Search for Continuous Gravitational-wave Signals in Pulsar Timing Residuals: A New Scalable Approach with Diffusive Nested Sampling

Yu-Yang Songsheng[1,2] , Yi-Qian Qian[3], Yan-Rong Li[1] , Pu Du[1] , Jie-Wen Chen[3], Yan Wang[3] ,
Soumya D. Mohanty[4] , and Jian-Min Wang[1,2,5]

[1] Key Laboratory for Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Beijing 100049, People's Republic of China; wangjm@ihep.ac.cn
[2] University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, People's Republic of China
[3] MOE Key Laboratory of Fundamental Physical Quantities Measurements, Hubei Key Laboratory of Gravitation and Quantum Physics, PGMF, Department of Astronomy and School of Physics, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
[4] Department of Physics and Astronomy, The University of Texas Rio Grande Valley, One West University Blvd., Brownsville, Texas 78520, USA
[5] National Astronomical Observatories of China, Chinese Academy of Sciences, 20A Datun Road, Beijing 100020, People's Republic of China

Abstract

Detecting continuous nanohertz gravitational waves (GWs) generated by individual close binaries of supermassive black holes (CB-SMBHs) is one of the primary objectives of pulsar timing arrays (PTAs). The detection sensitivity is slated to increase significantly as the number of well-timed millisecond pulsars will increase by more than an order of magnitude with the advent of next-generation radio telescopes. Currently, the Bayesian analysis pipeline using parallel tempering Markov Chain Monte Carlo has been applied in multiple studies for CB-SMBH searches, but it may be challenged by the high dimensionality of the parameter space for future large-scale PTAs. One solution is to reduce the dimensionality by maximizing or marginalizing over uninformative parameters semianalytically, but it is not clear whether this approach can be extended to more complex signal models without making overly simplified assumptions. Recently, the method of diffusive nested (DNest) sampling has shown capability in coping with high dimensionality and multimodality effectively in Bayesian analysis. In this paper, we apply DNest to search for continuous GWs in simulated pulsar timing residuals and find that it performs well in terms of accuracy, robustness, and efficiency for a PTA including $\mathcal{O}(10^2)$ pulsars. DNest also allows a simultaneous search of multiple sources elegantly, which demonstrates its scalability and general applicability. Our results show that it is convenient and also highly beneficial to include DNest in current toolboxes of PTA analysis.

Unified Astronomy Thesaurus concepts: Pulsar timing method (1305); Gravitational waves (678); Nested sampling (1894)

1. Introduction

The first observation of the gravitational-wave (GW) signal from a binary black hole merger by LIGO in 2015 has opened up a new window for our exploration of the universe (Abbott et al. 2016). Since then, dozens of GW events from mergers of stellar-mass compact objects have been detected by ground-based interferometers (Abbott et al. 2019, 2021), which transforms our understanding of stellar evolution (e.g., Marchant et al. 2016), equation of state of extremely dense matter (e.g., Abbott et al. 2018), etc. With ground-based interferometers, we can only observe the high-frequency regime ($\sim$10–10^3 Hz) of the GW universe. Several missions for space-based interferometers, such as eLISA (Amaro-Seoane et al. 2017), Taiji (Hu & Wu 2017), and TianQin (Luo et al. 2016), are in progress to open up the low-frequency regime of 10^{-4}–10^{-1} Hz. For ultra-low-frequency GWs ($\sim$$10^{-9}$–$10^{-6}$ Hz), a natural galactic-scale detector comprising a network of millisecond pulsars (MSPs), called a pulsar timing array (PTA), is the most promising way (Sazhin 1978; Foster & Backer 1990; Jenet et al. 2006).

The times of arrival (TOAs) of the radio pulses from a rotating MSP can be measured with high precision. Sophisticated timing models that account for pulsar system dynamics, pulsar-observatory astrometry, dispersion delay and general relativity effects, can be fitted to the observed TOAs of pulses (Edwards et al. 2006; Luo et al. 2021). Timing residuals are defined as the differences between the observed TOAs and the TOAs predicted by the best-fit model. The unmodeled effect of a GW passing between Earth and pulsars is to disturb the background spacetime and modulate the propagation time of radio pulses, thus appearing coherently across the timing residuals of an array of pulsars, allowing it to be distinguished from noise or other incoherent unmodeled effects.

Searching for the stochastic GW background generated by numerous unresolved close binaries of supermassive black holes (CB-SMBHs) is one of the major scientific objectives of PTAs. Thanks to the unique quadrupole property of the gravitational radiation, the correlation between timing residuals from a pair of pulsars varies with their angular separation in a distinctive way (Hellings & Downs 1983). Recently, the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) has presented the "12.5 yr" pulsar timing data set of 47 MSPs (Alam et al. 2021). A common-spectrum stochastic process, described by a power law, has been found in the timing behaviors across all pulsars, whereas quadrupole spatial correlation of timing residuals between pulsars is still not significant and smoking gun evidence for the GW background needs further data (Arzoumanian et al. 2020a). Besides NANOGrav, major PTA programs in operation include the European PTA (Desvignes et al. 2016) and Parkes PTA (Kerr et al. 2020), and approximately 100 pulsars have been timed with high precision for GW detection totally. The number will increase dramatically with the operation of next-generation large-scale radio telescope, especially the Five-hundred-meter Aperture Spherical Telescope (FAST; Lee 2016) and the Square Kilometer Array (SKA; Weltman et al. 2020), making a substantial leap in PTA's sensitivity.

Another important goal of PTA programs is searching for signals of continuous GWs from individual CB-SMBHs by fitting the signal model to the timing residuals of pulsars. Such searches have been conducted in multiple studies using a variety of methods, but no individual sources have been identified yet (Zhu et al. 2014; Babak et al. 2016; Aggarwal et al. 2019). It is expected that PTAs including more pulsars timed with higher precision in the future will be sensitive enough to find signals of single resolvable CB-SMBHs in timing residuals. However, the data analysis with a large number of pulsars will be extremely challenging because of the following reason. Because wavelengths of GWs from CB-SMBHs are usually less than uncertainties of distances of pulsars to Earth, each pulsar in the PTA will introduce an unknown pulsar phase parameter in the signal model, making the dimension of the parameter space too high to explore easily. Ellis (2013) introduced a fully Bayesian data analysis pipeline and applied a parallel tempering Markov Chain Monte Carlo (MCMC; see Metropolis et al. 1953; Hastings 1970; Swendsen & Wang 1986; Sharma 2017) method to sample the posterior probability distribution of model parameters. NANOGrav applied the method to the observed timing residuals of about 40 pulsars in their recent studies (Aggarwal et al. 2019; Arzoumanian et al. 2020b), but testing with more pulsars, which will certainly be the case in the near future, is needed. Similarly, Zhu et al. (2016) introduced a more efficient frequentist framework and employed particle swarm optimization (PSO; see Eberhart & Kennedy 1995; Wang & Mohanty 2010) to search over the parameter space that includes the unknown pulsar phases for the maximum of the likelihood. Another approach is to reduce the dimension of parameter space and therefore sampling complexity by maximizing (Babak & Sesana 2012; Ellis et al. 2012; Wang et al. 2015) or marginalizing (Taylor et al. 2014; Wang et al. 2017) the likelihood function over some model parameters (especially uninformative pulsar phases) semianalytically. It had been used to search for signals of continuous GWs by EPTA (Babak et al. 2016) and tested with an extremely large-scale PTA based on the SKA-era PTA that contains 10^3 pulsars (Wang & Mohanty 2017). However, the maximization and marginalization rely largely on the analytical form of the signal model, making it not straightforward to generalize to CB-SMBHs with elliptical or evolving orbits without introducing simplified assumptions. A common assumption is that the pulsar terms do not add up as coherently as Earth terms and so can be discarded for simplicity. Given that assumption, Babak & Sesana (2012) and Petiteau et al. (2013) developed a maximum-likelihood-based method with an implementation of the genetic algorithm (Holland 1975) to resolve multiple CB-SMBHs. As a further step, Bécsy & Cornish (2020) used transdimensional Bayesian inference implemented by reversible jump MCMC (Green 1995) to search for isolated sources and stochastic GW background jointly in PTA data.

To establish a general Bayesian framework for detecting continuous GW with large-scale PTAs, we notice that the diffusive nested sampling (DNest) method proposed by Brewer et al. (2011) would be an appropriate choice. DNest can effectively solve problems arising from high dimensions, multimodal distributions, highly correlated parameters, and phase changes compared to other sampling methods (Brewer & Foreman-Mackey 2018). It has been successfully applied to reconstruct the broad-line region model from reverberation mapping and spectroastrometry data in the research of active galactic nuclei where the dimension of the parameter space can exceed 100 (e.g., Pancoast et al. 2014; Li et al. 2018; Wang et al. 2020).

In this work, we develop a code named TRAINS[6] (Songsheng 2021) to implement the Bayesian inference with DNest, aiming to model the timing residuals induced by GWs from CB-SMBHs. We test the capability of DNest in searching for nanohertz GW signals generated by a mock population of CB-SMBHs in simulated pulsar timing residuals of a large-scale PTA containing 10^2–10^3 pulsars. We fit the signal model to the simulated data and compare probability distributions of model parameters to their input values. For a blind search of a single source, DNest with a fully Bayesian framework can overcome the problem of high dimension caused by pulsar phases and perform as well as that with a marginalization technique in terms of accuracy, robustness, and efficiency. We further apply the method to search for multiple sources simultaneously and find that it can successfully identify several strongest sources across a wide range of locations and frequencies. As more and more electromagnetic observational signatures are applied to find CB-SMBH candidates, the targeted search is becoming increasingly important for reliable estimation of orbital parameters by breaking up degeneracies between them (Wang & Li 2020). As a fully Bayesian method, TRAINS can be easily generalized to analyze pulsar timing residuals and electromagnetic data jointly for targeted searches.

This paper is structured as follows. Section 2 introduces the framework used to generate mock data and sample the probability distribution of model parameters. Section 3 presents searching results in cases of single sources, multiple sources, and targeted sources. Discussions are provided in Section 4, and conclusions are summarized in the last section.

2. Methodology

2.1. Signal Model

The Doppler response to GWs is originally given by Estabrook & Wahlquist (1975), which is applied to a binary source first in Wahlquist (1987). We rederive the timing residuals of a pulsar caused by GWs from a CB-SMBH in Appendix A for readers' convenience, but also refer to Aggarwal et al. (2019) and references therein. Here we only present the main results. The response of a pulsar to a GW source is described by the antenna pattern functions F^+ and $F^\times$,

$$F^+ = \frac{1}{2} \frac{(\hat{N} \cdot \hat{p})^2 - (\hat{E} \cdot \hat{p})^2}{1 + \hat{p} \cdot \hat{k}}$$
$$F^\times = -\frac{(\hat{E} \cdot \hat{p})(\hat{N} \cdot \hat{p})}{1 + \hat{p} \cdot \hat{k}}, \tag{1}$$

where $\hat{k}$ is the unit vector pointing from the source to the observer, $\hat{p}$ is the unit vector pointing from the observer to the pulsar, $\hat{E}$ and $\hat{N}$ are two orthogonal normalized basic vectors in the plane perpendicular to $\hat{k}$, and pointing to the direction of increasing R.A. and decl. respectively.

The effect of the GW on pulsar's residuals measured by the observer at time t can be expressed as

$$s(t) = F^+ \Delta s_+(t) + F^\times \Delta s_\times(t), \tag{2}$$

where $\Delta s_{+,\times}(t) \equiv s_{+,\times}(t) - s_{+,\times}(t_p)$ is the difference between the so-called Earth term $s_{+,\times}(t)$ and pulsar term $s_{+,\times}(t_p)$,
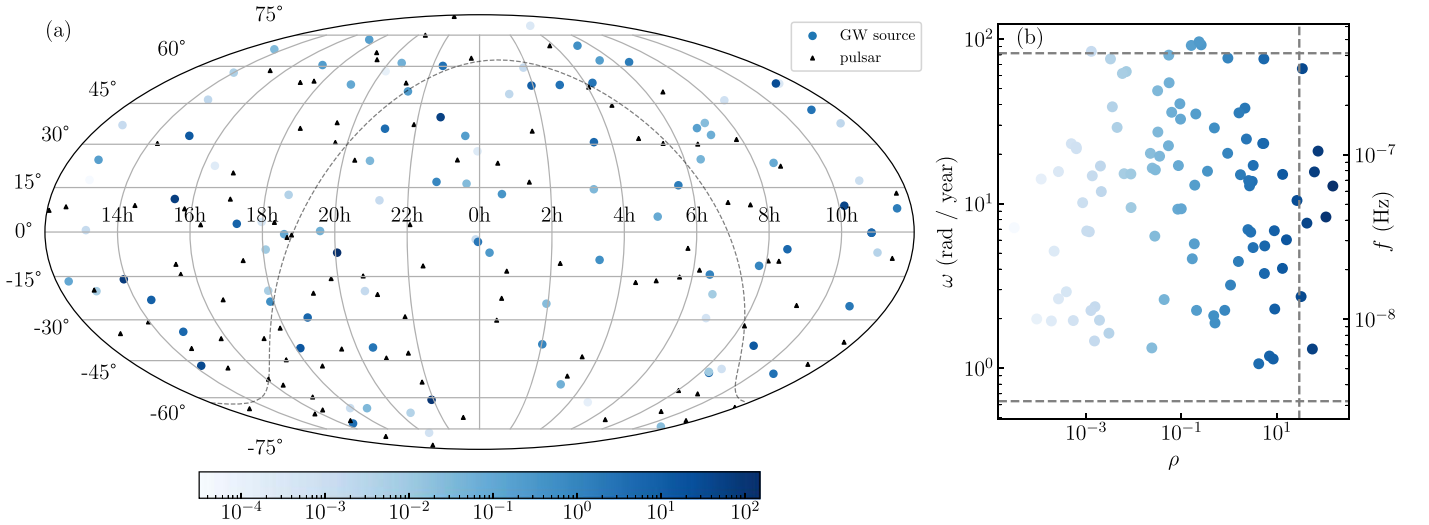

Figure 1. Positions, strengths, and frequencies of simulated GW sources. (a) Distribution of pulsars and CB-SMBHs on the celestial sphere. Black triangles trace positions of pulsars while blue dots trace those of CB-SMBHs. The darkness of blue, as indicated by the color bar, represents the S/N of the GW signal. The dashed gray line denotes the Galactic plane. (b) Distribution of S/Ns and frequencies of GW sources. S/Ns and frequencies span from 3×10^{-5} to 1.5×10^{2} and from 5×10^{-9} Hz to 5×10^{-7} Hz, respectively. The upper and lower dashed lines denote $f = 1/(2\Delta t)$ (Nyquist frequency) and $f = 1/(2T)$ respectively, where δt and T are the cadence and duration of the sampling, respectively. The vertical dashed line marks $\rho = 30$.

induced by the GW at Earth and the pulsar, respectively. The pulsar time t_p is related to Earth time t as

$$t_p = t - d_p(1 + \hat{p} \cdot \hat{k})/c \equiv t - \tau_p, \qquad (3)$$

where d_p is the distance of the pulsar to the observer and c is the speed of light.

For a CB-SMBH in a circular orbit, $s_{+,\times}(t)$ is given by

$$s_+(t) = \frac{(G\mathcal{M})^{5/3}}{c^4 D_L \omega(t)^{1/3}} [\cos 2\psi (1 + \cos^2 \iota) \sin 2\varphi(t)$$
$$- 2\sin 2\psi \cos \iota \cos 2\varphi(t)],$$

$$s_\times(t) = \frac{(G\mathcal{M})^{5/3}}{c^4 D_L \omega(t)^{1/3}} [\sin 2\psi (1 + \cos^2 \iota) \sin 2\varphi(t)$$
$$+ 2\cos 2\psi \cos \iota \cos 2\varphi(t)], \qquad (4)$$

where G is the gravitational constant, $\mathcal{M}$ and D_L are the redshifted chirp mass and luminosity distance of the CB-SMBH, respectively, ι is the inclination angle of the orbital plane, and ψ is the GW polarization angle. The observed orbital angular frequency ω of the CB-SMBH evolves slowly as the GW radiates the energy of the binary away gradually,

$$\omega(t) = \omega_0 \left(1 - \frac{t}{t_{\mathrm{merge}}}\right)^{-3/8}, \qquad (5)$$

where $\omega_0 \equiv \omega(t = 0)$ is the initial orbital angular frequency and t_{merge} is the merger time,

$$t_{\mathrm{merge}} = \frac{5c^5}{256}(G\mathcal{M})^{-5/3}\omega^{-8/3},$$

$$= 4.4 \times 10^4 \left(\frac{\mathcal{M}}{10^9 \, M_\odot}\right)^{-5/3} \left(\frac{\omega}{10^{-8} \, \mathrm{Hz}}\right)^{-8/3} \mathrm{yr}. \qquad (6)$$

The variation of the orbital phase with time $\varphi(t)$ is given by

$$\varphi(t) = \varphi_0 + \frac{8\omega_0 t_{\mathrm{merge}}}{5}\left[1 - \left(1 - \frac{t}{t_{\mathrm{merge}}}\right)^{5/8}\right], \qquad (7)$$

where $\varphi_0 \equiv \varphi(t = 0)$ is the initial orbital phase. For the pulsar term evaluated at the pulsar time t_p, we also define $\omega_p \equiv \omega(-\tau)$ and $\varphi_p \equiv \varphi(-\tau_p)$, and we have

$$\omega(t_p) = \omega_p \left(1 - \frac{t}{t_{\mathrm{merge}} + \tau_p}\right)^{-3/8},$$

$$\varphi(t_p) = \varphi_p + \frac{8\omega_p(t_{\mathrm{merge}} + \tau_p)}{5} \qquad (8)$$

$$\times \left[1 - \left(1 - \frac{t}{t_{\mathrm{merge}} + \tau_p}\right)^{5/8}\right].$$

If the time span of the PTA program is much shorter than the merger time t_{merge}, orbital angular frequencies in the Earth and pulsar terms can be treated as constants, respectively, and variations of orbital phases can be approximated as $\varphi(t) = \varphi_0 + \omega_0 t$ and $\varphi(t_p) = \varphi_p + \omega_p t$, respectively. If the light-travel time between Earth and the pulsar is also much shorter than the merger time, we can further assume $\omega_0 \approx \omega_p$ and $\varphi_p \approx \varphi_0 - \omega_0 \tau_p$.

2.2. Simulation of Pulsar Timing Residuals

The PTA used for our simulation is composed of 100 pulsars, provided by the pulsar simulation for the SKA (Smits et al. 2009). Their positions are marked by the black triangles in Figure 1(a). Timing uncertainties of MSPs depend on rotation periods, flux densities, and profile widths of pulsars, as well as integration times, bandwidths, and sensitivities of telescopes (Wang & Mohanty 2018). For simplicity, we assume the typical timing uncertainty of each pulsar follows a white Gaussian noise with rms value of 10^{-7} s, and it is added to the timing residuals caused by GWs. Other sources of noise, such as pulsar spin noise, dispersion measure variation, and stochastic GW background, have complicated correlations in space or time (Tiburzi et al. 2016) and could impact CB-SMBH searches in practice (Bécsy & Cornish 2020). As a proof of concept of the DNest method, we ignore them here, and

integrating them into our model will be subject of a future work.

The GWs are generated by 100 CB-SMBHs uniformly distributed on the celestial sphere, as shown by the blue circles in Figure 1 (a). The luminosity distance (D_L) of the sources ranges from 100 to 1000 Mpc. The cubic of the distance is generated uniformly to achieve a homogeneous distribution of sources in space. The redshifted chirp mass ($\mathcal{M}$) of the CB-SMBHs follows a log-uniform distribution, and its minimum and maximum are $10^6 M_\odot$ and $10^{10} M_\odot$, respectively. The observed angular frequencies of GWs (ω_{gw}) are set to range from 1 to 100 rad yr^{-1} log-uniformly. As a result, the characteristic amplitude of timing residuals caused by a single source,

$$
\zeta = \frac{(G\mathcal{M})^{5/3}}{c^4 D_L (\pi f_{gw})^{1/3}} = 5.5 \times 10^{-8} \left(\frac{\mathcal{M}}{10^9 M_\odot} \right)^{5/3}
$$
$$
\left(\frac{D_L}{100 \text{ Mpc}} \right)^{-1} \left(\frac{f_{gw}}{10^{-8} \text{ Hz}} \right)^{-1/3} \text{ s}, \tag{9}
$$

will span a wide range from 2.6×10^{-6} to 1.2×10^{-14} s. The cosine of the inclination angle (ι) between the binary orbital plane and the plane of the sky and the GW polarization angle (ψ) are uniformly drawn from $[-1, 1]$ and $[0, \pi]$, respectively. Finally, we also assume the initial orbital phase (Φ_0) is uniformly distributed from 0 to π.

We emphasize here that our mock sample of CB-SMBHs is quite artificial and does not represent realistic populations in the universe. However, this is acceptable because our major goal is to demonstrate the capability of our searching algorithm in large-scale PTAs rather than to make robust predictions for realistic observations. We also ignore the evolution of binary orbits in order to compare the method with the technique of maximization and marginalization, but a test of the method in the regime of evolving orbits will be presented in Appendix B.

Given the information on pulsars and CB-SMBHs, timing residuals for each pulsar can be calculated by the signal model of nonevolving binaries. The timing residual data are sampled with the cadence of 14 days for a period of 5 yr. The GW angular frequencies of most sources are below the Nyquist frequency of the sampling to avoid frequency leak, as shown in Figure 1(b). However, the frequency reach of a PTA-based GW search can be extended far beyond the Nyquist frequency associated with the single pulsar's cadence by exploiting asynchronous observations from multiple pulsars (Wang et al. 2021), which will be explored in the future.

To quantify the relative strength of the signal for each source, we define the network signal-to-noise ratio (S/N) ρ for a source i as

$$
\rho_i = \left[\sum_{I=0}^{N_p} \sum_{k=0}^{N_t} \left(\frac{s_{iI}(t_k)}{\sigma_I} \right)^2 \right]^{1/2}, \tag{10}
$$

where $s_{iI}(t_k)$ is the timing residual generated by source i for pulsar I at time t_k, σ_I is the timing uncertainty of pulsar I, and N_p and N_t are the number of pulsars and observation times, respectively. In our simulated data, S/Ns of CB-SMBHs span from 1.3×10^{-5} to 1.5×10^2, as shown in Figure 1(b), and eight sources have S/Ns above 30.

Table 1
Parameters Used in the CB-SMBH Model

Parameters	Meanings	Prior ranges	Prior probability
α	R.A. of the CB-SMBH	$(0, 2\pi)$	uniform
δ	Decl. of the CB-SMBH	$(-\pi/2, \pi/2)$	$\sin\delta$ uniform
ι	Inclination of the orbital plane	$(0, \pi)$	$\cos\iota$ uniform
ψ	GW polarization angle	$(0, \pi)$	uniform
φ_0	Initial orbital phase	$(0, \pi)$	uniform
ζ	Amplitude of timing residuals	$(10^{-15}, 10^{-6})$ s	log uniform
ω_{gw}	Observed angular frequency of the GW	$(1, 100)$ rad yr^{-1}	log uniform
φ_p	pulsar phase	$(0, \pi)$	uniform

2.3. Bayesian Analysis

In order to fit the nonevolving CB-SMBH model with N_s sources to the data, we have seven free parameters for each source, as summarized in Table 1. Note that the chirp mass $\mathcal{M}$ and luminosity distance D_L of the source are degenerate, and only the amplitude of timing residuals ζ is included in the fitting. When calculating timing residuals for each pulsar, we need to figure out a pulsar phase for each pair of source and pulsar,

$$
\varphi_p = \varphi_0 - \frac{\omega_{gw} d_p (1 - \cos\theta)}{2c}, \tag{11}
$$

where θ is the opening angle between the source and the pulsar subtended at the observer. However, the uncertainties of d_p are much larger than the GW wavelengths when dealing with real data. We therefore treat each pulsar phase as a free parameter with a uniform prior in our mock data analysis (Corbin & Cornish 2010).

Now, we have $N_s(7 + N_p)$ free parameters in our models. The likelihood function of data set $\mathcal{D}$ for parameter $\{\Theta\}$ in model M is therefore

$$
P(\mathcal{D}|\Theta, M) = \prod_{I=1}^{N_p} \prod_{k=1}^{N_t} \frac{1}{\sqrt{2\pi\sigma^2}}
$$
$$
\exp\left\{ -\frac{[r_I(t_k) - r_{I,M}(t_k, \Theta)]^2}{2\sigma_I^2} \right\}, \tag{12}
$$

where N_t is the number of data points in timing residuals for each pulsar, σ_I is the timing uncertainty of the pulsar, and $r_I(t_k)$ and $r_{I,M}(t_k, \Theta)$ are the timing residual for pulsar I at timing t_k of the data and model, respectively. In light of Bayes' theorem, the posterior probability distribution for Θ is given by

$$
P(\Theta|\mathcal{D}, M) = \frac{P(\Theta|M)P(\mathcal{D}|\Theta, M)}{P(\mathcal{D}|M)}, \tag{13}
$$

where $P(\Theta|M)$ is the prior distribution of the model parameter and $P(\mathcal{D}|M)$ is a normalization factor.

For $N_s = 1$, if the seven parameters of the source s are given, the likelihood function, Equation (12), can be maximized analytically over the pulsar phase parameters ϕ (Wang et al. 2015). If we use the new likelihood function

$$
P(\mathcal{D}|s, M) = \max_{\{\phi\}} P(\mathcal{D}|\Theta, M), \quad \Theta = \{s, \phi\} \tag{14}
$$

instead of Equation (12), the dimension of the parameter space will be reduced from $N_p + 7$ to 7. In such a case, traditional optimization or sampling algorithm, such as PSO and MCMC, can be applied to find the parameter s of the source.

A more mathematically rigorous way to get the probability distribution of a subset s of model parameters Θ is to marginalize nuisance parameters ϕ out, namely,

$$P(s|\mathscr{D}, M) = \int P(\Theta|\mathscr{D}, M)d\phi$$
$$= \frac{P(s|M)}{P(\mathscr{D}|M)} \int P(\mathscr{D}|\Theta, M)P(\phi|M)d\phi. \quad (15)$$

Here, we assume the prior of the source parameters s and that of pulsar phase parameters ϕ are independent. When $N_s = 1$, the integration at the right-hand side of Equation (15) can be worked out semianalytically and integrated numerically (Wang et al. 2017). Again, the dimension of the parameter space reduces to 7.

However, if $N_s > 1$, or the orbit of the CB-SMBH is elliptical or evolving, the maximization or marginalization of the likelihood function over pulsar phases cannot be performed analytically. The dimension of the parameter space will be very large ($\mathcal{O}(10^2)$–$\mathcal{O}(10^3)$) as the number of available pulsars in the timing array increases. In this case, the DNest method (Brewer et al. 2011)[7] will be an appropriate choice to tackle the challenge of high dimensionality.

2.4. Diffusive Nested Sampling

To evaluate the evidence Z of a model M for data set $\mathscr{D}$:

$$Z \equiv P(\mathscr{D}|M) = \int P(\mathscr{D}|\Theta, M)P(\Theta|M)d\Theta, \quad (16)$$

Skilling (2004) proposed the nested sampling method. Note that

$$P(M|\mathscr{D}) = \frac{P(\mathscr{D}|M)P(M)}{P(\mathscr{D})}. \quad (17)$$

It is generally reasonable to assign equal priors for models under consideration, as a result, the evidence can be directly used for model selection.

Nested sampling first samples n particles Θ_i from the prior $P(\Theta|M)$ and evaluates the likelihood $L(\Theta_i) \equiv P(\mathscr{D}|\Theta_i, M)$ of each point. The particle with the lowest likelihood L_1 is recorded and replaced by a new one drawn from the prior but under a constraint $L(\Theta) > L_1$ via MCMC. Mathematically, the new prior can be expressed as

$$p_j(\Theta) = \frac{P(\Theta|M)}{X_j} \begin{cases} 1 \text{ if } P(\mathscr{D}|\Theta, M) > L_j \\ 0 \text{ if } P(\mathscr{D}|\Theta, M) <= L_j \end{cases},$$
$$X_j \equiv \int_{L(\Theta)>L_j} P(\Theta|M)d\Theta \quad (18)$$

(here $j = 1$). Then we record the minimum of likelihoods L_2 and repeat the process to build up nested likelihood levels $L_1 < L_2 < \cdots$. It can be proved that the prior probability

---

enclosed by jth level X_j has an expectation of $E(X_j) = \exp(-j/n)$. Consequently, the evidence can be approximated by

$$Z \approx \sum_j L_j(X_{j-1} - X_j). \quad (19)$$

As particles are constrained to higher levels with higher likelihoods, a posterior sample of Θ can also been obtained from those recorded points with the lowest likelihoods in each iteration (Skilling 2004).

However, classical nested sampling performs unsatisfactorily when sampling multimodal or highly correlated distributions. Particles may be stuck in local maximum and fail to explore the whole parameter space if a distribution has isolated "islands" with high likelihoods. DNest makes improvements by assigning a label j to each particle indicating which particular level it is currently constrained by, where $j = 0, 1, \cdots, j_{max}$ and j_{max} is the current top level. Instead of sampling all points according to the restricted prior $p_{j_{max}}$, DNest samples particles' positions Θ and labels j concurrently by a mixed distribution:

$$p(\Theta, j) = w_j p_j(\theta), \quad (20)$$

where w_j is the chosen weighting scheme. In the process of creating levels, the exponentially decaying weights

$$w_j \propto \exp\left(\frac{j - j_{max}}{\Lambda}\right) \quad (21)$$

are adopted, where Λ is a backtrack length controlling how far particles can diffuse to lower levels. As a result, particles can diffuse to lower levels where the parameter space is more connected so that particles can easily explore the whole space. The weights are adjusted to become uniform once the desired number of levels has been generated, to further sample the posterior distribution of Θ.

If the levels created are insufficient, we may not see the peak in the series $L_j(X_{j-1} - X_j)$, i.e., we have not found the level that contains the most posterior probability to obtain a large enough posterior sample. In this case, we can introduce a "temperature" T and modify the likelihood of each level to $\tilde{L}_j = L_j \exp(-T)$. If the peak appears in the series $\tilde{L}_j(X_{j-1} - X_j)$, we can get a posterior sample of the modified posterior probability, which is a broadened version of the original one. The uncertainties of model parameters would be overestimated, but it can still extract the information of parameters from data as much as possible even though the standard sampling with $T = 1$ fails to find out the spiky peak of the posterior probability.

3. Results

3.1. Blind Search: a Single Source

First, we fit the CB-SMBH model containing one source to the pulsar timing residuals using three different methods: (1) treating pulsar phases as free parameters (FP); (2) maximizing the likelihood function over pulsar phases (MP); and (3) averaging the likelihood function over pulsar phases (AP). The posterior probability distributions in the three cases are all sampled by the DNest algorithm. The CB-SMBH with the highest S/N ($\rho = 151$) is identified as we expected. Posterior distributions of major parameters of the source are illustrated in Figure 2. Uncertainties and biases of the recovered source's position are also listed in the first row of Table 2.
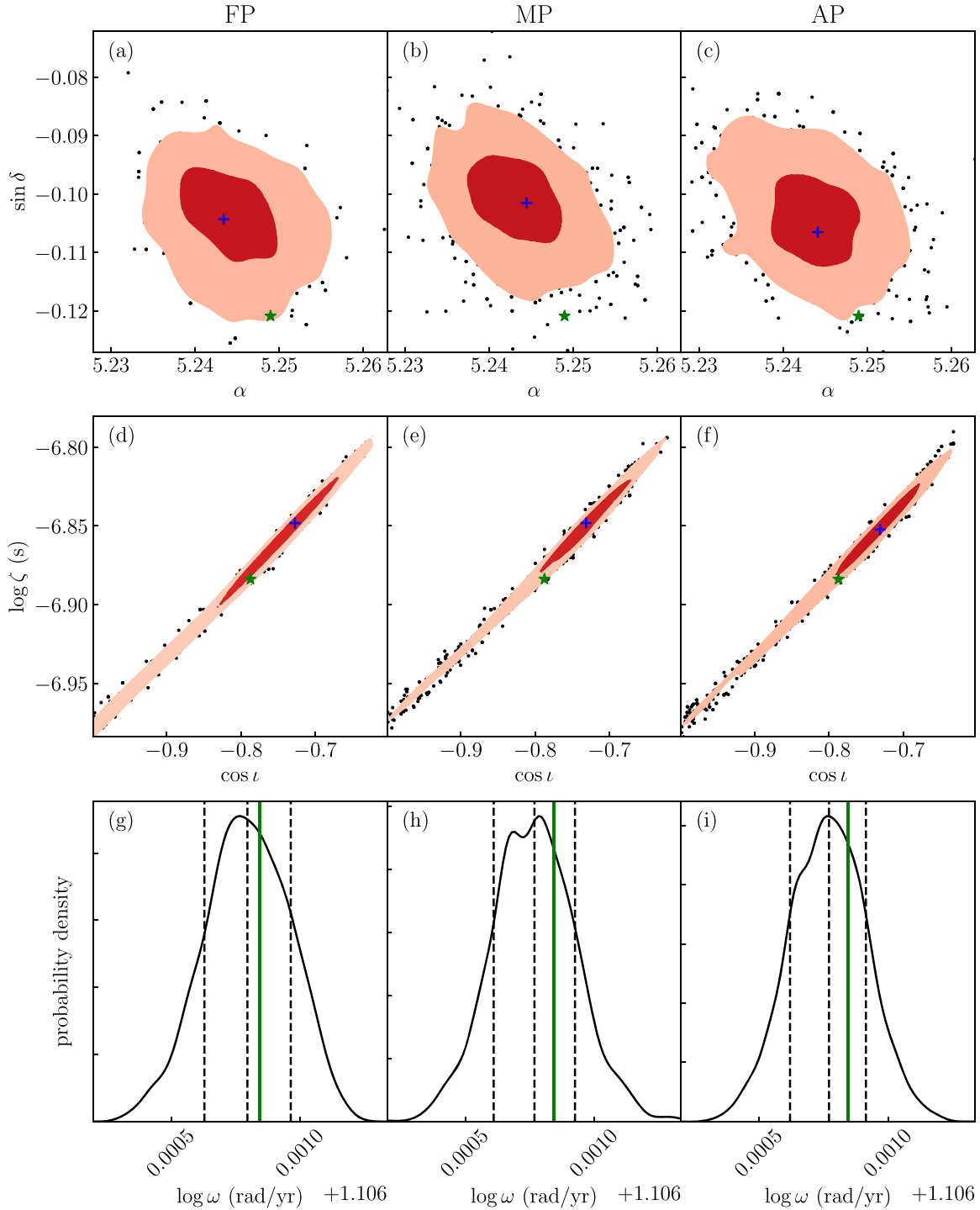
Figure 2. Posterior probability distributions of some model parameters of the GW source. The first, second, and third columns are distributions reconstructed by the FP, MP, and AP methods, respectively. The first row shows the distributions of the right accession and decl. of the source. The red and pink colors represent the 1σ and 2σ regions of the distribution. Black dots are posterior samples outside the 2σ region. The green star marks the input parameters, while the blue cross indicates the peak of the distribution. The second row illustrates the distributions of the inclination and amplitude of the source. The last row shows probability density functions of the angular frequency of the source. The dashed lines are the 16%, 50% and 84% quantiles, while the green line is the input frequency. All probability density functions are calculated using the `Gaussian_kde` method of the `scipy.stats` module (Virtanen et al. 2020) of `Python` from posterior samples reconstructed by the DNest method.

The first row of Figure 2 shows the distribution of R.A. and decl. of the source. For the FP and AP methods, the actual location of the source lies near or within the 2σ region (enclosing 86% of the total probability) of the posterior sample, while the deviation in the MP method is a little bit larger. The second row shows the distribution of the amplitude of timing

residuals and the inclination of the orbital plane. The input value is at or near the edge of the 1σ contour (enclosing 40% of the total probability), and the long tail of the distribution might be caused by the strong correlation between inclinations and amplitudes. The last row displays the probability density of the angular frequency of the GW. The input frequencies lies within
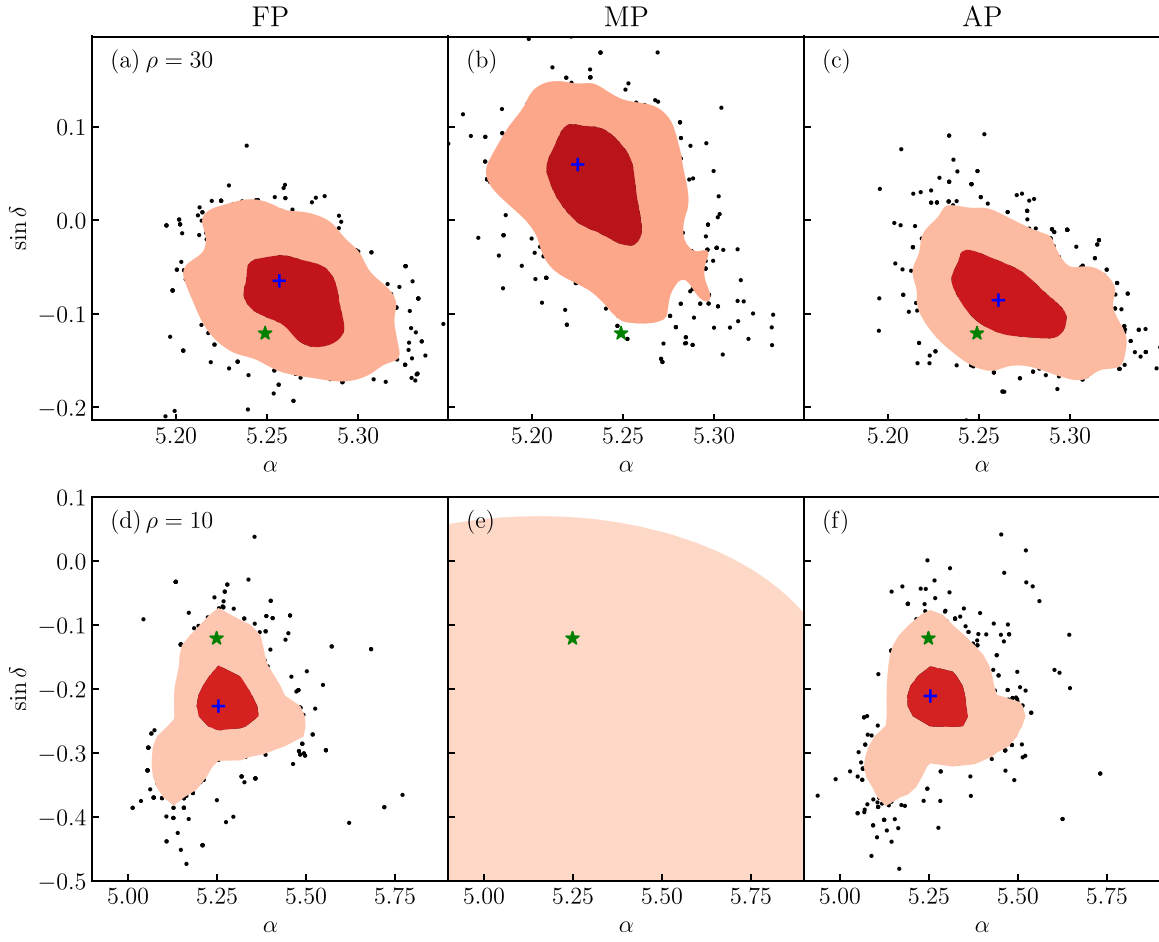
Figure 3. Posterior probability distributions of the R.A. and decl. of the GW source. The first, second, and third columns are the distributions reconstructed by the FP, MP, and AP methods, respectively. The red and pink colors represent the 1σ and 2σ regions of the distribution. Black dots are posterior samples outside the 2σ region. The green star marks the input parameters, while the blue cross indicates the peak of the distribution. In the first row, the S/N of the source is 30, while in the second row, the S/N is only 10. Note that in panel (e), the distribution is very broad and only a portion is shown.

Table 2
Uncertainties and Biases of the Recovered Source's Position for Different S/Ns and Searching Methods

S/N	Uncertainty (deg²)			Bias (deg)		
	FP	MP	AP	FP	MP	AP
151	1.75	1.61	1.78	1.00	1.14	0.87
30	52.9	65.5	49.4	3.25	10.4	2.14
10	223	11026	233	6.07	130	5.17

Note. The uncertainty of the source's position is defined as the size of the 2σ region of the posterior sample (enclosing 86% of the total probability), and the bias is defined as the angular distance between the input location and the center of the posterior distribution.

the 1σ range of the distribution, and the uncertainty is only 2×10^{-4} dex, making it the best-determined model parameters. In a word, DNest can effectively cope with high dimensionality owing to pulsar phases when the S/N of the GW source is relatively high.

In order to test the performance of DNest in search of a relatively weak source, we increase the timing uncertainties of all pulsars to 5×10^{-7} s, bringing down the S/N of the strongest source to 30. As can be seen from the first row of Figure 3 and the second row of Table 2, the source can still be located, but with larger uncertainties and biases. Posterior

distributions of CB-SMBH's location obtained by the FP and AP methods are similar. Distribution obtained by the MP method is a little deformed and has a slightly larger uncertainty and bias.

We further increase the timing uncertainties to 1.5×10^{-6} s, reducing the S/N to 10. The result is illustrated in the second row of Figure 3 and the third row of Table 2. Both FP and AP methods work well, but the posterior distribution recovered by the MP method is extremely scattered, failing to locate the source. The failure of MP in the case of low S/N may be related to the ill posedness of the inverse problem of estimating the large number of pulsar phase parameters (e.g., Wang et al. 2015). The likelihood function is expected to be highly degenerate over the pulsar phases and contain strong secondary maxima. As S/N decreases, the locations of such secondary maxima are more likely to become the global maximum under small perturbations from the noise in the data. The jumping of locations of global maximum to radically different values will lead to a large bias and uncertainty in parameter estimation. We conclude that the performance of DNest with FP in high-dimensional parameter space is still competitive even if the signal of the source is weak.

In the process of getting the results above, the FP and MP methods are the most efficient, but the MP method performs poorly in the case of weak signals. The AP method takes about three times as long as the FP and MP methods. The FP method
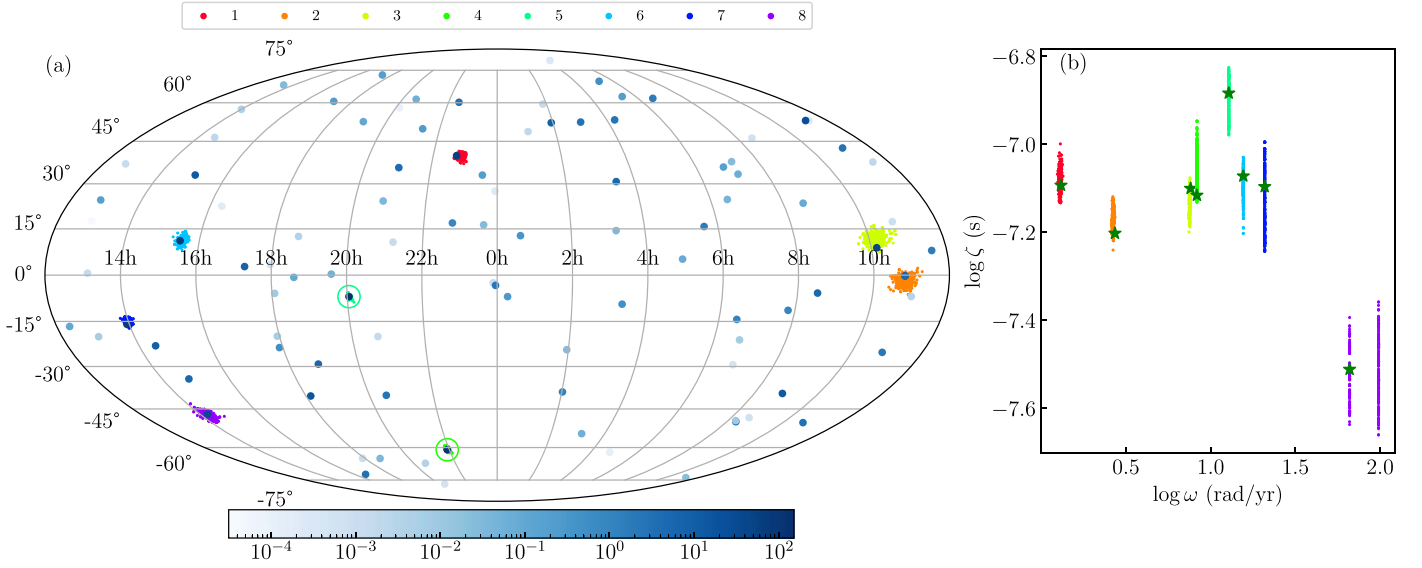
Figure 4. Posterior samples of some model parameters of all GW sources found by DNest. (a) Posterior samples of locations of different GW sources on the celestial sphere. Large blue dots mark positions of CB-SMBHs. The darkness of blue, as indicated by the color bar, represents the S/N of the GW signal. Small colorful dots mark posterior samples of right accessions and declinations of different sources. Eight sources with the highest S/Ns are identified. The posterior distributions of two sources are too concentrated to recognize in the figure. We draw two circles around them to indicate their positions. (b) Posterior samples of amplitudes and angular frequencies of different GW sources. Green stars mark true values while colorful dots mark posterior samples. Different sources are sorted by their frequencies.

with DNest is not only efficient and robust, but also applicable to a wider range of cases without approximations, such as elliptical orbits, slowly evolving orbits, and especially the simultaneous search of multiple sources. We emphasize here that MP and AP are originally derived from the maximum likelihood approach, which produces point estimates of parameters rather than samples from their posterior distributions. Maximization can be performed with fast heuristic optimizers, making point estimation with MP and AP much more efficient.

3.2. Blind Search: Multiple Sources

In this section, we try to locate all CB-SMBHs with $\rho > 30$ simultaneously from pulsar timing residuals. To this end, we include eight GW sources in our model, each with its own pulsar phases, increasing the dimension of the parameter space to 856. We note that if we switch the order of sources in the model, the signal remains unchanged. To avoid this ambiguity, we demand that $\omega_1 < \cdots < \omega_i < \cdots < \omega_8$, where ω_i is the angular frequency of the GW generated by the ith source. This can be achieved by sorting the angular frequencies of all sources after updating the particles' positions in parameter space via MCMC.

Posterior samples of model parameters of all GW sources found by DNest are shown in Figure 4. As illustrated in Figure 4(a), eight sources with the highest S/Ns are identified. The higher the S/N, the more accurate the location of the GW source. Figure 4(b) presents posterior samples of the amplitudes and angular frequencies of these sources. Angular frequencies of different sources are well separated in posterior samples, implying that it is appropriate to eliminate ambiguity of exchanging sources by sorting frequencies. There are double peaks in the posterior sample of the frequency of the eighth source, and the minor peak matches the input frequency. The frequency of the eighth source is close to the Nyquist frequency of the data sampling, raising the problem of aliasing. We also notice that the frequency uncertainties of the first two sources

are relatively large when compared to those of other sources, probably because the signals they produce change too slowly to have enough cycles for precise measurement of periods.

An interesting fact happens if we "flatten" the posterior probability distribution artificially by setting the "temperature" of DNest to $T = 2.3$. The posterior probability density reaches its maximum when parameters of the eight sources in our model match those of the eight strongest sources. However, the probability also has a local maximum if parameters of the eighth strongest source are replaced by those of the ninth strongest one. This minor peak can be sampled if we "flatten" the posterior probability distribution properly. In the posterior samples, some points correspond to the major peak while others to the minor peak. Therefore, parameters of a particular source may have different positions in these two kinds of parameter vectors. For example, as shown in Figure 5(a), the source at $\sim$(16 hr, 15°) is sampled by both blue and cyan dots.

Sources in the posterior sample can be clustered by k-means algorithm (Lloyd 1982) according to their right ascensions, declinations, and frequencies. The result is shown in Figure 5(b). Evidently, nine sources are identified, including the source at the upper-right corner which have not been identified in Figure 4. The S/N of the newly identified source is 26. By increasing the "temperature" of DNest properly, we can even find more sources than there are in the model, but at a cost of larger uncertainties.

Generally speaking, DNest shows its great power in multisource search, despite the huge dimension of the parameter space.

3.3. Targeted Search

In a blind search, the strongest sources are always found first. To search for a relatively weak source in timing residuals, we have three methods: (1) Search for the weak source with strong sources simultaneously; (2) search for strong sources first, then subtract their signals and search for the weak source; (3) use the electromagnetic information to conduct targeted
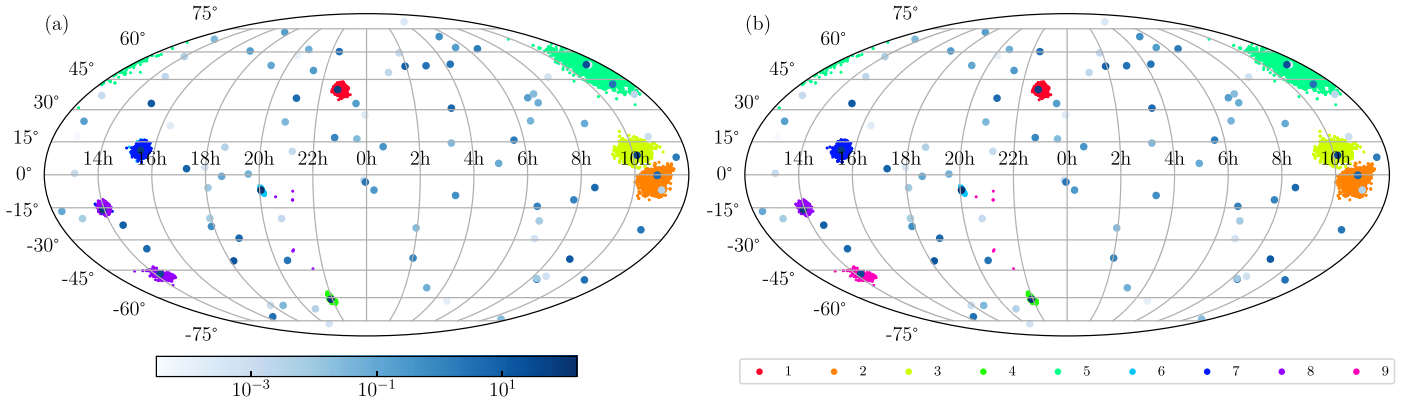
Figure 5. Posterior samples of locations of all GW sources found by DNest. Large blue dots mark positions of CB-SMBHs. Small colorful dots mark posterior samples of right accessions and declinations of different sources. (a) Before clustering, colors of small dots only represent labels of the parameters in the model, which may not correspond to different sources one to one, i.e., one color may correspond to multiple sources, or different colors may correspond to one source. (b) After clustering, colors of small dots correspond to different sources one to one. The clustering is performed by the KMeans method of the Python module sklearn. cluster (Pedregosa et al. 2011).
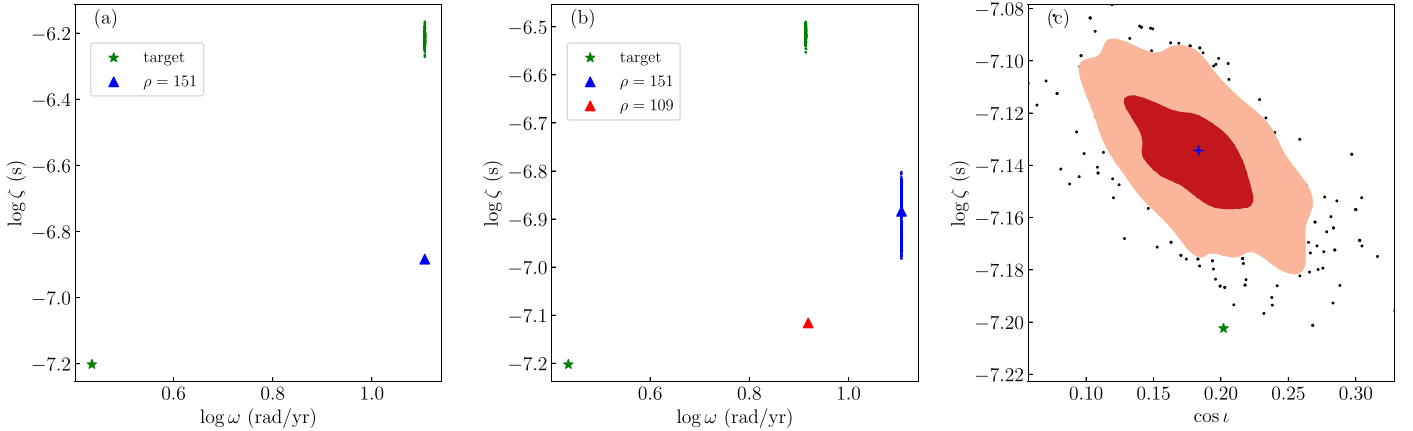

Figure 6. Posterior samples of part of the parameters in a targeted search. (a) Posterior samples obtained with fixed R.A. and decl. in the model. Green dots are posterior samples of the amplitude and angular frequency of the target while the green star marks the true values. The blue triangle indicates the amplitude and frequency of the source with the highest S/N. (b) Posterior samples obtained by including one more source in the model. Green and blue dots are posterior samples of the target and the strongest source, respectively. The red triangle marks the second strongest source. (c) Posterior samples obtained with fixed R.A., decl., and angular frequency in the model. The red and pink colors represent the 1σ and 2σ regions of the distribution of the inclination and amplitude. Black dots are posterior samples outside the 2σ region. The green star marks the input values, while the blue cross indicates the peak of the distribution.

search for known CB-SMBH candidates. The first two methods spend most of their time in resolving strong sources. If we are only interested in known candidates, the last method will be the most efficient one.

Currently, there are more than 100 CB-SMBH candidates identified by characteristic electromagnetic information, such as periodic light curves (e.g., Graham et al. 2015; Charisi et al. 2016) and emission-line profiles (e.g., Bon et al. 2012; Eracleous et al. 2012; Li et al. 2016, 2019). The detectability of CB-SMBHs traced by periodic light curves with realistic PTAs has also been studied (Xin et al. 2021). Unlike a blind search, the location of the source to be identified is already known from electromagnetic observations. As an example, we try to search for the source at (11 hr, 0°), i.e., the source identified by orange dots in Figure 4(a). The S/N of the source is about 32.

First, we fix the R.A. and decl. of the source in the model and fit the model to the timing residuals using the DNest method. The result is shown in Figure 6(a). The posterior sample of the amplitude and frequency of the source does not match those of the target at all. The reconstructed amplitude is

an order of magnitude larger than the true value, and the reconstructed frequency is close to that of the source with the highest S/N. In the process of fitting, the model attempts to match the overall amplitude and main periodic characteristics of timing residuals, though with misplaced phases. The difference of this "fake" signal from the data is still much smaller than the difference of the true signal with a small amplitude from the data. To avoid the impact of the strongest source on the targeted search, we include one more source in the signal model. The result is shown in Figure 6(b). The strongest source has been successfully resolved. However, the reconstructed amplitude of the target is still an order of magnitude larger, and the recovered frequency matches that of the second strongest source. We can infer that the target can be resolved only if all stronger sources are included in the signal model.

Now we further fix the frequency of the source in the model, as the orbital period of the CB-SMBH candidate can be constrained by periodic signals in electromagnetic observations. The result is shown in Figure 6(c). The true amplitude and inclination of the target lie outside the 2σ region of the
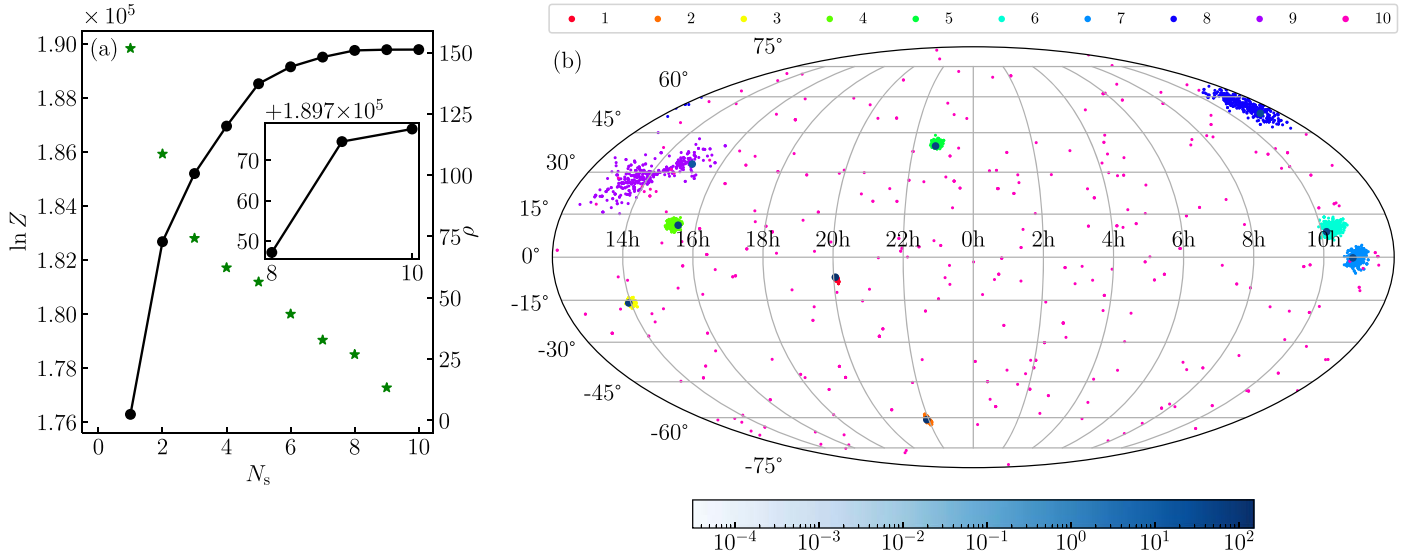
Figure 7. (a) Logarithm of evidence of signal models with different numbers of GW sources. The black line describes the variation of Bayesian evidence as the number of GW sources in the signal model increases. Green stars mark S/Ns of the 10 sources used to generate the timing residuals. (b) Posterior samples of locations of all GW sources found by DNest. Large blue dots and small colorful dots mark positions of input CB-SMBHs and their posterior distributions respectively.

posterior sample, but the difference between the true amplitude and the peak value of the posterior distribution is still within 0.1 dex.

The location of the CB-SMBH candidate can always be determined precisely from electromagnetic observations. However, the orbital period of the candidate cannot be determined directly if it is identified by nonperiodic characteristics such as the velocity-delay map from reverberation mapping (e.g., Wang et al. 2018) or differential phase curves from spectro-astrometry (e.g., Songsheng et al. 2019). Consequently, we should fit the CB-SMBH model to time residuals and electromagnetic data jointly to search for the GW signal. DNest is also suitable for such joint analysis (e.g., Wang et al. 2020).

4. Discussions

4.1. Determining the Number of GW Sources

A major concern when searching for multiple sources simultaneously is to determine the number of GW sources in the model. It can be considered a problem of model selection in the Bayesian framework. Fortunately, as a variant of the nested sampling method, DNest can evaluate the evidence of the model directly (see Equation (16)). Therefore, we can increase the number of GW sources in the model gradually and sample the probability distribution of model parameters with DNest until the evidence reaches its peak value. In such iterations, the information obtained in the present runs can be used in the next run. To be specific, when adding a new source to the model, the ranges of parameters of previously found sources can be further restricted according to their sampled posteriors. In practice, we can set the upper (lower) bound of the parameter to be its mean plus (minus) five times the standard deviation, and the likelihood beyond this range can be ignored. We note that the new prior density of the parameter will be L/L' times higher, where L and L' are the original and modified span of the parameter's value, respectively. As a result, the actual evidence of the model will be V/V' times lower, where V and V' is the original and modified prior volume, respectively. In this way, the

the information obtained with the previous model is being utilized, accelerating the sampling significantly. However, if there are GW sources with very close S/N, the posterior sample of the newly added source in the model may have multiple peaks, corresponding to different sources. Clustering algorithms must be applied to separate them in order to get real means and standard deviations of sources' parameters.

As a preliminary test, we avoid this complexity by selecting nine CB-SMBHs with quite different S/Ns from our mock sample to generate pulsar timing residuals. Then evidence for signal models with different number of sources can be evaluated as we described above. The result is shown in Figure 7. The model evidence grows gradually as the number of sources in the model increases, but the increase rate drops with the decrease of the S/N of the newly added source. We note that the evidence ratio between the 10 source and 9 source model is about 23, which seems to supports the 10 source model strongly (Jeffreys 1998). However, as shown in Figure 7(b), 9 input sources have all been successfully identified, while the 10th source is extremely scattered around the sky, indicating that the data set is best described by the 9 source model. The Bayesian evidence does not penalize the model including the excess source as much as we expected, probably because all parameters except the amplitude ζ of the excess source are unconstrained by the data (Taylor 2021). Therefore, to include a new source in the model, we may require the evidence ratio to be larger than 100 to draw a decisive conclusion.

We also note that a maximum-likelihood-based method has been developed to resolve multiple CB-SMBHs with PTAs (Babak & Sesana 2012; Petiteau et al. 2013). The algorithm ignores the contribution from pulsar terms and consequently discards the pulsar phase parameters. The likelihood function is then marginalized over inclinations, polarization angles, initial phases, and amplitudes of sources and so only depends on locations and frequencies of sources. As the dimension of the parameter space is reduced to $3N_s$, a genetic algorithm can be applied to locate the maximum of likelihood efficiently. The number of sources in the signal model is also increased
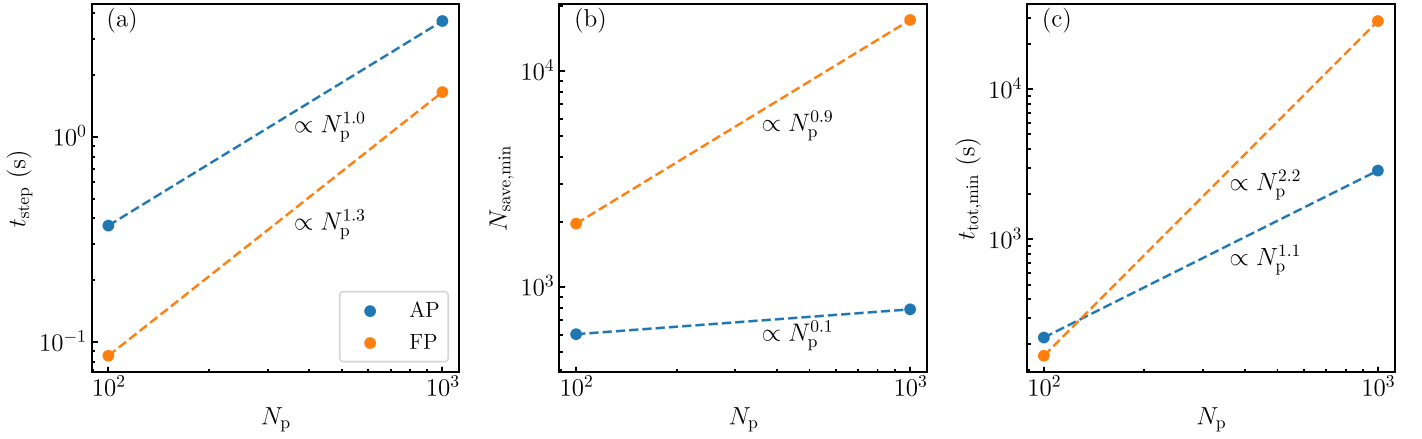
Figure 8. The time consumed by the DNest algorithm as the number of pulsars N_p increases. (a) The time consumed by saving one sample is t_{step}, which is the product of the time to calculate likelihood t_{lh} and the length of the Markov Chain N_{mc} for sampling the restricted prior ($t_{\text{step}} \equiv N_{\text{mc}} t_{\text{lh}}$). For the AP method, it increases linearly with N_p as the time of calculating the likelihood is proportional to N_p. For the FP method, the number of free parameters is proportional to N_p, while the time to calculate the likelihood increases slightly with N_p, leading to a superlinear dependence on N_p. (b) The minimum number of saved samples to build enough levels and locate the region with the maximum of the posterior distribution. This number increases very slowly with N_p for the AP method as uncertainties of the parameters decrease gradually with N_p, but it increases nearly linearly with N_p for the FP method because the dimension of the parameter space is proportional to N_p. (c) The minimum time needed by the algorithm, which is the product of the time consumed by saving one sample and the minimum number of saved samples. It increases linearly and quadratically with N_p for the AP and FP methods, respectively. The algorithm was tested on two chips of Intel Xeon CPU E5-2690 v4 (all of the 28 cores are used).

gradually until the likelihood can no longer be improved significantly. Bécsy & Cornish (2020) also discards pulsar terms in their signal model but treats the number of GW sources as a free parameter. The variable-dimension parameter space is sampled by a transdimensional MCMC sampler `BayesHopper`, and the number of sources is determined by Bayes factors automatically. A possible stochastic GW background can also be integrated into the model to search for standout sources and unresolved background jointly. We may first use these methods with simplified signal models to determine the number of sources and estimate their parameters roughly. Then we can apply the DNest method with full signal models to sample the probability distribution of model parameters accurately.

4.2. Computational Complexity

The time consumed by DNest t_{tot} mainly depends on three factors: (1) the time of calculating likelihood for given parameters t_{lh}, (2) the length of the Markov Chain when sampling the restricted prior distribution N_{mc}, and (3) the total number of particles saved in the process N_{save}. Roughly, we have $t_{\text{tot}} \approx N_{\text{save}} N_{\text{mc}} t_{\text{lh}}$.

First, t_{lh} is proportional to the number of sources in the model N_s and the number of pulsars in the array N_p. Second, N_{mc} is proportional to the number of free parameters in the model, i.e., $N_s(N_p + 7)$. When the number of pulsars is large, the number of free parameters can be approximated as $N_s N_p$. The lower limit of N_{save} depends on the number of levels and the speed of creating levels and is not easy to estimate. During the process of DNest, parameter points are "compressed" to higher levels gradually until they reach those levels that contain most of the posterior probability. We define the ratio of the volume of the prior space to the volume occupied by the posterior sample as the compression rate. Thus, the minimum number of levels equals the logarithm of the compression rate, and so proportional to the dimension of the parameter space at first sight. However, the compression rate contributed by different parameters varies a lot, making the situation more complex. For example, though the

number of pulsar phase parameters surpass that of source parameters largely, the compression rate contributed by them are comparable, because source parameters can be constrained with much smaller uncertainties. Further more, the speed of creating levels may also change. We also note that when we update the phase of a pulsar during sampling with the FP method, only the timing residual of this pulsar needs to be recalculated. Similarly, we can record the timing residuals generated by each source independently and update the residuals of a source only when its associated parameters are being updated. As a result, t_{lh} will not increase a lot when N_p and N_s increase.

When N_p changes, we can conclude that $t_{\text{tot}} \propto N_{\text{save}} N_p$ for the FP, MP, and AP methods. The lower limit of N_{save} increases slowly with N_p for the MP and AP methods because the uncertainties of the source parameters are decreased as the S/N of the source increases with N_p. As for the FP method, the lower limit of N_{save} increases more quickly because the dimension of parameter space also increases with N_p. In order to find out the actual dependence of the total time on N_p, we increase the number of pulsars to 1000 and redo the calculations. The result is shown in Figure 8. For the AP method, we have $t_{\text{step}} \equiv t_{\text{lh}} N_{\text{mc}} \propto N_p$ and $N_{\text{save, min}} \propto N_p^{0.1}$, leading to $t_{\text{tot, min}} \propto N_p^{1.1}$. For the FP method, we have $t_{\text{step}} \propto N_p^{1.3}$ and $N_{\text{save, min}} \propto N_p^{0.9}$, resulting in $t_{\text{tot, min}} \propto N_p^{2.2}$.

When N_s changes, we have $t_{\text{tot}} \propto N_{\text{save}} N_s$ for the FP method. The lower limit of N_{save} also increases with N_s as the dimension of the parameter space is proportional to N_s. Therefore, the computational complexity is approximately $\mathcal{O}(N_s^2)$. The computation is affordable as long as N_s is about 10–20. For a larger N_s, we can conduct a hierarchical search as in the previous subsection. If the ranges of parameters of previously found sources can be restricted, the computational complexity for the model with N_s sources is closer to $\mathcal{O}(N_s)$ rather than $\mathcal{O}(N_s^2)$, because the information from previous runs expedites the level building.

5. Conclusion

In this paper, we conduct a mock data analysis to test the performance of DNest method in searching for continuous GW

signals in pulsar timing residuals. For a PTA containing 10^2 pulsars, a Bayesian framework with DNest can overcome the issues of high dimension caused by pulsar phases and performs as well as that with the marginalization technique in light of the accuracy, robustness, and efficiency in search of single sources. The probability distribution of model parameters can still be sampled effectively even if the number of pulsars increases to 10^3. The method can also be used to search for multiple sources simultaneously. Several strongest sources across a wide range of locations and frequencies can be successfully identified with the method. The method does not depend on the analytical form of the signal model and therefore can be generalized to CB-SMBHs with evolving and elliptical orbits or even more complicated models.

Appendix A
Pulsar Timing Residuals Induced by GWs Generated by CB-SMBH

We first derive GWs emitted by a binary black hole, following the approach of Wahlquist (1987). We establish an inertial frame centered at the observer, with the z-axis pointing to the north celestial pole and the x-axis to the equinox, as shown in Figure 9(a). For a CB-SMBH with right accession α, decl. δ, and distance D, the unit vector pointing from the center of mass of the CB-SMBH to the observer is

$$\hat{\boldsymbol{k}} = -\cos\delta\cos\alpha\hat{\boldsymbol{x}} - \cos\delta\sin\alpha\hat{\boldsymbol{y}} - \sin\delta\hat{\boldsymbol{z}}. \quad (A1)$$

We choose

$$\hat{\boldsymbol{E}} = -\sin\alpha\hat{\boldsymbol{x}} + \cos\alpha\hat{\boldsymbol{y}} \text{ and}$$
$$\hat{\boldsymbol{N}} = -\sin\delta\cos\alpha\hat{\boldsymbol{x}} - \sin\delta\sin\alpha\hat{\boldsymbol{y}} + \cos\delta\hat{\boldsymbol{z}} \quad (A2)$$

as two orthogonal normalized basic vectors in the plane perpendicular to $\hat{\boldsymbol{k}}$. $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{N}}$ point to the direction of increasing R.A. and decl., respectively.

Next, we establish an inertial frame centered at the center of mass of the CB-SMBH, as shown in Figure 9(b). We assume that the intersection of the orbital plane and the celestial sphere is

$$\hat{\boldsymbol{u}} = \cos\psi\hat{\boldsymbol{E}} + \sin\psi\hat{\boldsymbol{N}}, \quad (A3)$$

and the angle between them is ι. So, the orthogonal to the line of nodes in the orbital plane is

$$\hat{\boldsymbol{v}} = \cos\iota(\sin\psi\hat{\boldsymbol{E}} - \cos\psi\hat{\boldsymbol{N}}) + \sin\iota\hat{\boldsymbol{k}}. \quad (A4)$$

Note that $(\hat{\boldsymbol{u}} \times \hat{\boldsymbol{v}}) \cdot \hat{\boldsymbol{k}} = \cos\iota$.

Assuming a nonevolving circular orbit for simplicity, displacements of the primary and secondary black holes in the Newtonian limit are

$$\boldsymbol{r}_1 = -D\hat{\boldsymbol{k}} + \frac{m_2}{m_1 + m_2}\boldsymbol{r}, \boldsymbol{r}_2 = -D\hat{\boldsymbol{k}} - \frac{m_1}{m_1 + m_2}\boldsymbol{r}, \quad (A5)$$

where $\boldsymbol{r} = a[\cos\varphi\hat{\boldsymbol{u}} + \sin\varphi\hat{\boldsymbol{v}}]$ is the displacement of the primary black hole relative to the secondary, $\varphi = \omega_r t + \varphi_0$ is the orbital phase, m_1 and m_2 are mass of the two black holes, a is their separation, and $\omega_r = [G(m_1 + m_2)]^{1/2}a^{-3/2}$ is the angular velocity of the binary in the rest frame. The quadrupole moment for the system is therefore

$$\boldsymbol{Q} = m_1\boldsymbol{r}_1 \otimes \boldsymbol{r}_1 + m_2\boldsymbol{r}_2 \otimes \boldsymbol{r}_2$$
$$= MD^2\hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}} + \mu\boldsymbol{r} \otimes \boldsymbol{r}, \quad (A6)$$

where $M = m_1 + m_2$ and $\mu = m_1 m_2/(m_1 + m_2)$ are the total mass and reduced mass, respectively.

Considering the GW generated by the binary and propagating in the direction $\hat{\boldsymbol{k}}$, we may use the projection tensor $\boldsymbol{I} - \hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}}$ to obtain the transverse components of $\boldsymbol{Q}$, where $\boldsymbol{I}$ is the identity tensor:

$$\boldsymbol{Q}^T \equiv (\boldsymbol{I} - \hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}}) \cdot \boldsymbol{Q} \cdot (\boldsymbol{I} - \hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}})$$
$$= \mu\boldsymbol{r}^T \otimes \boldsymbol{r}^T, \quad (A7)$$

where

$$\boldsymbol{r}^T = (\boldsymbol{I} - \hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}}) \cdot \boldsymbol{r} = a(\cos\varphi\cos\psi + \cos\iota\sin\varphi\sin\psi)\hat{\boldsymbol{E}}$$
$$+ a(\cos\varphi\sin\psi - \cos\iota\sin\varphi\cos\psi)\hat{\boldsymbol{N}}. \quad (A8)$$

Subtracting the trace then gives $\boldsymbol{Q}^{\mathrm{TT}}$ as

$$\boldsymbol{Q}^{\mathrm{TT}} = \mu\left[\boldsymbol{r}^T \otimes \boldsymbol{r}^T - \frac{1}{2}(\boldsymbol{r}^T \cdot \boldsymbol{r}^T)(\boldsymbol{I} - \hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}})\right]$$
$$= Q_+\boldsymbol{e}_+ + Q_\times\boldsymbol{e}_\times, \quad (A9)$$

where

$$Q_+ = \frac{1}{2}\mu a^2[\cos 2\psi(\cos^2\varphi - \cos^2\iota\sin^2\varphi)$$
$$+ \sin 2\psi\cos\iota\sin 2\varphi], \quad (A10)$$

$$Q_\times = \frac{1}{2}\mu a^2[\sin 2\psi(\cos^2\varphi - \cos^2\iota\sin^2\varphi)$$
$$- \cos 2\psi\cos\iota\sin 2\varphi], \quad (A11)$$

and

$$\boldsymbol{e}_+ = \hat{\boldsymbol{E}} \otimes \hat{\boldsymbol{E}} - \hat{\boldsymbol{N}} \otimes \hat{\boldsymbol{N}}, \boldsymbol{e}_\times = \hat{\boldsymbol{E}} \otimes \hat{\boldsymbol{N}} + \hat{\boldsymbol{N}} \otimes \hat{\boldsymbol{E}} \quad (A12)$$

The GW in transverse traceless gauge (TT gauge) is therefore (Misner et al. 1973)

$$\boldsymbol{h}^{\mathrm{TT}}(t, \boldsymbol{x}) = \frac{2G}{c^4 D}\ddot{\boldsymbol{Q}}^{\mathrm{TT}}|_{t'} = h_+\boldsymbol{e}_+ + h_\times\boldsymbol{e}_\times, \quad (A13)$$

with

$$h_+ = \frac{2(G\mathcal{M}_r)^{5/3}\omega_r^{2/3}}{c^4 D}[-\cos 2\psi(1 + \cos^2\iota)\cos 2\varphi$$
$$- 2\sin 2\psi\cos\iota\sin 2\varphi],$$

$$h_\times = \frac{2(G\mathcal{M}_r)^{5/3}\omega_r^{2/3}}{c^4 D}[-\sin 2\psi(1 + \cos^2\iota)\cos 2\varphi$$
$$+ 2\cos 2\psi\cos\iota\sin 2\varphi]. \quad (A14)$$
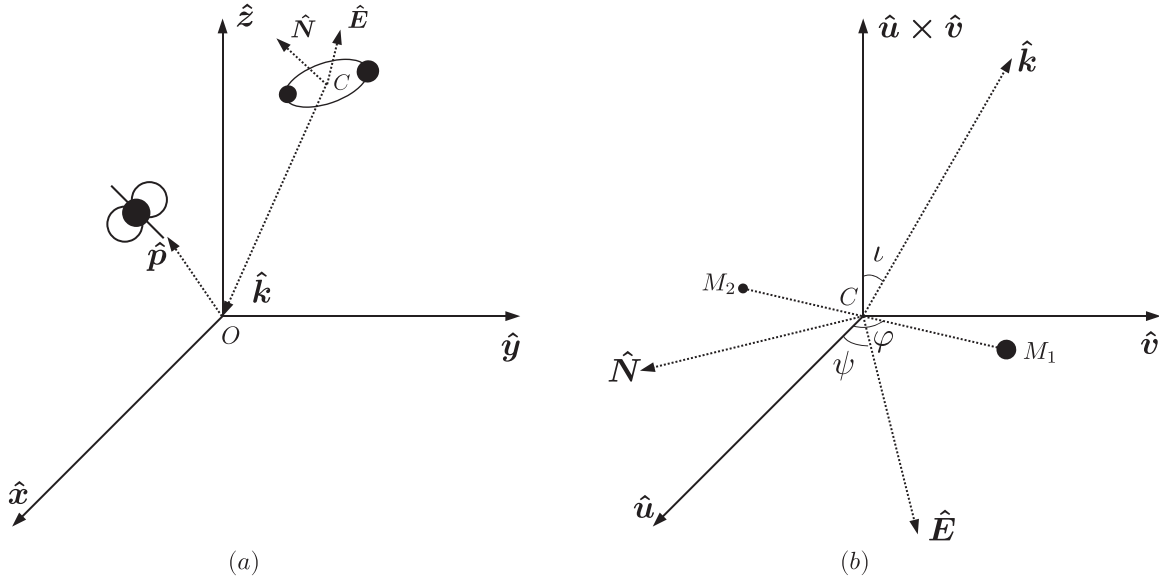
(a) (b)

Figure 9. Coordinate systems for the CB-SMBH and pulsar. (a) An inertial frame centered at the observer O. The z-axis points to the north celestial pole and the x-axis to the equinox. C is the location of the center of mass of the CB-SMBH, and $\hat{k}$ is the unit vector pointing from C to O. $\hat{E}$ and $\hat{N}$ are the basis vectors of the R.A. and decl. of the equatorial coordinates. Unit vectors $\hat{k}$, $\hat{E}$, and $\hat{N}$ are perpendicular with each other. $\hat{p}$ is the unit vector pointing from the observer to the pulsar. (b) An inertial frame centered at the center of mass of the CB-SMBH C. M_1 and M_2 are the primary and secondary black holes, respectively. The unit vector $\hat{u}$ lies in the orbital plane and perpendicular to $\hat{k}$. The unit vector $\hat{v}$ also lies in the orbital plane and perpendicular to $\hat{u}$. The inclination angle ι is defined as the angle between $\hat{k}$ and $\hat{u} \times \hat{v}$. The polarization angle ψ is defined as the angle between $\hat{u}$ and $\hat{E}$. The phase angle ϕ is defined as the angle between $\hat{u}$ and CM_1.

Here, $\mathcal{M}_r \equiv (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ is the rest-frame chirp mass and $t' = t - |x + \hat{k}D|/c$ is the retarded time.

If we take the expansion of the universe into consideration and assume that the redshift of the CB-SMBH is z, Equation (A14) should be revised to

$$
\begin{aligned}
h_+ &= \frac{2(G\mathcal{M})^{5/3}\omega^{2/3}}{c^4 D_L}[-\cos 2\psi(1 + \cos^2 \iota)\cos 2\varphi \\
&\quad - 2\sin 2\psi \cos \iota \sin 2\varphi], \\
h_\times &= \frac{2(G\mathcal{M})^{5/3}\omega^{2/3}}{c^4 D_L}[-\sin 2\psi(1 + \cos^2 \iota)\cos 2\varphi \\
&\quad + 2\cos 2\psi \cos \iota \sin 2\varphi],
\end{aligned}
$$
$$(A15)$$

where D_L is the luminosity distance of the CB-SMBH, $\mathcal{M} = (1 + z)\mathcal{M}_r$ is the redshifted chirp mass, $\omega = (1 + z)^{-1}\omega_r$ is the observed orbital angular frequency, and $\varphi = \varphi_0 + \omega t$ is the observed orbital phase (Holz & Hughes 2005).

If we further consider the evolution of the binary orbit due to the radiation of GWs, the merger time in the observer's frame is

$$
\begin{aligned}
t_{\mathrm{merge}} &= \frac{5c^5}{256}(G\mathcal{M})^{-5/3}\omega^{-8/3} \\
&= 4.4 \times 10^4 \left(\frac{M}{10^9 M_\odot}\right)^{-5/3}\left(\frac{\omega}{10^{-8}\mathrm{Hz}}\right)^{-8/3}\mathrm{yr}.
\end{aligned}
$$
$$(A16)$$

On such a timescale, the angular velocity of the binary is no longer constant but varies with time as

$$
\omega(t) = \omega_0\left(1 - \frac{t}{t_{\mathrm{merge}}}\right)^{-3/8},
$$
$$(A17)$$

where ω_0 is the observed angular frequency at the initial time. The evolution of the orbital phase will be modified to

$$
\begin{aligned}
\varphi(t) &= \varphi_0 + \int_0^t \omega(t')dt' \\
&= \varphi_0 + \frac{8\omega_0 t_{\mathrm{merge}}}{5}\left[1 - \left(1 - \frac{t}{t_{\mathrm{merge}}}\right)^{5/8}\right],
\end{aligned}
$$
$$(A18)$$

which reduces to $\varphi(t) = \varphi_0 + \omega_0 t$ when $|t| \ll t_{\mathrm{merge}}$.

To derive the impact of GWs on the propagation of radio pulses from pulsars, we follow the approach in Estabrook & Wahlquist (1975). As the wavelength of the GW is much shorter than the distance between the observer and pulsars in the Galaxy, which is further shorter than the distance to the CB-SMBH, the GW can be well approximated by a plane wave propagating in the direction $\hat{k}$. So the spacetime possesses three Killing vectors:

$$
\xi_a = (0, \hat{E}), \; \xi_b = (0, \hat{N}) \text{ and } \xi_c = (1, \hat{k}).
$$
$$(A19)$$

Now, suppose there is a pulsar in direction $\hat{p}$, emitting a photon toward us. The frequency of the photon is ν in the rest frame of the pulsar. Thus, the four-wavevector of the photon to $\mathcal{O}(h)$ is

$$
k^\mu = \nu\left(1, -\left(\delta_{ij} - \frac{1}{2}h_{ij}\right)\hat{p}_j\right),
$$
$$(A20)$$

or as a covariant vector

$$
k_\mu = -\nu\left(1, \left(\delta_{ij} + \frac{1}{2}h_{ij}\right)\hat{p}_j\right).
$$
$$(A21)$$

Due to the properties of Killing vectors and geodesics, the quantities $k_\mu \xi^\mu$ must be conserved during the propagation of the photon. When the photon reaches us, the frequency has

shifted to $\nu + \delta\nu$ and direction to $\hat{\boldsymbol{p}} + \delta\hat{\boldsymbol{p}}$. We have

$$
\begin{aligned}
&\nu\left(\hat{p}_i\hat{E}_i + \frac{1}{2}h_{ij}(x)\hat{E}_i\hat{p}_j\right) \\
&= (\nu + \delta\nu)\left(\left(\hat{p}_i + \delta\hat{p}_i\right)\hat{E}_i + \frac{1}{2}h_{ij}(y)\hat{E}_i(\hat{p}_j + \delta\hat{p}_j)\right) \\
&\nu\left(\hat{p}_i\hat{N}_i + \frac{1}{2}h_{ij}(x)\hat{N}_i\hat{p}_j\right) \\
&= (\nu + \delta\nu)\left(\left(\hat{p}_i + \delta\hat{p}_i\right)\hat{N}_i + \frac{1}{2}h_{ij}(y)\hat{N}_i(\hat{p}_j + \delta\hat{p}_j)\right) \\
&\nu(1 + \hat{k}_i\hat{p}_i) = (\nu + \delta\nu)(1 + \hat{k}_i(\hat{p}_i + \delta\hat{p}_i)) \\
&(\hat{p}_i + \delta\hat{p}_i)(\hat{p}_j + \delta\hat{p}_j) = 1,
\end{aligned}
\tag{A22}
$$

where x and y are the spacetime coordinates of when the photon is emitted and received, respectively.

To the first order of h, the solution to Equation (A22) gives

$$
\frac{\delta\nu}{\nu} = F^+\Delta h_+ + F^\times\Delta h_\times,
\tag{A23}
$$

where

$$
\begin{aligned}
F^+ &= \frac{1}{2}\frac{(\hat{\boldsymbol{N}} \cdot \hat{\boldsymbol{p}})^2 - (\hat{\boldsymbol{E}} \cdot \hat{\boldsymbol{p}})^2}{1 + \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{k}}}, \\
F^\times &= -\frac{(\hat{\boldsymbol{E}} \cdot \hat{\boldsymbol{p}})(\hat{\boldsymbol{N}} \cdot \hat{\boldsymbol{p}})}{1 + \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{k}}},
\end{aligned}
\tag{A24}
$$

and

$$
\begin{aligned}
\Delta h_{+,\times} &= h_{+,\times}(t, 0) - h_{+,\times}(t - d_p/c, d_p\hat{\boldsymbol{p}}) \\
&= h_{+,\times}(t, 0) - h_{+,\times}(t - d_p(1 + \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{k}})/c, 0).
\end{aligned}
\tag{A25}
$$

Here, d_p is the distance of the pulsar. We also define the pulsar lag $\tau_p \equiv d_p(1 + \hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{k}})/c$ and pulsar time $t_p \equiv t - \tau_p$, respectively.

The timing residual induced by the GW for an observer at Earth is

$$
s(t) = -\int\frac{\delta\nu}{\nu}dt = F^+\Delta s_+ + F^\times\Delta s_\times,
\tag{A26}
$$

where

$$
\begin{aligned}
\Delta s_{+,\times} &= s_{+,\times}(t) - s_{+,\times}(t_p), \\
s_+(t) &= \frac{(G\mathcal{M})^{5/3}}{c^4 D_L \omega(t)^{1/3}}[\cos 2\psi(1 + \cos^2\iota)\sin 2\varphi(t) \\
&\quad - 2\sin 2\psi\cos\iota\cos 2\varphi(t)], \\
s_\times(t) &= \frac{(G\mathcal{M})^{5/3}}{c^4 D_L \omega(t)^{1/3}}[\sin 2\psi(1 + \cos^2\iota)\sin 2\varphi(t) \\
&\quad + 2\cos 2\psi\cos\iota\cos 2\varphi(t)],
\end{aligned}
\tag{A27}
$$

For the pulsar term $s_{+,\times}(t_p)$, we also note that

$$
\begin{aligned}
\omega(t_p) &= \omega_p\left(1 - \frac{t}{t_{\text{merge}} + \tau_p}\right)^{-3/8}, \\
\varphi(t_p) &= \varphi_p + \frac{8\omega_p(t_{\text{merge}} + \tau_p)}{5} \\
&\quad \times\left[1 - \left(1 - \frac{t}{t_{\text{merge}} + \tau_p}\right)^{5/8}\right].
\end{aligned}
\tag{A28}
$$

where $\omega_p \equiv \omega(-\tau_p)$ and $\varphi_p \equiv \varphi(-\tau_p)$.

Appendix B
Evolution of Binary Orbits

We have ignored the evolution of binary orbits in our mock data analysis. However, the detectable CB-SMBH usually has a chirp mass larger than $10^9 M_\odot$ and an orbital period less than 10 yr, and thereby its merger time will be comparable to light-travel times between the Earth and pulsars, making frequencies of the Earth term and pulsar term different. Fortunately, our method can be generalized to the case of evolving orbits directly by further including merger times of sources and distances of pulsars in the model. For simplicity, we assume that relative uncertainties of pulsar distances given by electromagnetic observations are 20% and using Gaussian priors for them. The total dimension of the parameter space will be $N_s(8 + N_p) + N_p$.

We note that merger times of a fair fraction of CB-SMBHs generated by the method in Section 2.2 are quite short ($\lesssim$ 10 yr), and so they are hardly caught by a typical PTA program. Therefore, we draw the amplitude of timing residuals ζ and merger time t_{merge} rather than chirp mass and luminosity distance log-uniformly when generating mock populations of CB-SMBHs. The ranges of ζ and t_{merge} are $[10^{-15}, 10^{-7}]$ s and $[10^2, 10^8]$ yr, respectively. Then we perform a blind search for a single target via the DNest method, as done in Section 3.1. The result is shown in Figure 10.

The source with the highest S/N ($\rho \sim 120$) is successfully identified. Figure 10(a) shows the distribution of R.A. and decl. of the source. The actual location of the source lies near the 2σ region of the posterior sample. The 2σ region of the posterior sample has a size of $\sim 3\,\text{deg}^2$, and the angular distance between the source and distribution center is about $0°.7$. The second panel shows the distribution of the amplitude of timing residuals and the inclination of the orbital plane. The two parameters are still highly correlated, and the uncertainty of the amplitude is about 0.1 dex. The third panel displays the distribution of the angular frequency of the GW and merger time of the CB-SMBH, which are weakly correlated. The frequency can still be constrained precisely. Though the merger time is much longer than the time span of the PTA program, it can be determined quite accurately by detecting the frequency difference between the Earth term and pulsar terms. This shows the potential of CB-SMBHs with $t_{\text{merge}} \lesssim 10^4$ yr to be standard sirens for cosmology if their electromagnetic counterparts can be identified. In a word, DNest also performs well when the evolution of binary orbits is taken into account.
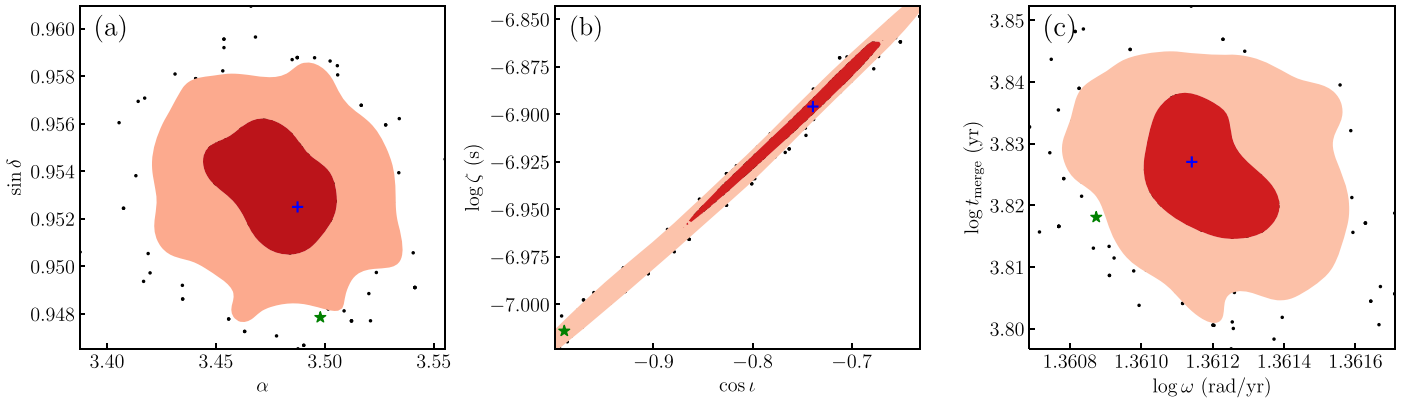
Figure 10. Posterior probability distributions of some model parameters of the GW source with an evolving orbit. (a) Distributions of the right accession and decl. of the source. The red and pink colors represent the 1σ and 2σ regions of the distribution. Black dots are posterior samples outside the 2σ region. The green star marks the input parameters, while the blue cross indicates the peak of the distribution. (b) Distributions of the inclination and amplitude of the source. (c) Distributions of the angular frequency and merger time of the source.

ORCID iDs

Yu-Yang Songsheng ⓘ https://orcid.org/0000-0003-4042-7191
Yan-Rong Li ⓘ https://orcid.org/0000-0001-5841-9179
Pu Du ⓘ https://orcid.org/0000-0002-5830-3544
Yan Wang ⓘ https://orcid.org/0000-0001-8990-5700
Soumya D. Mohanty ⓘ https://orcid.org/0000-0002-4651-6438

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, PhRvL, 121, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, PhRvX, 9, 031040
Abbott, R., Abbott, T. D., Abraham, S., et al. 2021, ApJL, 913, L7
Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2019, ApJ, 880, 116
Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2021, ApJS, 252, 4
Amaro-Seoane, P., Audley, H., Babak, S., et al. 2017, arXiv:1702.00786
Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020a, ApJL, 905, L34
Arzoumanian, Z., Baker, P. T., Brazier, A., et al. 2020b, ApJ, 900, 102
Babak, S., Petiteau, A., Sesana, A., et al. 2016, MNRAS, 455, 1665
Babak, S., & Sesana, A. 2012, PhRvD, 85, 044034
Bécsy, B., & Cornish, N. J. 2020, CQGra, 37, 135011
Bon, E., Jovanović, P., Marziani, P., et al. 2012, ApJ, 759, 118
Brewer, B. J., & Foreman-Mackey, D. 2018, J. Stat. Softw., 86, 1
Brewer, B. J., Pártay, L. B., & Csányi, G. 2011, Stat. Comput., 21, 649
Charisi, M., Bartos, I., Haiman, Z., et al. 2016, MNRAS, 463, 2145
Corbin, V., & Cornish, N. J. 2010, arXiv:1008.1782
Desvignes, G., Caballero, R. N., Lentati, L., et al. 2016, MNRAS, 458, 3341
Eberhart, R., & Kennedy, J. 1995, in MHS'95. Proc. of the 6th Int. Symp. on
    Micro Machine and Human Science (Piscataway, NJ: IEEE), 39
Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, MNRAS, 372, 1549
Ellis, J. A. 2013, CQGra, 30, 224004
Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012, ApJ, 756, 175
Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, ApJS, 201, 23
Estabrook, F. B., & Wahlquist, H. D. 1975, GReGr, 6, 439
Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300
Graham, M. J., Djorgovski, S. G., Stern, D., et al. 2015, MNRAS, 453, 1562
Green, P. J. 1995, Bimka, 82, 711
Hastings, W. K. 1970, Bimka, 57, 97
Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39
Holland, J. H. 1975, Adaptation in Natural and Artificial Systems. An
    Introductory Analysis with Applications to Biology, Control and Artificial
    Intelligence (Ann Arbor, MI: Univ. of Michigan Press)
Holz, D. E., & Hughes, S. A. 2005, ApJ, 629, 15
Hu, W.-R., & Wu, Y.-L. 2017, Natl. Sci. Rev., 4, 685
Jeffreys, H. 1998, The Theory of Probability, 432 (Oxford: Clarendon)
Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, ApJ, 653, 1571
Kerr, M., Reardon, D. J., Hobbs, G., et al. 2020, PASA, 37, e020

Lee, K. J. 2016, in ASP Conf. Ser. 502, Frontiers in Radio Astronomy and
    FAST Early Sciences Symp. 2015, ed. L. Qain & D. Li (San Francisco, CA:
    ASP), 19
Li, Y.-R. 2020, LiyrAstroph/CDNest: CDNest: A diffusive nested sampling
    code in C, v0.2.0, Zenodo, doi:10.5281/zenodo.3884449
Li, Y.-R., Songsheng, Y.-Y., Qiu, J., et al. 2018, ApJ, 869, 137
Li, Y.-R., Wang, J.-M., Ho, L. C., et al. 2016, ApJ, 822, 4
Li, Y.-R., Wang, J.-M., Zhang, Z.-X., et al. 2019, ApJS, 241, 33
Lloyd, S. 1982, ITIT, 28, 129
Luo, J., Chen, L.-S., Duan, H.-Z., et al. 2016, CQGra, 33, 035010
Luo, J., Ransom, S., Demorest, P., et al. 2021, ApJ, 911, 45
Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T. M., & Moriya, T. J.
    2016, A&A, 588, A50
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., &
    Teller, E. 1953, JChPh, 21, 1087
Misner, C. W., Thorne, K. S., & Wheeler, J. A. 1973, Gravitation, Vol. 999
    (San Francisco, CA: W. H. Freeman and Company)
Pancoast, A., Brewer, B. J., Treu, T., et al. 2014, MNRAS, 445, 3073
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res.,
    12, 2825
Petiteau, A., Babak, S., Sesana, A., & de Araújo, M. 2013, PhRvD, 87, 064036
Sazhin, M. V. 1978, SvA, 22, 36
Sharma, S. 2017, ARA&A, 55, 213
Skilling, J. 2004, in AIP Conf. Ser. 735, Bayesian Inference and Maximum
    Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, &
    U. V. Toussaint (Melville, NY: AIP), 395
Smits, R., Kramer, M., Stappers, B., et al. 2009, A&A, 493, 1161
Songsheng, Y.-Y. 2021, TRAINS: Timing Residuals Analysis Integrated with
    Nested Sampling, v0.1.1, Zenodo, doi: 10.5281/zenodo.5230415
Songsheng, Y.-Y., Wang, J.-M., Li, Y.-R., & Du, P. 2019, ApJ, 881, 140
Swendsen, R. H., & Wang, J.-S. 1986, PhRvL, 57, 2607
Taylor, S., Ellis, J., & Gair, J. 2014, PhRvD, 90, 104028
Taylor, S. R. 2021, arXiv:2105.13270
Tiburzi, C., Hobbs, G., Kerr, M., et al. 2016, MNRAS, 455, 4339
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wahlquist, H. 1987, GReGr, 19, 1101
Wang, J.-M., & Li, Y.-R. 2020, RAA, 20, 160
Wang, J.-M., Songsheng, Y.-Y., Li, Y.-R., Du, P., & Zhang, Z.-X. 2020,
    NatAs, 4, 517
Wang, J.-M., Songsheng, Y.-Y., Li, Y.-R., & Yu, Z. 2018, ApJ, 862, 171
Wang, Y., & Mohanty, S. D. 2010, PhRvD, 81, 063002
Wang, Y., & Mohanty, S. D. 2017, PhRvL, 118, 151104
Wang, Y., & Mohanty, S. D. 2018, JPhCS, 957, 012003
Wang, Y., Mohanty, S. D., & Cao, Z. 2021, ApJL, 907, L43
Wang, Y., Mohanty, S. D., & Jenet, F. A. 2015, ApJ, 815, 125
Wang, Y., Mohanty, S. D., & Qian, Y.-Q. 2017, JPhCS, 840, 012058
Weltman, A., Bull, P., Camera, S., et al. 2020, PASA, 37, e002
Xin, C., Mingarelli, C. M. F., & Hazboun, J. S. 2021, ApJ, 915, 97
Zhu, X. J., Hobbs, G., Wen, L., et al. 2014, MNRAS, 444, 3709
Zhu, X. J., Wen, L., Xiong, J., et al. 2016, MNRAS, 461, 1317