25%, and 90%, 40% and 20% respectively, and are shown by the dashed black lines. The star shows the fractional uncertainty from the NANOGrav 12.5 yr analysis.

et al. 2016) with GWB amplitudes representative of typical values predicted in the literature:  $A_{\text{GWB}} = [0.4^{+0.26}_{-0.16}, 1.5^{+0.3}_{-0.3}, 4.0^{+3.3}_{-1.8}] \times 10^{-15}$ , which were also examined in the NANOGrav 11 yr GWB analysis (Arzoumanian et al. 2018). At initial detection, the amplitude measurement will have a fractional uncertainty of 44%, which will be sufficient to distinguish the first model from the other two. Models two and three are distinguishable with a fractional uncertainty of  $\sim 37\%$  on the amplitude, occurring near the 17 yr slice of our simulated data. Thus, almost immediately after the initial detection of the GWB, we expect to be able to clearly distinguish between typical models in the literature.

### 5.3.1. Constraining dynamical influences on the GWB spectral shape

A realistic, astrophysical GWB from SMBHBs is, however, expected to deviate noticeably from the pure power-law spectrum we have assumed so far. To quantify the influence of these deviations on our detection prospects, we use astrophysically-motivated simulations of SMBHB populations to inject three non-power-law GWB spectra into the synthesized data.

To calculate plausible and self-consistent GWB spectra, we generate full mock populations of SMBH binaries. We employ two entirely independent approaches to simulate SMBHB populations. In the first, we anchor the population on the galaxy-galaxy merger rate derived from the Illustris cosmological simulations (Vogelsberger et al. 2014a; Genel et al. 2014; Vogelsberger et al. 2014b). While the Illustris population of SMBH mergers has been used extensively (Kelley et al. 2017a,c, 2018), we have generalized our approach to allow for a more flexible range of binary properties, and thus result-



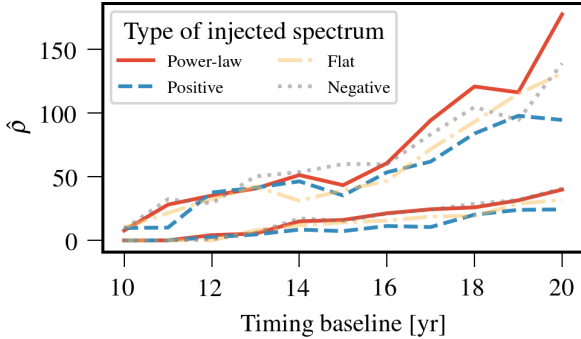
**Figure 5.** Per-frequency spectral modeling of the GWB spectrum for the 20 yr slice. Each frequency bin shows the median strain and uncertainty across ten realizations, while the injected GWB spectrum is shown by the solid lines. The lowest few frequency bins closely track the shape of the injected GWB, with the lowest frequency bin being the strongest discriminator between the three types of injected spectra. The high frequency bins are dominated by the white noise in the PTA and thus are insensitive to the GWB.

ing GWB spectra. In particular, we have chosen binary hardening rates and masses so as to produce i) a GWB spectrum that is very nearly a power-law with a relatively steep “negative” spectral index at low frequencies (Fig. 5: grey), and ii) a strain spectrum that is nearly “flat” at low frequencies (Fig. 5: yellow).

The second method uses spectra originally generated for the astrophysical inference performed on the NANOGrav 11yr dataset (Arzoumanian et al. 2018). This method uses a semi-analytic model (Simon & Burke-Spolaor 2016) to simulate a binary population using observational-based measurements of the galaxy

stellar mass function, galaxy merger rate, and SMBH mass–host galaxy relation. The eccentricity distribution and binary hardening rate is incorporated over a wide range of parameters (Taylor et al. 2017). Model parameters are chosen to produce iii) a GW strain spectrum that is attenuated strongly enough to turn over, producing a “positive” spectral index at low frequencies (Fig. 5: blue).

Overall, the amplitude of the simulated spectra is calibrated primarily by the distribution of SMBH masses, and the low-frequency spectral index is increased from the fiducial  $-2/3$  by increasing the environmental hardening rate, mostly due to coupling with the nuclear stellar environment (Simon & Burke-Spolaor 2016; Kelley et al. 2017c). For both methodologies, we have calibrated all three realistic spectra to have an amplitude of  $A_{\text{GWB}} = 1 \times 10^{-15}$  at a frequency of  $1 \text{ yr}^{-1}$ , and are shown by the solid lines in Fig. 5.



**Figure 6.** Detection significance for a variety of injected GWB spectra that are all modeled using a pure power-law spectrum in the detection pipeline. The two sets of curves represent the median 95% credible interval upper and lower bounds on the S/N. We can see that power-law modeling of the GWB is robust for making a detection of the GWB.

The total S/N recovered using our standard detection pipeline (assuming a pure power-law strain spectrum) is shown for these realistically-modeled GWB spectra in Fig. 6. Even in the “positive” model, where a significant amount of GW strain is attenuated at low frequencies, the measured S/N is unaffected at the 15–17 yr slices (where initial detection is predicted to happen) and is only decreased by  $\sim 50\%$  at the 20 yr slice. This point is crucial, since it demonstrates that the assumption of a pure power-law strain spectrum allows for a high S/N detection, despite omitting the detailed spectral structure. Additionally, even significant attenuation of the GWB signal at low frequencies does not represent a significant obstacle to detection. Any eventual reduction in

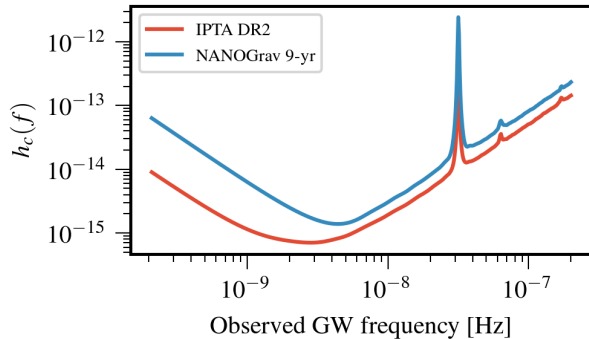
the recovered S/N compared to the predicted S/N can then be used to interpret the presence of a turnover in the GWB spectrum.

However, many of the astrophysical details about SMBHB environments and demographics are missed by power-law analyses. The additional encoded information can be extracted using per-frequency modeling of the GWB (Lentati et al. 2013; Taylor et al. 2013). In this approach, the strain in each frequency bin is modeled independently, allowing us to measure the spectral shape of the underlying GWB. A per-frequency spectral analysis of 20 yrs of data is shown in Fig. 5 for our three astrophysically-motivated injected GWB spectra, with points and uncertainties representing strain measured independently in each frequency bin. We see that the lowest few bins are dominated by the injected GWB and closely track their spectral shapes. In a followup analysis we will explore full astrophysical model inference to determine how accurately astrophysical parameters can be extracted from realistic GWBs. However, the results shown here are very encouraging in that even models with some degeneracy (i.e., in the overall signal amplitude,  $A_{\text{GWB}}$ ) can be disentangled within the next several years.

## 6. ACCELERATING PTA MILESTONES

The S/N for a GWB signal in PTA data can be raised by increasing the time-span of the dataset, or by increasing the number of monitored pulsars (Siemens et al. 2013). However, new pulsars usually cannot be added into a PTA immediately upon discovery, and generally require multiple years of timing data to assess their appropriateness for PTA analysis. A solution is to combine the data already collected by PTAs around the world into a single dataset, and analyze this joint dataset for a GWB signal. Beyond NANOGrav, there are currently two other PTAs apart that have decades-worth of timing data on millisecond pulsars: the Parkes Pulsar Timing Array (PPTA, Hobbs 2013) and the European Pulsar Timing Array (EPTA, Kramer & Champion 2013). Together, these PTAs, along with emerging PTAs in India (InPTA, Joshi et al. 2018), China (CPTA, Lee 2016) and South Africa (SAPTA, Bailes et al. 2016), form the International Pulsar Timing Array (IPTA, Hobbs et al. 2010). To date, the IPTA has published two data releases (Verbiest et al. 2016; Perera et al. 2019), which result in an increase in the number of pulsars relative to individual regional PTAs at that time, while also increasing the time-span of data for pulsars that are common across the three PTAs. The IPTA has analyzed one of these datasets for GW signals (Verbiest et al. 2016;

Lentati et al. 2016), while the analysis of the latest data release (Perera et al. 2019) is currently in progress.



**Figure 7.** The sensitivity curve for IPTA dr2 and the NANOGrav 9 yr dataset, which was one of the components going into the construction of dr2. These sensitivity curves were constructed using HASASIA, and only used the white noise parameters released with IPTA dr2. The IPTA dr2 is more sensitive at all frequencies than the NANOGrav 9 yr dataset.

To quantify the improvement offered by the IPTA over NANOGrav-only data, we use the IPTA’s second data release (Perera et al. 2019), DR2, and compare its sensitivity to that of the NANOGrav subset of the DR2 dataset, i.e., the NANOGrav 9 yr dataset (NANOGrav Collaboration et al. 2015) (NG9). We estimate the sensitivity using HASASIA (Hazboun et al. 2019c,a), a software package that can compute the per-frequency sensitivity of any PTA given the white and red noise properties of the pulsars. We use the noise properties of the pulsars provided with the public DR2 data, but do not include intrinsic red noise since these parameters are highly covariant with a common process and would require a full GWB detection analysis for accurate calculation.

The per-frequency sensitivities of IPTA DR2 and NG9 are shown in Fig. 7. IPTA DR2 is more sensitive across all frequencies than NG9. Consequently, IPTA DR2 achieves a total-signal significance that is a factor of 30 higher than NG9. This corresponds to a decrease in the fractional uncertainty on the the GWB amplitude and spectral index by factors of 4 and 5.5 respectively. Note that these are only estimates since (as previously mentioned) no intrinsic red noise was accounted for in the sensitivity curves, which would have the effect of reducing each of the S/N values. However, given these estimates, an analysis of a combined IPTA dataset offers the opportunity to detect a GWB and extract the underlying astrophysics of the source population at a faster

pace and with higher significance than can be achieved with any individual PTA.

## 7. CONCLUSION

The recent NANOGrav evidence of a common-spectrum process is a tantalizing hint of the first signs of a GWB with growing significance in PTA data. If so, we have developed a road-map for the next several years that includes three key PTA scientific milestones: (I) robust detection of the GWB through measuring Hellings-&-Downs-correlated timing delays across the PTA; (II) discriminating the origin of the GWB as being from either a population of supermassive black-hole binaries or exotica; and (III) unveiling the demographic and dynamical properties of supermassive black-hole binaries as encoded in the shape of the recovered GWB strain spectrum. All of these milestones could be achievable by the middle of this decade under conservative assumptions, and even sooner under the auspices of the IPTA.

## ACKNOWLEDGMENTS

This work has been carried out by the NANOGrav collaboration, which is part of the International Pulsar Timing Array. The NANOGrav project receives support from National Science Foundation (NSF) Physics Frontiers Center award number 1430284. This work made use of the Super Computing System (Spruce Knob) at West Virginia University (WVU), which are funded in part by the National Science Foundation EPSCoR Research Infrastructure Improvement Cooperative Agreement #1003907, the state of West Virginia (WV EPSCoR via the Higher Education Policy Commission) and WVU. Some of our models utilized the IllustrisTNG simulations public data release (<http://www.tng-project.org/>; Nelson et al. 2019), and sublink merger trees (Rodriguez-Gomez et al. 2015).

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. JS and MV acknowledge support from the JPL RTD program. SBS acknowledges support for this work from NSF grants #1458952 and #1815664. SBS is a CIFAR Azrieli Global Scholar in the Gravity and the Extreme Universe program. SRT acknowledges support from NSF grant AST-#2007993, and a Dean’s Faculty Fellowship from Vanderbilt University’s College of Arts & Science. TTP acknowledges support from the MTA-ELTE Extragalactic Astrophysics Research Group, funded by the Hungarian Academy of Sciences (Magyar Tudományos Akadémia), that was used during the development of this research.

TD and ML acknowledge NSF AAG award number 2009468.

*Software:* ENTERPRISE (Ellis et al. 2019), enterprise\_extensions (Taylor et al. 2018), hasasia

(Hazboun et al. 2019b), libstempo (Vallisneri 2020), matplotlib (Hunter 2007), PTMCMC (Ellis & van Haasteren 2017)

## REFERENCES

- Accadia, T., Acernese, F., Alshourbagy, M., et al. 2012, *Journal of Instrumentation*, 7, 3012, doi: [10.1088/1748-0221/7/03/P03012](https://doi.org/10.1088/1748-0221/7/03/P03012)
- Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021, *ApJS*, 252, 4, doi: [10.3847/1538-4365/abc6a0](https://doi.org/10.3847/1538-4365/abc6a0)
- Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, *PhRvD*, 79, 084030, doi: [10.1103/PhysRevD.79.084030](https://doi.org/10.1103/PhysRevD.79.084030)
- Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2016, *ApJ*, 821, 13, doi: [10.3847/0004-637X/821/1/13](https://doi.org/10.3847/0004-637X/821/1/13)
- Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2018, *ApJ*, 859, 47, doi: [10.3847/1538-4357/aabd3b](https://doi.org/10.3847/1538-4357/aabd3b)
- Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, *ApJL*, 905, L34, doi: [10.3847/2041-8213/abd401](https://doi.org/10.3847/2041-8213/abd401)
- Bailes, M., Barr, E., Bhat, N. D. R., et al. 2016, in *MeerKAT Science: On the Pathway to the SKA*, 11. <https://arxiv.org/abs/1803.07424>
- Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, *nat*, 287, 307, doi: [10.1038/287307a0](https://doi.org/10.1038/287307a0)
- Blasi, S., Brdar, V., & Schmitz, K. 2021, *Phys. Rev. Lett.*, 126, 041305, doi: [10.1103/PhysRevLett.126.041305](https://doi.org/10.1103/PhysRevLett.126.041305)
- Burke-Spolaor, S., Taylor, S. R., Charisi, M., et al. 2019, *A&A Rv*, 27, 5, doi: [10.1007/s00159-019-0115-7](https://doi.org/10.1007/s00159-019-0115-7)
- Chamberlin, S. J., Creighton, J. D. E., Siemens, X., et al. 2015, *PhRvD*, 91, 044048, doi: [10.1103/PhysRevD.91.044048](https://doi.org/10.1103/PhysRevD.91.044048)
- Charisi, M., Bartos, I., Haiman, Z., et al. 2016, *MNRAS*, 463, 2145, doi: [10.1093/mnras/stw1838](https://doi.org/10.1093/mnras/stw1838)
- Chen, Y.-C., Liu, X., Liao, W.-T., et al. 2020, *arXiv e-prints*, arXiv:2008.12329. <https://arxiv.org/abs/2008.12329>
- De Luca, V., Franciolini, G., & Riotto, A. 2021, *Phys. Rev. Lett.*, 126, 041303, doi: [10.1103/PhysRevLett.126.041303](https://doi.org/10.1103/PhysRevLett.126.041303)
- Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, *ApJ*, 762, 94, doi: [10.1088/0004-637X/762/2/94](https://doi.org/10.1088/0004-637X/762/2/94)
- Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, *MNRAS*, 458, 3341, doi: [10.1093/mnras/stw483](https://doi.org/10.1093/mnras/stw483)
- Detweiler, S. 1979, *ApJ*, 234, 1100, doi: [10.1086/157593](https://doi.org/10.1086/157593)
- Ellis, J., & Lewicki, M. 2021, *Phys. Rev. Lett.*, 126, 041304, doi: [10.1103/PhysRevLett.126.041304](https://doi.org/10.1103/PhysRevLett.126.041304)
- Ellis, J., & van Haasteren, R. 2017, *jellis18/PTMCMCSampler*: Official Release, doi: [10.5281/zenodo.1037579](https://doi.org/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*. <http://ascl.net/1912.015>
- Ferrarese, L., & Merritt, D. 2000, *ApJL*, 539, L9, doi: [10.1086/312838](https://doi.org/10.1086/312838)
- Folkner, W. M., & Park, R. S. 2018, *Planetary ephemeris DE438 for Juno*, Tech. Rep. IOM 392R-18-004, Jet Propulsion Laboratory, Pasadena, CA
- Foster, R. S., & Backer, D. C. 1990, *ApJ*, 361, 300, doi: [10.1086/169195](https://doi.org/10.1086/169195)
- Gebhardt, K., Bender, R., Bower, G., et al. 2000, *ApJL*, 539, L13, doi: [10.1086/312840](https://doi.org/10.1086/312840)
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, 445, 175, doi: [10.1093/mnras/stu1654](https://doi.org/10.1093/mnras/stu1654)
- Graham, M. J., Djorgovski, S. G., Stern, D., et al. 2015, *MNRAS*, 453, 1562, doi: [10.1093/mnras/stv1726](https://doi.org/10.1093/mnras/stv1726)
- Grishchuk, L. P. 2005, *Physics Uspekhi*, 48, 1235, doi: [10.1070/PU2005v048n12ABEH005795](https://doi.org/10.1070/PU2005v048n12ABEH005795)
- Harry, G. M., & LIGO Scientific Collaboration. 2010, *Classical and Quantum Gravity*, 27, 084006, doi: [10.1088/0264-9381/27/8/084006](https://doi.org/10.1088/0264-9381/27/8/084006)
- Hazboun, J., Romano, J., & Smith, T. 2019a, *The Journal of Open Source Software*, 4, 1775, doi: [10.21105/joss.01775](https://doi.org/10.21105/joss.01775)
- . 2019b, *The Journal of Open Source Software*, 4, 1775, doi: [10.21105/joss.01775](https://doi.org/10.21105/joss.01775)
- Hazboun, J. S., Romano, J. D., & Smith, T. L. 2019c, *PhRvD*, 100, 104028, doi: [10.1103/PhysRevD.100.104028](https://doi.org/10.1103/PhysRevD.100.104028)
- Hellings, R. W., & Downs, G. S. 1983, *ApJL*, 265, L39, doi: [10.1086/183954](https://doi.org/10.1086/183954)
- Hobbs, G. 2013, *Classical and Quantum Gravity*, 30, 224007, doi: [10.1088/0264-9381/30/22/224007](https://doi.org/10.1088/0264-9381/30/22/224007)
- Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010, *Classical and Quantum Gravity*, 27, 084013, doi: [10.1088/0264-9381/27/8/084013](https://doi.org/10.1088/0264-9381/27/8/084013)
- Hunter, J. D. 2007, *Computing in Science and Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Jaffe, A. H., & Backer, D. C. 2003, *ApJ*, 583, 616, doi: [10.1086/345443](https://doi.org/10.1086/345443)
- Joshi, B. C., Arumugasamy, P., Bagchi, M., et al. 2018, *Journal of Astrophysics and Astronomy*, 39, 51, doi: [10.1007/s12036-018-9549-y](https://doi.org/10.1007/s12036-018-9549-y)

- Kelley, L. Z., Blecha, L., & Hernquist, L. 2017a, MNRAS, 464, 3131, doi: [10.1093/mnras/stw2452](https://doi.org/10.1093/mnras/stw2452)
- Kelley, L. Z., Blecha, L., Hernquist, L., Sesana, A., & Taylor, S. R. 2017b, MNRAS, 471, 4508, doi: [10.1093/mnras/stx1638](https://doi.org/10.1093/mnras/stx1638)
- . 2017c, MNRAS, 471, 4508, doi: [10.1093/mnras/stx1638](https://doi.org/10.1093/mnras/stx1638)
- . 2018, MNRAS, 477, 964, doi: [10.1093/mnras/sty689](https://doi.org/10.1093/mnras/sty689)
- Kerr, M., Reardon, D. J., Hobbs, G., et al. 2020, PASA, 37, e020, doi: [10.1017/pasa.2020.11](https://doi.org/10.1017/pasa.2020.11)
- Kramer, M., & Champion, D. J. 2013, Classical and Quantum Gravity, 30, 224009, doi: [10.1088/0264-9381/30/22/224009](https://doi.org/10.1088/0264-9381/30/22/224009)
- Kuroyanagi, S., Takahashi, T., & Yokoyama, S. 2021, JCAP, 2021, 071, doi: [10.1088/1475-7516/2021/01/071](https://doi.org/10.1088/1475-7516/2021/01/071)
- Lasky, P. D., Mingarelli, C. M. F., Smith, T. L., et al. 2016, Physical Review X, 6, 011035, doi: [10.1103/PhysRevX.6.011035](https://doi.org/10.1103/PhysRevX.6.011035)
- Lee, K. J. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain & D. Li, 19
- Lentati, L., Alexander, P., Hobson, M. P., et al. 2013, PhRvD, 87, 104021, doi: [10.1103/PhysRevD.87.104021](https://doi.org/10.1103/PhysRevD.87.104021)
- Lentati, L., Shannon, R. M., Coles, W. A., et al. 2016, MNRAS, 458, 2161, doi: [10.1093/mnras/stw395](https://doi.org/10.1093/mnras/stw395)
- Liu, T., Gezari, S., Ayers, M., et al. 2019, ApJ, 884, 36, doi: [10.3847/1538-4357/ab40cb](https://doi.org/10.3847/1538-4357/ab40cb)
- McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2014, ApJ, 789, 156, doi: [10.1088/0004-637X/789/2/156](https://doi.org/10.1088/0004-637X/789/2/156)
- NANOGrav Collaboration, Arzoumanian, Z., Brazier, A., et al. 2015, ApJ, 813, 65, doi: [10.1088/0004-637X/813/1/65](https://doi.org/10.1088/0004-637X/813/1/65)
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, Computational Astrophysics and Cosmology, 6, 2, doi: [10.1186/s40668-019-0028-x](https://doi.org/10.1186/s40668-019-0028-x)
- Ölmez, S., Mandic, V., & Siemens, X. 2010, PhRvD, 81, 104028, doi: [10.1103/PhysRevD.81.104028](https://doi.org/10.1103/PhysRevD.81.104028)
- Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, MNRAS, 490, 4666, doi: [10.1093/mnras/stz2857](https://doi.org/10.1093/mnras/stz2857)
- Phinney, E. S. 2001, arXiv e-prints, astro. <https://arxiv.org/abs/astro-ph/0108028>
- Rajagopal, M., & Romani, R. W. 1995, ApJ, 446, 543, doi: [10.1086/175813](https://doi.org/10.1086/175813)
- Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, MNRAS, 449, 49, doi: [10.1093/mnras/stv264](https://doi.org/10.1093/mnras/stv264)
- Romano, J. D., & Cornish, N. J. 2017, Living Reviews in Relativity, 20, 2, doi: [10.1007/s41114-017-0004-1](https://doi.org/10.1007/s41114-017-0004-1)
- Romano, J. D., Hazboun, J. S., Siemens, X., & Archibald, A. M. 2020, arXiv e-prints, arXiv:2012.03804. <https://arxiv.org/abs/2012.03804>
- Rosado, P. A., Sesana, A., & Gair, J. 2015, MNRAS, 451, 2417, doi: [10.1093/mnras/stv1098](https://doi.org/10.1093/mnras/stv1098)
- Sampson, L., Cornish, N. J., & McWilliams, S. T. 2015, PhRvD, 91, 084055, doi: [10.1103/PhysRevD.91.084055](https://doi.org/10.1103/PhysRevD.91.084055)
- Sazhin, M. V. 1978, Soviet Astronomy, 22, 36
- Sesana, A., Shankar, F., Bernardi, M., & Sheth, R. K. 2016, MNRAS, 463, L6, doi: [10.1093/mnrasl/slw139](https://doi.org/10.1093/mnrasl/slw139)
- Sesana, A., Vecchio, A., & Colacino, C. N. 2008, MNRAS, 390, 192, doi: [10.1111/j.1365-2966.2008.13682.x](https://doi.org/10.1111/j.1365-2966.2008.13682.x)
- Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, Classical and Quantum Gravity, 30, 224015, doi: [10.1088/0264-9381/30/22/224015](https://doi.org/10.1088/0264-9381/30/22/224015)
- Simon, J., & Burke-Spolaor, S. 2016, ApJ, 826, 11, doi: [10.3847/0004-637X/826/1/11](https://doi.org/10.3847/0004-637X/826/1/11)
- Taylor, S. R., Baker, P. T., Hazboun, J. S., Simon, J. J., & Vigeland, S. J. 2018, enterprise extensions. [https://github.com/nanograv/enterprise\\_extensions](https://github.com/nanograv/enterprise_extensions)
- Taylor, S. R., Gair, J. R., & Lentati, L. 2013, PhRvD, 87, 044035, doi: [10.1103/PhysRevD.87.044035](https://doi.org/10.1103/PhysRevD.87.044035)
- Taylor, S. R., Simon, J., & Sampson, L. 2017, prl, 118, 181102, doi: [10.1103/PhysRevLett.118.181102](https://doi.org/10.1103/PhysRevLett.118.181102)
- Taylor, S. R., Vallisneri, M., Ellis, J. A., et al. 2016, ApJL, 819, L6, doi: [10.3847/2041-8205/819/1/L6](https://doi.org/10.3847/2041-8205/819/1/L6)
- Taylor, S. R., van Haasteren, R., & Sesana, A. 2020, PhRvD, 102, 084039, doi: [10.1103/PhysRevD.102.084039](https://doi.org/10.1103/PhysRevD.102.084039)
- Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, MNRAS, 455, 4339, doi: [10.1093/mnras/stv2143](https://doi.org/10.1093/mnras/stv2143)
- Vallisneri, M. 2020, libstempo: Python wrapper for Tempo2. <http://ascl.net/2002.017>
- Vallisneri, M., Taylor, S. R., Simon, J., et al. 2020, ApJ, 893, 112, doi: [10.3847/1538-4357/ab7b67](https://doi.org/10.3847/1538-4357/ab7b67)
- Vaskonen, V., & Veermäe, H. 2021, Phys. Rev. Lett., 126, 051303, doi: [10.1103/PhysRevLett.126.051303](https://doi.org/10.1103/PhysRevLett.126.051303)
- Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016, MNRAS, 458, 1267, doi: [10.1093/mnras/stw347](https://doi.org/10.1093/mnras/stw347)
- Vigeland, S. J., Islo, K., Taylor, S. R., & Ellis, J. A. 2018, PhRvD, 98, 044003, doi: [10.1103/PhysRevD.98.044003](https://doi.org/10.1103/PhysRevD.98.044003)
- Vigeland, S. J., & Siemens, X. 2016, PhRvD, 94, 123003, doi: [10.1103/PhysRevD.94.123003](https://doi.org/10.1103/PhysRevD.94.123003)
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, nat, 509, 177, doi: [10.1038/nature13316](https://doi.org/10.1038/nature13316)
- . 2014b, MNRAS, 444, 1518, doi: [10.1093/mnras/stu1536](https://doi.org/10.1093/mnras/stu1536)