# Searching for gravitational wave burst in pulsar-timing-array data with piecewise linear functions

Heling Deng, Bence Bécsy, and Xavier Siemens
*Department of Physics, Oregon State University, Corvallis, Oregon 97331, USA*

Neil J. Cornish
*eXtreme Gravity Institute, Department of Physics, Montana State University, Bozeman, Montana 59717, USA*

Dustin R. Madison
*Department of Physics, University of the Pacific, 3601 Pacific Avenue, Stockton, California 95211, USA*

(Received 19 July 2023; accepted 27 September 2023; published 27 November 2023)

Transient gravitational waves (aka gravitational wave bursts) within the nanohertz frequency band could be generated by a variety of astrophysical phenomena such as the encounter of supermassive black holes, the kinks or cusps in cosmic strings, or other as-yet-unknown physical processes. Radio pulses emitted from millisecond pulsars could be perturbed by passing gravitational waves; hence, the correlation of the perturbations in a pulsar timing array can be used to detect and characterize burst signals with a duration of $\mathcal{O}(1\text{–}10)$ years. We propose a fully Bayesian framework for the analysis of the pulsar-timing-array data, where the burst waveform is generically modeled by piecewise straight lines, and the waveform parameters in the likelihood can be integrated out analytically. As a result, with merely three parameters (in addition to those describing the pulsars' intrinsic and background noise), one is able to efficiently search for the existence and the sky location of a burst signal. If a signal is present, the posterior of the waveform can be found without further Bayesian inference. We demonstrate this model by analyzing simulated datasets containing a stochastic gravitational wave background and a burst signal generated by the parabolic encounter of two supermassive black holes.

DOI: 10.1103/PhysRevD.108.102007

## I. INTRODUCTION

Millisecond pulsars are highly magnetized rotating neutron stars with periods $\mathcal{O}(1\text{–}10)$ milliseconds. Beams of electromagnetic radiation emitting from the magnetic poles rotate about the star's spinning axis and may hit us as radio pulses once per period. Known to have very stable rotations, millisecond pulsars are highly sensitive probes of their environments, including gravitational waves (GWs). These waves cannot be inferred by observing a single pulsar, but correlations among an array of pulsars can in principle be hunted down. In particular, the detection of Hellings and Downs (HD) correlations would be an unambiguous signature of a stochastic gravitational wave background (SGWB) [1]. Pulsar-timing-array (PTA) observations so far typically have a sampling interval of weeks and span over $\mathcal{O}(10)$ years, implying a sensitive GW frequency range of around 1–100 nHz.

Recently, the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [2] reported compelling evidence for a nHz SGWB in our Universe [3]. The analysis of the NANOGrav 15-year dataset shows a red noise process that has a spectrum common among all pulsars and that is spatially correlated among pulsar pairs in a manner consistent with HD correlations. In the meantime, the Chinese Pulsar Timing Array (CPTA), the European Pulsar Timing Array (EPTA) [including data from the Indian PTA (InPTA)], and the Parkes Pulsar Timing Array (PPTA) have also reported these correlations at varying levels of significance [4–6].

A promising source of SGWB in PTA is the combined emission from an ensemble of inspiraling supermassive black hole (SMBH) binaries. Most galaxies have an SMBH with mass $10^6$–$10^{10}M_\odot$ sitting at the center [7]. When two galaxies merge, two SMBHs may find each other and form a binary, emitting gravitational radiation for a time span much larger than the PTA observation period. The investigation of constraints on SMBH binaries from the NANOGrav 15-year dataset can be found in Ref. [8]. An SGWB in the nHz band could also be generated by physics in the early Universe, such as inflation [9–12], phase transitions [13–23], and cosmic strings [24–29]. Up-to-date PTA constraints on new physics can be found in Refs. [30–32].

Besides the stochastic background, events with strong GW emissions from certain sky locations could be detected

 

individually. An example under active search is continuous waves emitted from the brightest SMBH binary, the detection of which would provide direct evidence of the existence of SMBH binaries [33–36]. Several searches have been carried out over the years, setting increasingly stringent upper bounds on these sources [37–45].

In this paper, we shall focus on searching for the strongest GW burst (or GW transient) with a duration comparable to the PTA observation period. Bursts of this kind could come from a variety of physical processes: the encounter of two SMBHs, cosmic string cusps or kinks [46–48], or the GW memory effect—i.e., a permanent deformation of spacetime after a violent event, such as the merger of two SMBHs (the merger itself also emits GWs, but they are not in the nHz band) [49–55]. Furthermore, there may exist other phenomena generating bursts with unknown waveforms.

A number of methods of searching for and reconstructing a generic burst signal in the PTA data have been proposed and developed over the years.[1] In Ref. [59], an analytical hybrid (frequentist-Bayesian) approach was adopted, where each data point was assigned a parameter characterizing the burst waveform, and a maximum *a posteriori* estimate was performed to fix the hyperparameters. This method was improved in Ref. [60] by a Bayesian nonparametric analysis. In Refs. [39,61,62], frequentist frameworks were developed (in the time domain or the frequency domain), where piecewise linear functions were introduced to describe the burst waveform, and a least-squares fitting process was performed to obtain the estimates of the waveform parameters. Recently, a Bayesian algorithm was implemented in Refs. [63,64], where a generic burst is modeled by the superposition of Morlet-Gabor wavelets.

In the present work, we present an efficient Bayesian method to search for the strongest signal from a burst with an unknown waveform in PTA data. Motivated by Ref. [62], we model the burst waveform (with two polarizations) with two piecewise linear functions in the time domain. In our model, the existence and the sky location of the burst can be determined with only three parameters (in addition to those describing the pulsars' intrinsic and common noise), with the parameters characterizing the burst waveform analytically integrated out. If a signal is indeed present in the data, its waveform can then be straightforwardly extracted without performing further Bayesian inference. We shall test this method by analyzing simulated PTA datasets that contain an SGWB.

The rest of the paper is organized as follows: In Sec. II, we describe how the burst waveform can be modeled by piecewise linear functions, which leads to a simple expression of the marginalized likelihood discussed in detail in Sec. III. In Sec. III, we also show how to detect the burst's sky location and reconstruct the waveform. The model will be demonstrated by the analyses of three simulated datasets in Sec. IV. Conclusions are summarized and discussed in Sec. V.

[1]Burst search around the kilohertz frequency band has been studied extensively in the context of ground-based interferometric GW detectors such as LIGO and Virgo [56–58].

## II. WAVEFORM MODELING

In pulsar timing, the times of arrival (TOAs) of radio pulses from a millisecond pulsar are measured and compared with predictions based on a timing model that describes the pulsar physics (e.g., the spin period, the spin-down rate, etc.). The differences are called timing residuals. While they mainly come from deviations in the timing model, white noise from measurement uncertainties and red noise from the pulsar's intrinsic instabilities, the residuals may also be disturbed by passing GWs, such as a possible stochastic background or deterministic signals from certain physical phenomena. In this section, we describe the residuals of a pulsar induced by a generic GW burst and introduce our model, where the waveform of the burst is approximated by piecewise straight lines. The residuals can then be expressed in a simple form.

### A. Timing residuals induced by burst

The location of the observer is set as the Solar System barycenter (SSB), which sits at the origin of Cartesian coordinates defined by orthonormal vectors $(\hat{x}, \hat{y}, \hat{z})$. The north celestial pole is in the $\hat{z}$ direction, and the vernal equinox is in the $\hat{x}$ direction. The sky location of the GW burst can then be determined by the polar and azimuthal angles $(\theta, \phi)$. Three useful orthonormal vectors are

$$\hat{\Omega} = -\sin\theta\cos\phi\hat{x} - \sin\theta\sin\phi\hat{y} - \cos\theta\hat{z}, \tag{1}$$

$$\hat{m} = -\sin\phi\hat{x} + \cos\phi\hat{y}, \tag{2}$$

$$\hat{n} = -\cos\theta\cos\phi\hat{x} - \cos\theta\sin\phi\hat{y} + \sin\theta\hat{z}, \tag{3}$$

where $\hat{\Omega}$ points from the GW source to the SSB, and $\hat{m}$ and $\hat{n}$ are vectors that are useful for describing the two polarization tensors of the source.

It can be shown that the GW brings two redshifting signatures to a pulsar's TOAs: perturbations to the timing residuals when the wave reaches the pulsar (leading to the "pulsar term"), and when it reaches the Earth (the "Earth term") [65]. Since the duration of a GW burst of interest (∼1–10 years) is much smaller than the time it takes for a radio pulse to travel from the pulsar to the Earth (thousands of years), it is unlikely that a pulsar's Earth term and pulsar term are both present in a PTA dataset; it is also unlikely that pulsar terms from different pulsars are correlated. On the other hand, all Earth terms show up within the same period. It is thus safe to neglect the pulsar terms.

The residuals of a single pulsar induced by a burst can then be written in the following form:

$$h(t) = F^{+}(\hat{\Omega})H^{+}(t) + F^{\times}(\hat{\Omega})H^{\times}(t). \tag{4}$$

Here, $H^{+,\times}(t)$ represent the perturbations to the residuals when the burst reaches the Earth at time $t$, and $F^{+,\times}$ are so-called antenna pattern functions that describe the response of an Earth-pulsar system to the GW signal, given by

$$F^{+}(\hat{\Omega}) = \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}}, \tag{5}$$

$$F^{\times}(\hat{\Omega}) = \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})}{1+\hat{\Omega}\cdot\hat{p}}, \tag{6}$$

where $\hat{p}$ is a unit vector pointing from the SSB to the pulsar.

### B. Waveform described by piecewise straight lines

By the previous subsection, the timing residuals of the $I$th pulsar caused by a GW burst reaching the Earth at time $t$ can be written as

$$h_I(t) = F_I^{+}H^{+}(t) + F_I^{\times}H^{\times}(t). \tag{7}$$

If the sky location of the pulsar is known, $F_I^{+,\times}$ are functions of the sky location of the burst $(\theta,\phi)$. For certain physical processes, such as GW memory effects or cosmic string cusps, the waveforms $H^{+}(t)$ and $H^{\times}(t)$ can be determined by theories. For a generic burst, however, one needs a signal model that can describe a wide variety of waveforms.

We will model $H^{+}(t)$ and $H^{\times}(t)$ with two independent piecewise linear functions. An illustration is shown in Fig. 1. To this end, we divide the $n_{\rm toas}$ recorded TOAs of the pulsar into $n_H - 1$ (not necessarily even) parts, where $n_H \ll n_{\rm toas}$. Let $\tau_\mu$ be the time at the $n_H$ grid points, with $\mu = 1, 2, \ldots, n_H$. We assign to each $\tau_\mu$ two quantities: $\mathsf{H}_\mu^{+}$ and $\mathsf{H}_\mu^{\times}$. The $2n_H$ quantities $(\,\mathsf{H}_1^{+}\ \ \mathsf{H}_2^{+}\ \cdots\ \mathsf{H}_{n_H}^{+}\ \ \mathsf{H}_1^{\times}\ \ \mathsf{H}_2^{\times}\ \cdots\ \mathsf{H}_{n_H}^{\times}\,) \equiv \mathbf{H}^{\top}$ will be the parameters characterizing the waveform: $H^{+}$ ($H^{\times}$) at any time $t$ can be estimated from the linear combination of the two neighboring quantities in $\mathsf{H}_\mu^{+}$ ($\mathsf{H}_\mu^{\times}$). For example, for $\tau_1 < t < \tau_2$, $H^{+}(t)$ can be estimated as

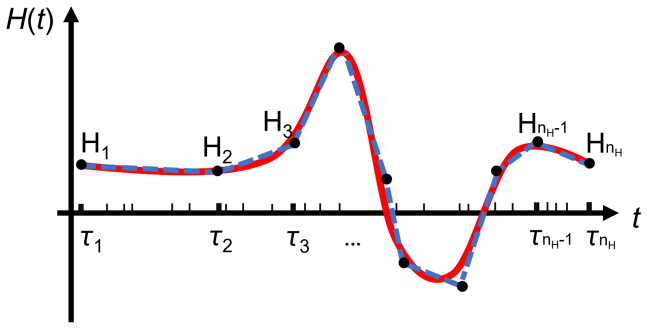


FIG. 1. Illustration of waveform $H(t)$ (red, solid curve) being approximated by piecewise straight lines (blue, dashed). The observation period is divided into $n_H - 1$ parts, each grid point being denoted by $\tau_\mu$, to which $n_H$ quantities $\mathsf{H}_\mu$ are assigned to represent $H(\tau_\mu)$.

$$\begin{aligned} H^{+}(t) &\approx \mathsf{H}_1^{+} + \frac{\mathsf{H}_2^{+} - \mathsf{H}_1^{+}}{\tau_2 - \tau_1}(t - \tau_1) \\ &= \begin{pmatrix} \frac{\tau_2 - t}{\tau_2 - \tau_1} & \frac{t - \tau_1}{\tau_2 - \tau_1} \end{pmatrix} \begin{pmatrix} \mathsf{H}_1^{+} \\ \mathsf{H}_2^{+} \end{pmatrix}. \end{aligned} \tag{8}$$

Let $\mathsf{t}_k$ be the list of TOAs, where $k = 1, 2, \ldots, n_{\rm toas}$. We have

$$\begin{pmatrix} H^{+}(\mathsf{t}_1) \\ H^{+}(\mathsf{t}_2) \\ \vdots \\ H^{+}(\mathsf{t}_{n_{\rm toas}-1}) \\ H^{+}(\mathsf{t}_{n_{\rm toas}}) \end{pmatrix} \approx \begin{pmatrix} \frac{\tau_2 - \mathsf{t}_1}{\tau_2 - \tau_1} & \frac{\mathsf{t}_1 - \tau_1}{\tau_2 - \tau_1} & 0 & \cdots & 0 & 0 \\ \frac{\tau_2 - \mathsf{t}_2}{\tau_2 - \tau_1} & \frac{\mathsf{t}_2 - \tau_1}{\tau_2 - \tau_1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{\tau_{n_H} - \mathsf{t}_{n_{\rm toas}-1}}{\tau_{n_H} - \tau_{n_H-1}} & \frac{\mathsf{t}_{n_{\rm toas}-1} - \tau_{n_H-1}}{\tau_{n_H} - \tau_{n_H-1}} \\ 0 & 0 & 0 & \cdots & \frac{\tau_{n_H} - \mathsf{t}_{n_{\rm toas}}}{\tau_{n_H} - \tau_{n_H-1}} & \frac{\mathsf{t}_{n_{\rm toas}} - \tau_{n_H-1}}{\tau_{n_H} - \tau_{n_H-1}} \end{pmatrix} \begin{pmatrix} \mathsf{H}_1^{+} \\ \mathsf{H}_2^{+} \\ \vdots \\ \mathsf{H}_{n_H-1}^{+} \\ \mathsf{H}_{n_H}^{+} \end{pmatrix} \equiv \mathsf{P}_I \mathbf{H}^{+}, \tag{9}$$

where $\mathsf{P}_I$ is an $n_{\rm toas} \times n_H$ matrix specific to the $I$th pulsar (since it depends on that pulsar's recorded TOAs). The burst signal can then be estimated as

$$h_I \approx F_I^{+}\mathsf{P}_I\mathbf{H}^{+} + F_I^{\times}\mathsf{P}_I\mathbf{H}^{\times} = \begin{pmatrix} F_I^{+}\mathsf{P}_I & F_I^{\times}\mathsf{P}_I \end{pmatrix} \begin{pmatrix} \mathbf{H}^{+} \\ \mathbf{H}^{\times} \end{pmatrix} \equiv \mathsf{S}_I\mathbf{H}, \tag{10}$$

where $\mathsf{S}_I \equiv (\,F_I^{+}\mathsf{P}_I \quad F_I^{\times}\mathsf{P}_I\,)$ is an $n_{\rm toas} \times 2n_H$ matrix. Let $\boldsymbol{h}$ be a concatenated vector composed of all pulsar residuals generated by the burst; we have

$$\boldsymbol{h} \approx \mathbf{S}\mathbf{H}. \tag{11}$$

Here we have defined $\mathbf{S} = (\,\mathsf{S}_1^\top \quad \mathsf{S}_2^\top \quad \cdots \quad \mathsf{S}_{n_{\rm psr}}^\top\,)^\top$, where $n_{\rm psr}$ is the number of pulsars under consideration. An advantage of using $\mathbf{H}$ to describe the waveform is that, compared with higher-order polynomials or Fourier series, locally bad TOAs do not contaminate the estimates of $H^{+,\times}(t)$ over large spans of data [62].

## III. LIKELIHOOD

The timing residuals induced by a burst modeled in the previous section will be used to construct the likelihood of the PTA data. It will be shown in this section that it is possible to analytically integrate out the waveform parameters $\mathbf{H}$. By so doing, one is able to determine the existence and the sky location of the burst without reconstructing the waveform. Such a marginalized likelihood allows an efficient Bayesian search.

### A. Likelihood without deterministic signals

PTA residuals are generated by various processes. In the absence of deterministic signals, the residuals are often modeled as the sum of contributions from timing model deviations, white noise and red noise (such as an SGWB) [66–69]:

$$\boldsymbol{r} = \boldsymbol{n} + \boldsymbol{M}\boldsymbol{\epsilon} + \boldsymbol{F}\boldsymbol{a}. \tag{12}$$

Here, $\boldsymbol{r}$ is a concatenated vector composed of all pulsar residuals; $\boldsymbol{n}$ contains white noise from the radiometer, instrumental effects, etc.; $\boldsymbol{M}$ is the timing model's design matrix basis, and $\boldsymbol{\epsilon}$ is a vector of the corresponding coefficients; $\boldsymbol{F}$ represents the Fourier basis of the red noise, and $\boldsymbol{a}$ is a vector of the corresponding amplitudes. Since $\boldsymbol{n}$ is expected to behave as white noise, the vector $\boldsymbol{r} - \boldsymbol{M}\boldsymbol{\epsilon} - \boldsymbol{F}\boldsymbol{a}$ obeys the Gaussian distribution. To simplify the notation, we shall write a zero-mean Gaussian (normal) distribution with covariance matrix $\boldsymbol{D}$ as

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{D}) \equiv \frac{\exp\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{D}^{-1}\boldsymbol{x}\right)}{\sqrt{\det(2\pi\boldsymbol{D})}}. \tag{13}$$

The PTA likelihood is then given by

$$\mathcal{L}(\boldsymbol{r}|\boldsymbol{b}) = \mathcal{N}(\boldsymbol{r} - \boldsymbol{T}\boldsymbol{b}|\boldsymbol{N}), \tag{14}$$

where $\boldsymbol{N}$ is the white noise covariance matrix, $\boldsymbol{T} = (\,\boldsymbol{M} \quad \boldsymbol{F}\,)$ and $\boldsymbol{b} = (\,\boldsymbol{\epsilon} \quad \boldsymbol{a}\,)^\top$. The prior on $\boldsymbol{b}$ can also be set as Gaussian:

$$\pi(\boldsymbol{b}|\boldsymbol{\eta}) = \mathcal{N}(\boldsymbol{b}|\boldsymbol{B}). \tag{15}$$

Here, the covariance matrix is given by $\boldsymbol{B} = \mathrm{diag}(\infty, \boldsymbol{\phi}(\boldsymbol{\eta}))$, where $\boldsymbol{\eta}$ contains hyperparameters that control $\boldsymbol{B}$. The timing model coefficients $\boldsymbol{\epsilon}$ are well constrained by observations; their inference is likelihood-dominated. Hence, we can impose on them a Gaussian prior of infinite variance. The covariance matrix of the Fourier coefficients $\boldsymbol{a}$ includes all possible intrinsic or common red noise processes:

$$\phi_{(Ii)(Jj)} = \langle a_{Ii} a_{Jj}\rangle = \delta_{ij}(\delta_{IJ}\varphi_{Ii} + \Gamma_{IJ}\Phi_i), \tag{16}$$

where $I$, $J$ range over pulsars and $i$, $j$ over Fourier components; $\delta_{ij}$ is the Kronecker delta; $\varphi_{Ii}$ describes the spectrum of intrinsic red noise in pulsar $I$; and $\Gamma_{IJ}\Phi_i$ describes processes with a common spectrum across all pulsars and interpulsar correlations. For an isotropic SGWB with HD correlations, $\Gamma_{IJ}$ is the HD function of pulsar angular separations, and $\Phi_i$ is usually assumed to obey a power law characterized by amplitude $A$ and spectral index $\gamma$:

$$\Phi_i = \frac{A^2}{12\pi^2}\frac{1}{T}\left(\frac{f_i}{1\ \mathrm{yr}^{-1}}\right)^{-\gamma} 1\ \mathrm{yr}^{-3}. \tag{17}$$

Here, $f_i$ is the frequency of the $i$th Fourier component and $T$ is the maximum TOA's extent. For an SWGB generated by inspiral SMBH binaries, $\gamma$ is expected to be $13/3$ [70].

The full hierarchical PTA posterior can then be written as

$$p(\boldsymbol{b}, \boldsymbol{\eta}|\boldsymbol{r}) = \mathcal{L}(\boldsymbol{r}|\boldsymbol{b})\pi(\boldsymbol{b}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}), \tag{18}$$

where $\pi(\boldsymbol{\eta})$ is the hyperprior on $\boldsymbol{\eta}$. Compared with the hyperparameters in $\boldsymbol{\eta}$—e.g., the amplitude ($A$) and spectral index ($\gamma$) of the SGWB power spectrum—parameters in $\boldsymbol{b}$—i.e., the design matrix coefficients $\boldsymbol{\epsilon}$ and the Fourier coefficients $\boldsymbol{a}$—are usually not of particular interest. Note that the hierarchical likelihood $\mathcal{L}(\boldsymbol{r}|\boldsymbol{b})\pi(\boldsymbol{b}|\boldsymbol{\eta})$ is a Gaussian function for $\boldsymbol{b}$. One can then integrate out these parameters analytically and obtain the marginalized likelihood that only depends on $\boldsymbol{\eta}$ [67,71]:

$$\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta}) = \int \mathcal{L}(\boldsymbol{r}|\boldsymbol{b})\pi(\boldsymbol{b}|\boldsymbol{\eta})\mathrm{d}\boldsymbol{\epsilon}\mathrm{d}\boldsymbol{a} = \mathcal{N}(\boldsymbol{r}|\boldsymbol{C}), \tag{19}$$

where $\boldsymbol{C} = \boldsymbol{N} + \boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^\top$, and we have used the Woodbury identity that gives $\boldsymbol{C}^{-1} = \boldsymbol{N}^{-1} + \boldsymbol{N}^{-1}\boldsymbol{T}(\boldsymbol{B}^{-1} + \boldsymbol{T}^\top\boldsymbol{N}^{-1}\boldsymbol{T})^{-1}\times \boldsymbol{T}^\top\boldsymbol{N}^{-1}$. This is practically the likelihood implemented within production-level GW search pipelines, such as ENTERPRISE[2] [72].

[2]https://github.com/nanograv/enterprise, https://github.com/nanograv/enterprise_extensions.

### B. Marginalized likelihood including the burst

In the previous subsection, we consider the PTA likelihood when only stochastic processes are present. When the burst is taken into account, residuals $\boldsymbol{r}$ in Eq. (19) should be replaced by $\boldsymbol{r}-\boldsymbol{h}\approx\boldsymbol{r}-\mathbf{SH}$. To simplify the notation, we define the inner product $[\![\boldsymbol{x}|\boldsymbol{y}]\!]=\boldsymbol{x}^\top\boldsymbol{C}^{-1}\boldsymbol{y}$. The likelihood in our "$\boldsymbol{h}=\mathbf{SH}$" model is then given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,\mathbf{H}) &= \mathcal{N}(\boldsymbol{r}-\mathbf{SH}|\boldsymbol{C})\\ &= \mathcal{N}(\boldsymbol{r}|\boldsymbol{C})\exp\left([\![\boldsymbol{r}|\mathbf{S}]\!]\mathbf{H}-\frac{1}{2}\mathbf{H}^\top[\![\mathbf{S}|\mathbf{S}]\!]\mathbf{H}\right),\end{aligned} \tag{20}$$

where the first part is simply the likelihood in the absence of deterministic signals [Eq. (19)], and the sky location of the burst $(\theta,\phi)$ only appears in the two inner products $[\![\boldsymbol{r}|\mathbf{S}]\!]$ and $[\![\mathbf{S}|\mathbf{S}]\!]$. Since we do not have any information about the burst waveform, elements in $\mathbf{H}$ can in principle take any real values. A reasonable prior on $\mathbf{H}$ is the Gaussian distribution:

$$\pi(\mathbf{H}|q)=\mathcal{N}(\mathbf{H}|\boldsymbol{Q}), \tag{21}$$

where the covariance matrix is defined to have the form $\boldsymbol{Q}=10^{2q}\boldsymbol{I}$, with $\boldsymbol{I}$ being a $2n_H\times 2n_H$ identity matrix.[3] The entries in $\boldsymbol{Q}$ have units $[\mathrm{s}^2]$. $\mathbf{H}$ plays a similar role to $\boldsymbol{b}$ in Eq. (15), and $q$ is similar to $\boldsymbol{\eta}$ (such as $A$ and $\gamma$). Note that elements in $\mathbf{H}$ are expected to be not much larger than $10^{-6}$ s, which is the order of magnitude of the residuals $\boldsymbol{r}$. Hence, the hyperparameter $q$ should have a value comparable to or smaller than $-6$.

With the likelihood $\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,\mathbf{H})$ and the prior $\pi(\mathbf{H}|q)$ (in addition to the trivial priors on other parameters), we are able to perform a Bayesian analysis in search of the posterior distributions for all model parameters. However, before searching for a burst signal with a PTA dataset, the first question one may ask is "Is there a burst in the data?" rather than "What is the waveform?" If we are only interested in whether a burst exists, note that since both $\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,\mathbf{H})$ and $\pi(\mathbf{H}|q)$ are Gaussian functions for $\mathbf{H}$, we can integrate out the waveform analytically and obtain the marginalized likelihood:

$$\begin{aligned}&\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,q)\\ &=\int\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,\mathbf{H})\pi(\mathbf{H}|q)\mathrm{d}\mathbf{H}\\ &=\int\mathcal{N}(\boldsymbol{r}|\boldsymbol{C})\frac{\exp\left[[\![\boldsymbol{r}|\mathbf{S}]\!]\mathbf{H}-\frac{1}{2}\mathbf{H}^\top([\![\mathbf{S}|\mathbf{S}]\!]+\boldsymbol{Q}^{-1})\mathbf{H}\right]}{\sqrt{\det(2\pi\boldsymbol{Q})}}\mathrm{d}\mathbf{H}\\ &=\mathcal{N}(\boldsymbol{r}|\boldsymbol{C})\frac{\exp\left(\frac{1}{2}[\![\boldsymbol{r}|\mathbf{S}]\!]\boldsymbol{\Sigma}^{-1}[\![\mathbf{S}|\boldsymbol{r}]\!]\right)}{\sqrt{\det(\boldsymbol{Q}\boldsymbol{\Sigma})}},\end{aligned} \tag{22}$$

where we have defined $\boldsymbol{\Sigma}\equiv[\![\mathbf{S}|\mathbf{S}]\!]+\boldsymbol{Q}^{-1}$ in the last step.[4] Therefore, without reconstructing the burst waveform, we are able to search for the noise parameters in $\boldsymbol{C}$ and the sky location of the burst in $\mathbf{S}$ (if there was indeed a burst). The hyperparameter $q$ in $\boldsymbol{Q}$ controls the prior of $\mathbf{H}$. If $q$ is fixed to be a small number—e.g., $q=-9$—only tiny values ($\lesssim 10^{-9}$ in the unit of [s]) would be assigned to $\mathbf{H}$, and hence the search is effectively equivalent to searching in a noise-only model. Therefore, the ratio of $q$'s prior and posterior at small $q$ gives the Savage-Dickey density ratio [73], which is equivalent to the Bayes factor, comparing our model to the model without deterministic signals. If $q$'s marginal posterior does not have support near the lower bound of its prior, our model is strongly preferred over the null model.

### C. Waveform reconstruction

If our model is favored over the noise-only model, one would then be interested in what the burst waveform looks like. In order to reconstruct the waveform, a straightforward way is to go back to the likelihood given by Eq. (20) and perform a Bayesian analysis over all parameters. Depending on how many pieces into which we divide the observation period, the number of parameters in $\mathbf{H}$ could be large, which makes the search computationally expensive. However, it turns out the cost can be reduced significantly if we exploit the samples drawn from the posterior based on Eq. (22) and search for the waveform parameters one by one.

Note again that the part containing $\mathbf{H}$ in Eq. (20) is (part of) a multivariate Gaussian distribution, which has the property that if some of the variables are integrated out, the rest also obey a multivariate Gaussian distribution. The means would be the corresponding means of the original distribution, and the covariance matrix would be the corresponding submatrix of the original one. For example, let $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{D})$ denote a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{D}$. If all variables except for $x_i$ are integrated out, the marginalized distribution of $x_i$ is

[3] Waveform parameters from a certain physical process should be correlated in some way, so the covariance matrix should have off-diagonal entries. However, considering that we do not have any *a priori* information about the signal, and that the number of pieces $n_H$ is much smaller than the number of TOAs, a diagonal $\boldsymbol{Q}$ may not be a bad assumption.

[4] Using the Woodbury identity, $\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,q)$ can be written in a more compact form: $\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta},\theta,\phi,q)=\mathcal{N}(\boldsymbol{r}|\boldsymbol{C}_q)$, where $\boldsymbol{C}_q\equiv\boldsymbol{C}+\mathbf{S}\boldsymbol{Q}\mathbf{S}^\top$. This is of the same form as the likelihood in Eq. (19); the difference is the extra term $\mathbf{S}\boldsymbol{Q}\mathbf{S}^\top$.

$$p(x_i) = \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{D}) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots = \mathcal{N}(x_i|\mu_i, D_{ii}), \tag{23}$$

where $\mu_i$ is the $i$th element of vector $\boldsymbol{\mu}$, and $D_{ii}$ represents the $i$th diagonal element in matrix $\boldsymbol{D}$. In our context, if, for example, we integrate out the last $2n_H - 1$ elements in $\mathbf{H} = (\,\mathsf{H}_1^+ \;\; \mathsf{H}_2^+ \;\; \cdots \;\; \mathsf{H}_{n_H}^+ \;\; \mathsf{H}_1^\times \;\; \mathsf{H}_2^\times \;\; \cdots \;\; \mathsf{H}_{n_H}^\times\,)^\top$, the marginalized likelihood becomes

$$\begin{aligned}\mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta}, \theta, \phi, q, \mathsf{H}_1^+) &= \int \mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta}, \theta, \phi, \mathbf{H}) \pi(\mathbf{H}|q) \mathrm{d}\mathsf{H}_2^+ \mathrm{d}\mathsf{H}_3^+ \ldots \mathrm{d}\mathsf{H}_{n_H}^\times \\ &= \mathcal{N}(\boldsymbol{r}|\boldsymbol{C}) \frac{\exp\left(\frac{1}{2}[\![\boldsymbol{r}|\mathbf{S}]\!]\boldsymbol{\Sigma}^{-1}[\![\mathbf{S}|\boldsymbol{r}]\!]\right)}{\sqrt{\det(\boldsymbol{Q}\boldsymbol{\Sigma})}} \int \mathcal{N}(\mathbf{H}|\boldsymbol{\Sigma}^{-1}[\![\mathbf{S}|\boldsymbol{r}]\!], \boldsymbol{\Sigma}) \mathrm{d}\mathsf{H}_2^+ \mathrm{d}\mathsf{H}_3^+ \ldots \mathrm{d}\mathsf{H}_{n_H}^\times \\ &= \mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta}, \theta, \phi, q) \mathcal{N}(\mathsf{H}_1^+|(\boldsymbol{\Sigma}^{-1}[\![\mathbf{S}|\boldsymbol{r}]\!])_1, \Sigma_{11}),\end{aligned} \tag{24}$$

where in the last step we have used Eq. (22). A similar expression applies to all other parameters in $\mathbf{H}$. This allows us to perform the search of $\mathbf{H}$ one element after another.

This can be achieved as follows: In performing a Bayesian analysis based on the likelihood given by Eq. (22), we have obtained the posterior $p(\boldsymbol{\eta}, \theta, \phi, q|\boldsymbol{r}) \propto \mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta}, \theta, \phi, q)$ from, e.g., MCMC sampling. To find the marginal posterior distribution of, say, $\mathsf{H}_1^+$, we need to integrate out the parameters $\boldsymbol{\eta}$, $\theta$, $\phi$, and $q$ in Eq. (24), which contains the distribution $\propto \mathcal{L}(\boldsymbol{r}|\boldsymbol{\eta}, \theta, \phi, q)$. To find the posterior density at $\mathsf{H}_1^+ = x$, we simply need to take the sum of $\mathcal{N}(x|(\boldsymbol{\Sigma}^{-1}[\![\mathbf{S}|\boldsymbol{r}]\!])_1, \Sigma_{11})$ over all the samples. In other words,

$$p(\mathsf{H}_1^+ = x) \propto \sum_j \mathcal{N}\left(x|(\boldsymbol{\Sigma}^{-1}[\![\mathbf{S}|\boldsymbol{r}]\!])_1^{(j)}, \Sigma_{11}^{(j)}\right), \tag{25}$$

where the superscript $(j)$ represents the $j$th sample in the chain. In principle, $x$ can take any value, but in the presence of a burst, we would expect that $\mathsf{H}_1^+$ only has support near $(\boldsymbol{\Sigma}_*^{-1}[\![\mathbf{S}_*|\boldsymbol{r}]\!])_1$ within the range $\sim\Sigma_{*11}$, where "$_*$" denotes the maximum *a posteriori* value obtained from $p(\boldsymbol{\eta}, \theta, \phi, q|\boldsymbol{r})$. Therefore, we can simply make a grid near this region and evaluate the posterior of $\mathsf{H}_1^+$. This process can be repeated for other parameters in $\mathbf{H}$.

In conclusion, there are two practical "versions" of likelihood in our model. If one wants to know whether a burst signal exists, we simply need to use Eq. (22), which efficiently gives the marginal posterior for $q$. If a burst signal is present, the sky location $(\theta, \phi)$ can also be tracked down from this analysis. If we are interested in what the burst looks like, we could then use Eqs. (24) and (25) to find the posterior distributions of $\mathbf{H}$ without a further Bayesian search. It will be shown in the next section how these two likelihoods are applied in simulated datasets.

## IV. ANALYSES OF SIMULATED DATASETS

In this section, we test our model by analyzing three simulated datasets, each consisting of 20 pulsars. Each pulsar has been observed for 10 years every 30 days. Since the timescale and the shape of the burst signal are unknown, the ideal number of grid points that divide the TOAs and the ideal grid spacing cannot be determined beforehand, and so they should be regarded as parameters. However, fixing the grid can significantly reduce the computational cost. In this work, we divide the PTA observation period into 20 even pieces ($n_H = 21$), leaving the effect of varying $n_H$ and the grid spacing to future work (see discussion in Sec. V).

For simplicity, all simulated residuals have the same constant white noise level of 0.5 μs and no intrinsic pulsar red noise. In addition, an SGWB is injected with a power spectrum given by Eq. (17), with $A = 4 \times 10^{-15}$ ($\log_{10} A \approx -14.398$) and $\gamma = 13/3$. Although the injected SGWB induces correlated red noise among pulsars, in the following analyses we treat the background as an uncorrelated common red noise process. This greatly reduces the computational cost, because the noise matrix $\boldsymbol{C}$ is now block diagonal, which allows $\boldsymbol{C}^{-1}$ to be computed block by block (or pulsar by pulsar). Consequently, the inner products $[\![\boldsymbol{r}|\mathbf{S}]\!]$ and $[\![\mathbf{S}|\mathbf{S}]\!]$ can be obtained rather efficiently. Ignoring pulsar correlations in $\boldsymbol{C}$ could bias the burst search by overestimating the significance of the potential signal, but since the correlations should have a smaller effect compared to the common spectrum, using a block-diagonal $\boldsymbol{C}$ is not expected to significantly affect the results in the simple scenarios we are considering in this work.

Following Ref. [59], a burst signal from the parabolic encounter of two SMBHs is injected. The waveform in Ref. [59] was obtained from the quadrupole formula

TABLE I. Prior distributions for $\log_{10} A$, $\gamma$, $\cos\theta$, $\phi$, and $q$ and the true values in the simulated datasets.

| Parameter | Prior range (uniform) | True value |
|---|---|---|
| $\log_{10} A$ | $[-18, -13]$ | $-14.398$ |
| $\gamma$ | $[0, 7]$ | $13/3$ |
| $\cos\theta$ | $[-1, 1]$ | 0.5 |
| $\phi$ | $[0, 2\pi]$ | 3 |
| $q$ | $[-9, -5]$ | |

applied to a parabolic Kepler orbit. It serves as a first approximation of the encounter event and as an exemplar for the purpose of demonstrating how our model works. The event under consideration is set to occur at sky location $(\cos\theta, \phi) = (0.5, 3)$. The two black holes have the same mass $10^9 M_\odot$, and the impact parameter is $2 \times 10^{11} M_\odot$ (where we have set $G = c = 1$). The injected signal is also set to sit in the middle of the observation period.

Similarly to Ref. [59], we shall test our model with datasets of different signal-to-noise ratios (SNRs, defined by $\mathrm{SNR} = \sqrt{[\![\boldsymbol{h}|\boldsymbol{h}]\!]}$). In the first two datasets, the encounter event occurs at different distances from us: 20 Mpc (strong signal, $\mathrm{SNR} \approx 14.7$) and 45 Mpc (weak signal, $\mathrm{SNR} \approx 6.5$). For each case, we first perform an analysis based on the noise-only model, with the likelihood given by Eq. (19), which only contains two parameters: $A$ and $\gamma$. Then, we search for the existence and sky location of the burst using the likelihood given by Eq. (22), which contains five parameters: $A$, $\gamma$, $\theta$, $\phi$, and $q$. Lastly, we reconstruct the waveform according to Eq. (25) and compare the results with the injected signals. In addition, we also test our model with a dataset that contains no burst signals. The priors on $\log_{10} A$, $\gamma$, $\cos\theta$, $\phi$, and $q$ are all set as uniform distributions, with bounds shown in Table I.

Besides the above three datasets that contain an SGWB, we shall also study datasets that contain a burst signal only and no SGWB.[5] The purpose is to test the possibility of the burst being mistaken as a background.

In what follows, the Bayesian inferences are achieved by Nestle, a Python implementation of the nested sampling algorithm [74–78] aiming at comparing models and generating samples from posterior distributions.[6]

### A. Strong signal (SNR ≈ 14.7)

We first consider a case where a strong signal is injected into the background. The source is placed at a distance of 20 Mpc. In Fig. 2, we show the postfit timing residuals[7] and

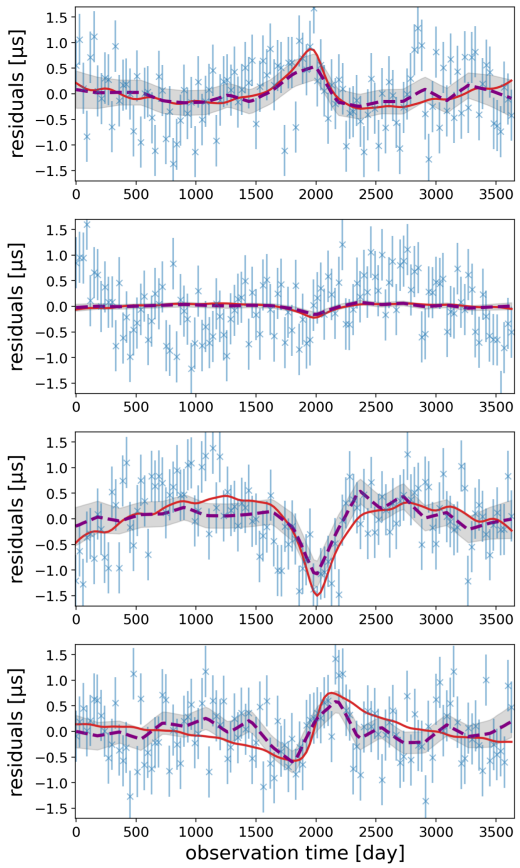


FIG. 2. Postfit residuals (blue data points with error bars) and perturbations from the injected strong burst (solid red curves) for four of the 20 pulsars. The reconstructed burst-induced residuals are shown as dashed, purple curves (with 90% confidence intervals), where the sky location of the burst is taken as the posterior medians for $\cos\theta$ and $\phi$.

the contribution from the injected burst for four of the 20 pulsars. We first analyze the dataset with the noise-only model, where the residuals are assumed to be generated by white noise and an SGWB. With the white noise parameters fixed, there are only two free parameters: the amplitude of the SGWB spectrum $A$ and the spectral index $\gamma$. Using the likelihood given by Eq. (19), we obtain the posterior distributions shown in Fig. 3. The median values ($\log_{10} A \approx -13.66$ and $\gamma \approx 2.79$) obviously deviate from the true values ($\log_{10} A \approx -14.398$ and $\gamma \approx 4.33$). In the presence of the strong signal, the noise-only model is not able to recover the SGWB faithfully. Since the noise-only model treats the burst signal as part of the background, the "detected" magnitude of the SGWB spectrum is larger than the true value.

We then perform a Bayesian analysis with the likelihood given by Eq. (22), which contains three additional parameters: $\theta$, $\phi$, and $q$. The posterior distributions are shown in Fig. 4. We can see that the SGWB parameters are now captured by our model, with median values and true values

[5]We would like to thank the anonymous referee for suggesting these tests.

[6]https://github.com/kbarbary/nestle.

[7]Here "postfit" means that the contribution in residuals fitted by the timing model, such as the quadratic components, has been removed.

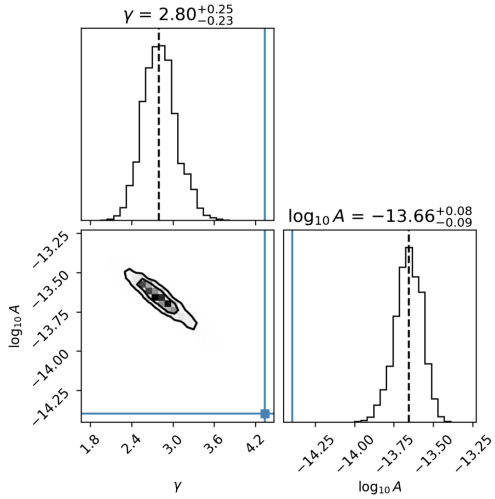


FIG. 3. Posterior distributions of the SGWB parameters $A$ and $\gamma$ in the noise-only model when a strong burst signal is present. The dashed lines represent the median values. The distributions are obviously incompatible with the true values in the simulated dataset ($\log_{10} A \approx -14.398$ and $\gamma \approx 4.33$), represented by blue lines.

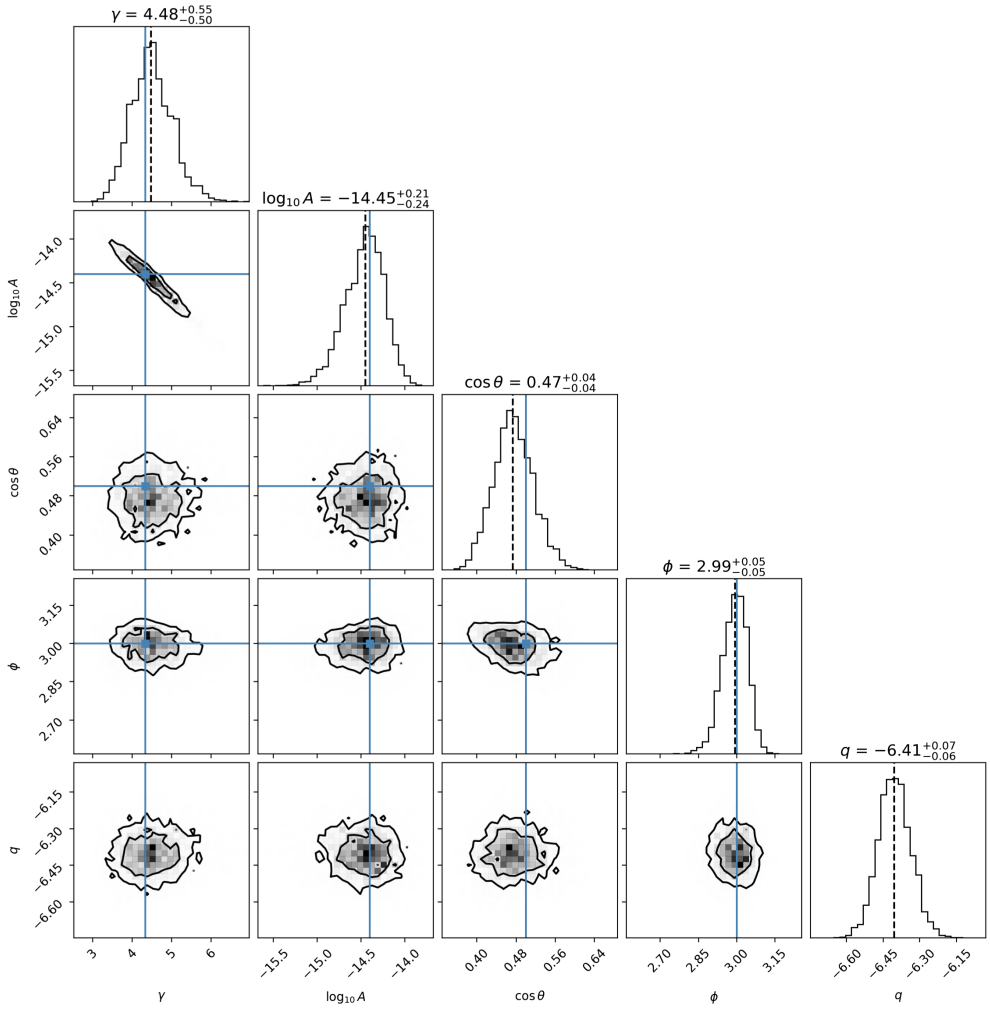


FIG. 4. Posterior distributions of parameters $\log_{10} A$, $\gamma$, $\cos\theta$, $\phi$, and $q$ in our model when a strong burst signal is present. The dashed lines represent the median values, and the blue solid lines denote the true values in the simulated dataset. The red noise parameters are recovered, and the location of the burst is detected. The existence of the burst has significant evidence, since the hyperparameter $q$ has no samples near the lower bound of its prior ($q_{\min} = -9$).

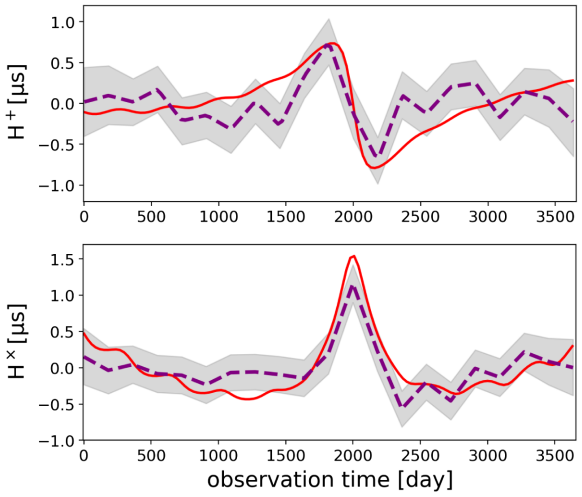


FIG. 5. Injected $H^+(t)$ and $H^\times(t)$ (postfit) for a strong-burst signal (red curves), the reconstructed piecewise linear functions $\mathsf{H}^+_\mu$ and $\mathsf{H}^\times_\mu$ (dashed, purple curves, median values) and 90% confidence intervals (shaded region), where the sky location of the burst is taken as the posterior medians for $\cos\theta$ and $\phi$.

being rather close to each other.[8] The hyperparameter $q$ is only sampled near $-6.4$, which indicates a large Savage-Dickey ratio. In fact, the Bayes factor for our model to the noise-only model (based on the Nestle results) is $\sim 10^{10}$, corresponding to overwhelming evidence for the presence of a signal. From the marginal posterior distributions of $\cos\theta$ and $\phi$, the injected burst is also perfectly localized on the sky map by our model.

Following the method described in Sec. III C, we then use the samples from the last paragraph to infer the burst waveform. The best-fit piecewise straight lines describing $H^+$ and $H^\times$ are shown in Fig. 5 by dashed, purple curves with 90% confidence intervals. The red curves are the postfit injected $H^+$ and $H^\times$. We also show in Fig. 2 the reconstructed burst-induced residuals for four pulsars.

### B. Weak signal (SNR ≈ 6.5)

Now we turn to a weaker signal from the encounter event occurring at a distance of 45 Mpc. Figure 6 shows the timing residuals and the contribution from the injected burst for four of the 20 pulsars (the same as those in Fig. 2). By analyzing the dataset with the noise-only model, we obtain the posteriors shown in Fig. 7. Unlike the case in the previous subsection, the distributions of $\log_{10} A$ and $\gamma$ here are compatible with the true values. The burst signal is so weak that the noise-only model is capable of detecting the injected SGWB.

[8]How well the median and true values agree depends on the realization of simulated data. With the background parameters (white noise level, $A$ and $\gamma$) fixed, different realizations of the stochastic feature (i.e., different sets of random number generators) can lead to statistical errors.

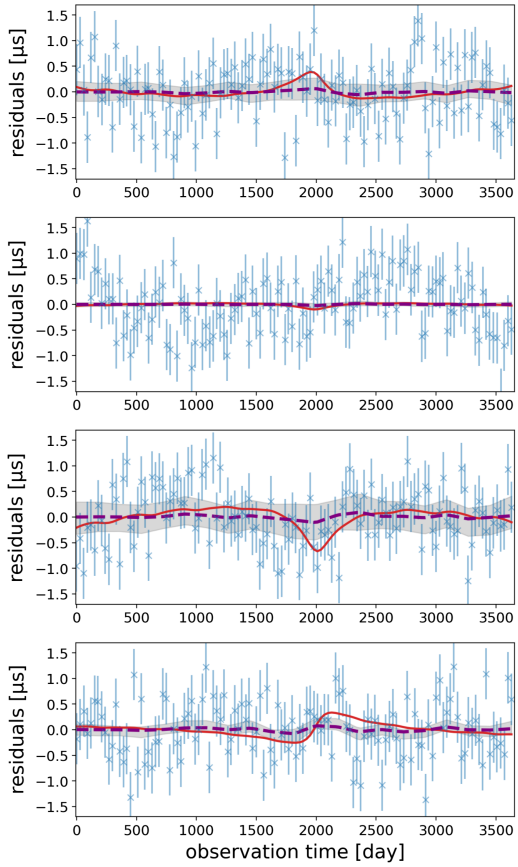


FIG. 6. Postfit residuals (blue data points with error bars) and perturbations from the injected weak burst (solid red curves) for four of the 20 pulsars. These are the same pulsars as those in Fig. 2. The difference in residuals for each pulsar comes from the distance of the encounter event. The reconstructed burst-induced residuals are shown as dashed, purple curves (with 90% confidence intervals), where the sky location of the burst is taken as the posterior medians for $\cos\theta$ and $\phi$.

We then search for the burst using our model with additional parameters $\theta$, $\phi$, and $q$. The posterior distributions are shown in Fig. 8. We can see that the SGWB spectrum is also well captured, with median values and true values almost overlapping. The distribution of $q$ has a tail extending to the lower bound of its prior. Both the Savage-Dickey ratio and the model evidences computed by Nestle give a Bayes factor of ∼2–3, corresponding to fairly weak evidence supporting our model. However, the sky location of the injected burst is captured by the two peaks in the posteriors of $\cos\theta$ and $\phi$. The reconstructed waveform is shown in Fig. 9. We also show in Fig. 6 the reconstructed burst-induced residuals for four pulsars. As expected, the injected signal is too weak to be fully characterized by our model.

### C. No signal

In the absence of deterministic signals, our model should be consistent with the noise-only model. Here we consider

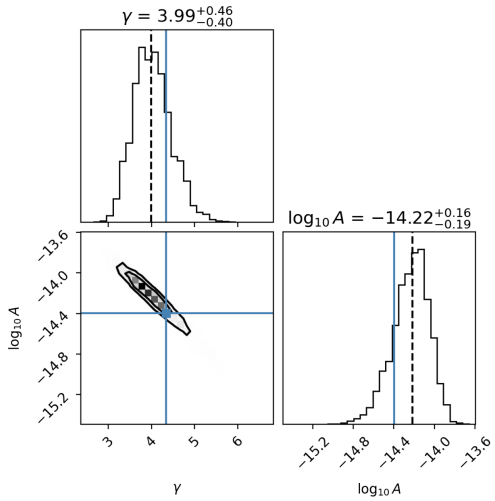


FIG. 7. Posterior distributions of the SGWB parameters $A$ and $\gamma$ in the noise-only model when a weak burst signal is present. The dashed lines represent the median values, and the blue solid lines denote the true values in the simulated dataset. For such a weak signal, the SGWB parameters are properly captured by our model.

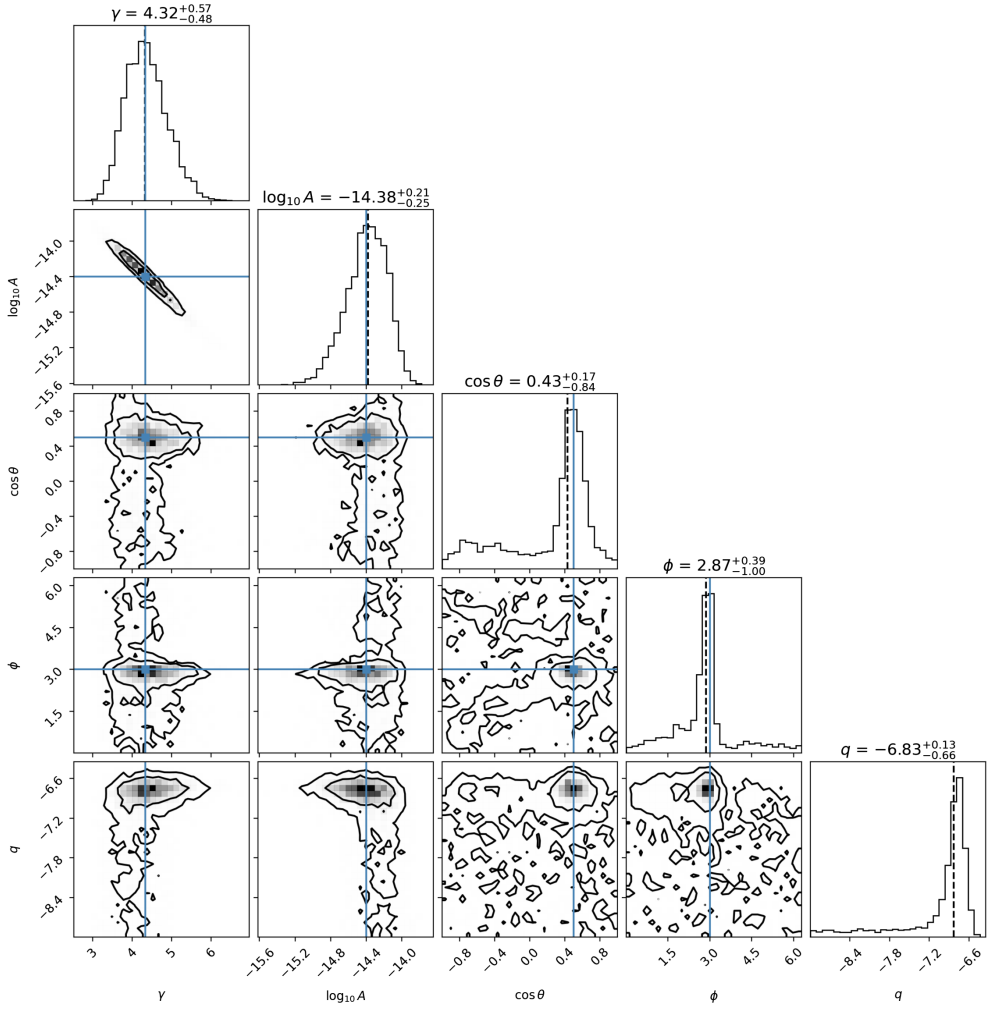


FIG. 8. Posterior distributions of parameters $\log_{10} A$, $\gamma$, $\cos\theta$, $\phi$, and $q$ in our model when a weak signal is present. The dashed lines represent the median values, and the blue solid lines denote the true values in the simulated dataset. The red noise parameters are recovered, and the location of the burst is detected. However, the evidence supporting the existence of the burst is not significant (see text).

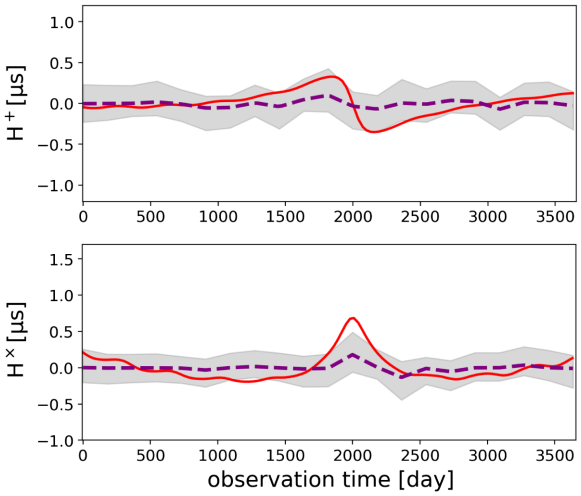


FIG. 9. Injected $H^+(t)$ and $H^\times(t)$ (postfit) for a weak-burst signal (red curves), the reconstructed piecewise linear functions $\mathsf{H}_\mu^+$ and $\mathsf{H}_\mu^\times$ (purple curves), and 90% confidence intervals (shaded region), where the sky location of the burst is taken as the posterior medians for $\cos\theta$ and $\phi$. For such a weak signal, our model is not able to characterize the burst waveform.

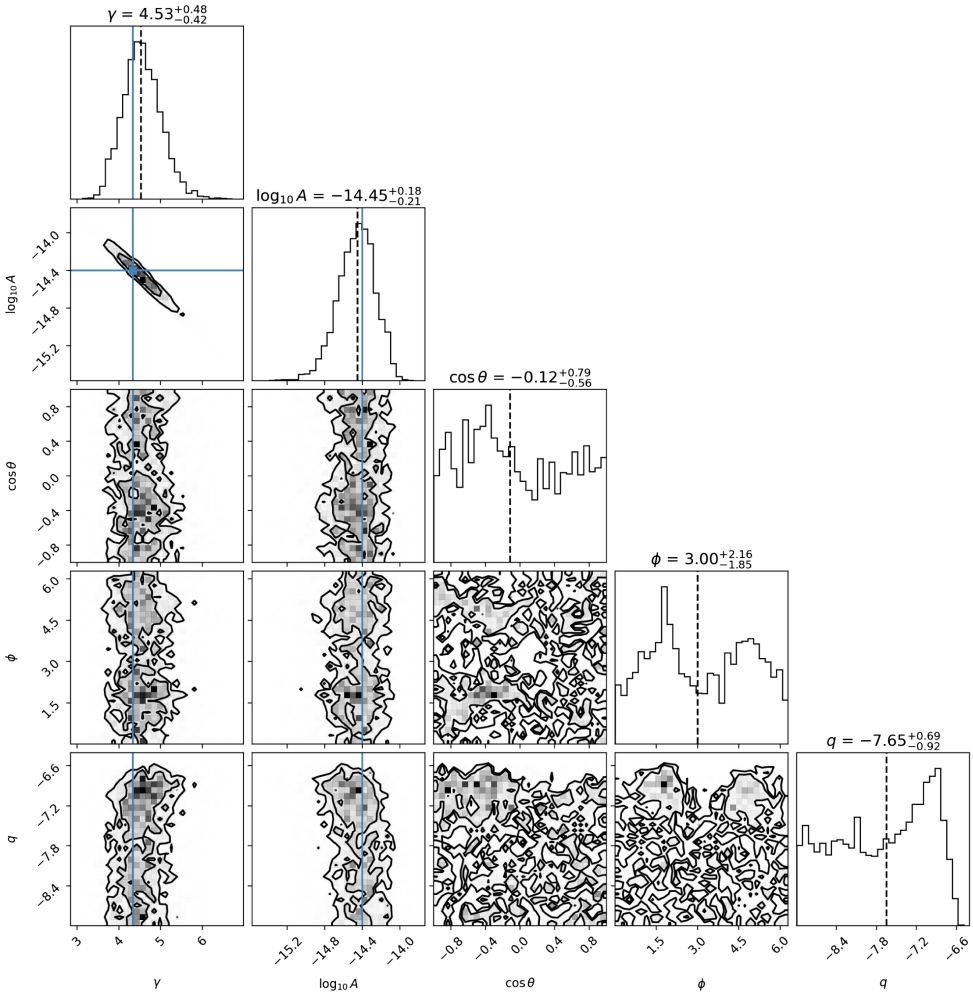


FIG. 10. Posterior distributions of parameters $\log_{10} A$, $\gamma$, $\cos\theta$, $\phi$, and $q$ when no burst signal is present. The dashed lines represent the median values, and the blue solid lines denote the true values in the simulated dataset. Our model does not detect any burst signal in this dataset. The distributions of the SGWB parameters are compatible with the true values.

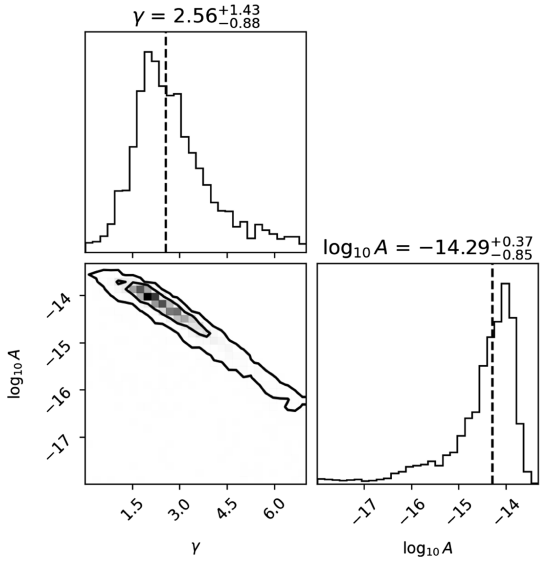


FIG. 11. Posterior distributions of the SGWB parameters $\log_{10} A$ and $\gamma$ in the noise-only model when only a burst signal and no SGWB is present. The dashed lines represent the median values. The model more or less sees a common red noise spectrum.

a dataset without a burst. The corner plots for our model are shown in Fig. 10. The Bayes factor for our model vs the noise-only model is ∼1, which means our model is capable of describing datasets that contain no evidence of a burst signal.

### D. No background

For completeness, in this subsection we consider the case where only a burst signal and no SGWB is present. Such a dataset is interesting, as it can show whether the transient signal could be mistaken as a background. For instance, in Ref. [79], a simulated dataset containing a continuous wave signal and no SGWB prefers an SGWB-only model.

In Figs. 11 and 12, we show the results based on a dataset where a burst signal the same as that in Sec. IV B is injected. Figure 11 is from the noise-only model, which more or less extracts a common red noise spectrum from the data. Figure 12 is from our model. By the Savage-Dickey ratio, it is obvious from the posteriors of $\log_{10} A$ and $q$ that our model or the burst-only model is strongly preferred over the noise-only model. We can also see that, compared to Fig. 8, now that the common red noise is removed from the data, the sky location of the burst source is clearly detected.

We have also analyzed datasets with various SNRs. At least in our case, where the burst comes from the encounter of two SMBHs, the presence of an SGWB is always disfavored, whereas the burst is either detected, or is drown out by white noise.

## V. CONCLUSIONS AND DISCUSSION

In this work, we have investigated a method for performing Bayesian analyses on PTA datasets to search for the strongest burst signals. The burst waveform is modeled by piecewise straight lines, which allows it the likelihood to have a simple form. Although the number of waveform parameters $\mathbf{H}$ could be large [of order $\mathcal{O}(10)$ depending on how the observation period is divided], these parameters can be analytically integrated out if their priors follow a Gaussian distribution. The resulting marginalized likelihood [Eq. (22)] has only three parameters in addition to the intrinsic and common noise parameters. Among the three parameters, $q$ controls the prior of the waveform parameters $\mathbf{H}$; its marginal posterior can immediately tell us whether our model is favored over the noise-only model. The other two parameters, $\theta$ and $\phi$, denote the sky location of the burst. If a signal is present, one can efficiently retrieve the posterior of the waveform by analyzing the MCMC samples of $q$, $\theta$, and $\phi$ based on the marginalized likelihood.

We tested this model by analyzing three simulated PTA datasets, the first two containing a burst signal generated by the parabolic encounter of two SMBHs, and the third containing no burst signals. For the strong signal (with SNR $\approx 14.7$), our model is strongly favored compared to the noise-only model; not only can the burst's sky location be detected, but its waveform can also be extracted to a reasonable accuracy. For a weak signal (with SNR $\approx 6.5$), although the waveform cannot be distinguished from the background, the marginal posteriors of the sky location peak near the true values. When the signal is absent, the Bayes factor for our model to the noise-only model becomes ∼1. For completeness, we have also analyzed datasets that do not contain an SGWB, but a burst signal only. The presence of the burst is always favored compared to a noise-only scenario.

Our model could be improved in several ways. As mentioned previously, in performing the piecewise linear fit, how the observation period is divided was predetermined. We fixed the grid point number $n_H$, with each segment containing a similar number of TOAs in each pulsar. In real data, however, TOAs are not evenly sampled in time, and they may vary significantly from one pulsar to another. We expect that the optimal number of grid points and the optimal spacing will both depend on the signal. A higher SNR signal requires a larger $n_H$, and a signal highly concentrated in a particular time span requires more points in that region and fewer elsewhere. Noting that varying the grid spacing adaptively tends to significantly increase the computational cost, a straightforward extension of the current framework is to treat $n_H$ as a free parameter while ensuring we have a sufficient number of TOAs for each piece, or to test different $n_H$'s and then perform model selection on $n_H$.

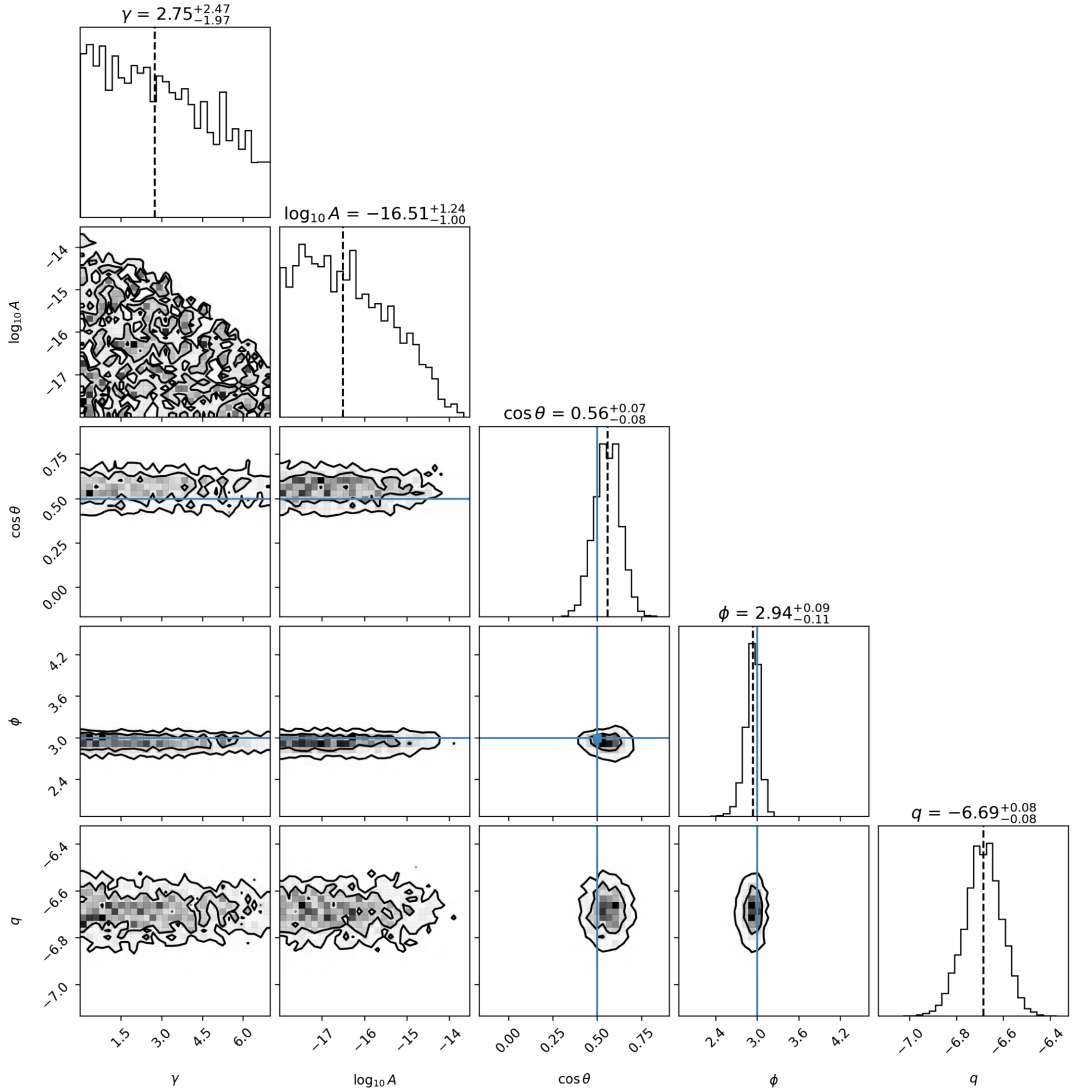


FIG. 12. Posterior distributions of parameters $\log_{10} A$, $\gamma$, $\cos\theta$, $\phi$, and $q$ when only a burst signal and no SGWB is present. The dashed lines represent the median values, and the blue solid lines denote the true values in the simulated dataset. Our model detects the sky location of the source and is strongly preferred over the noise-only model.

Finally, the burst may also be better characterized if the prior on the waveform parameters $\mathbf{H}$ is not controlled by a single hyperparameter. In the above analyses, we assume the prior $\pi(\mathbf{H}|q)$ is a Gaussian distribution with covariance $\boldsymbol{Q}$ determined by $q$ only. However, the amplitude of the waveform can differ between $H^+(t)$ and $H^\times(t)$, and it may vary significantly over the observation period. Roughly speaking, regions with a smaller amplitude require a smaller $q$. Hence, a natural extension of the current framework is to set a number of hyperparameters (e.g., $q_1, q_2, \ldots, q_n$, where $n \ll n_H$ in order not to lose the efficiency of the model) responsible for different polarization components and different time domains.

We plan on improving the model accordingly and applying it to the search for a gravitational wave burst signal in real datasets in the near future.

## ACKNOWLEDGMENTS

This project receives support from National Science Foundation (NSF) Physics Frontiers Center Grants No. 1430284 and No. 2020265.

[1] R. W. Hellings and G. S. Downs, Astrophys. J. Lett. **265**, L39 (1983).
[2] M. A. McLaughlin, Classical Quantum Gravity **30**, 224008 (2013).
[3] G. Agazie *et al.* (NANOGrav Collaboration), Astrophys. J. Lett. **951**, L8 (2023).
[4] H. Xu *et al.*, Res. Astron. Astrophys. **23**, 075024 (2023).
[5] J. Antoniadis *et al.* (EPTA Collaboration), arXiv:2306.16214.
[6] D. J. Reardon *et al.*, Astrophys. J. Lett. **951**, L6 (2023).
[7] J. Kormendy and L. C. Ho, Annu. Rev. Astron. Astrophys. **51**, 511 (2013).
[8] G. Agazie *et al.* (NANOGrav Collaboration), Astrophys. J. Lett. **952**, L37 (2023).
[9] L. P. Grishchuk, Zh. Eksp. Teor. Fiz. **67**, 825 (1974).
[10] A. A. Starobinsky, JETP Lett. **30**, 682 (1979).
[11] V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin, Phys. Lett. **115B**, 189 (1982).
[12] R. Fabbri and M. D. Pollock, Phys. Lett. **125B**, 445 (1983).
[13] A. Kosowsky, M. S. Turner, and R. Watkins, Phys. Rev. D **45**, 4514 (1992).
[14] A. Kosowsky, M. S. Turner, and R. Watkins, Phys. Rev. Lett. **69**, 2026 (1992).
[15] A. Kosowsky and M. S. Turner, Phys. Rev. D **47**, 4372 (1993).
[16] M. Kamionkowski, A. Kosowsky, and M. S. Turner, Phys. Rev. D **49**, 2837 (1994).
[17] C. Caprini, R. Durrer, and G. Servant, Phys. Rev. D **77**, 124015 (2008).
[18] S. J. Huber and T. Konstandin, J. Cosmol. Astropart. Phys. 09 (2008) 022.
[19] M. Hindmarsh, S. J. Huber, K. Rummukainen, and D. J. Weir, Phys. Rev. Lett. **112**, 041301 (2014).
[20] J. T. Giblin, Jr. and J. B. Mertens, J. High Energy Phys. 12 (2013) 042.
[21] J. T. Giblin and J. B. Mertens, Phys. Rev. D **90**, 023532 (2014).
[22] M. Hindmarsh, S. J. Huber, K. Rummukainen, and D. J. Weir, Phys. Rev. D **92**, 123009 (2015).
[23] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Phys. Rev. Lett. **127**, 251302 (2021).
[24] A. Vilenkin, Phys. Lett. **107B**, 47 (1981).
[25] T. Damour and A. Vilenkin, Phys. Rev. D **71**, 063510 (2005).
[26] W. Buchmuller, V. Domcke, and K. Schmitz, Phys. Lett. B **811**, 135914 (2020).
[27] J. Ellis and M. Lewicki, Phys. Rev. Lett. **126**, 041304 (2021).
[28] J. J. Blanco-Pillado, K. D. Olum, and J. M. Wachter, Phys. Rev. D **103**, 103512 (2021).
[29] M. Hindmarsh and J. Kume, J. Cosmol. Astropart. Phys. 04 (2022) 045.
[30] A. Afzal *et al.* (NANOGrav Collaboration), Astrophys. J. Lett. **951**, L11 (2023).
[31] J. Antoniadis *et al.*, arXiv:2306.16227.
[32] C. Smarra *et al.*, arXiv:2306.16228.
[33] A. Sesana, A. Vecchio, and M. Volonteri, Mon. Not. R. Astron. Soc. **394**, 2255 (2009).
[34] P. A. Rosado, A. Sesana, and J. Gair, Mon. Not. R. Astron. Soc. **451**, 2417 (2015).
[35] L. Z. Kelley, L. Blecha, L. Hernquist, A. Sesana, and S. R. Taylor, Mon. Not. R. Astron. Soc. **477**, 964 (2018).
[36] B. Bécsy, N. J. Cornish, and L. Z. Kelley, Astrophys. J. **941**, 119 (2022).
[37] D. R. B. Yardley *et al.*, Mon. Not. R. Astron. Soc. **407**, 669 (2010).
[38] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Astrophys. J. **794**, 141 (2014).
[39] X. J. Zhu *et al.*, Mon. Not. R. Astron. Soc. **444**, 3709 (2014).
[40] S. Babak *et al.*, Mon. Not. R. Astron. Soc. **455**, 1665 (2016).
[41] K. Aggarwal *et al.*, Astrophys. J. **880**, 116 (2019).
[42] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Astrophys. J. Lett. **951**, L28 (2023).
[43] M. Falxa *et al.* (IPTA Collaboration), Mon. Not. R. Astron. Soc. **521**, 5077 (2023).
[44] G. Agazie *et al.* (NANOGrav Collaboration), Astrophys. J. Lett. **951**, L50 (2023).
[45] J. Antoniadis *et al.*, arXiv:2306.16226.
[46] T. Damour and A. Vilenkin, Phys. Rev. Lett. **85**, 3761 (2000).
[47] T. Damour and A. Vilenkin, Phys. Rev. D **64**, 064008 (2001).
[48] N. Yonemaru *et al.*, Mon. Not. R. Astron. Soc. **501**, 701 (2021).
[49] J. M. Cordes and F. A. Jenet, Astrophys. J. **752**, 54 (2012).
[50] D. R. Madison, J. M. Cordes, and S. Chatterjee, Astrophys. J. **788**, 141 (2014).
[51] R. van Haasteren and Y. Levin, Mon. Not. R. Astron. Soc. **401**, 2372 (2010).
[52] Z. Arzoumanian *et al.* (NANOGrav Collaboration), Astrophys. J. **810**, 150 (2015).
[53] J. B. Wang *et al.*, Mon. Not. R. Astron. Soc. **446**, 1657 (2015).
[54] K. Aggarwal *et al.* (NANOGrav Collaboration), Astrophys. J. **889**, 38 (2020).
[55] J. Sun, P. T. Baker, A. D. Johnson, D. R. Madison, and X. Siemens, Astrophys. J. **951**, 121 (2023).
[56] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), Phys. Rev. D **104**, 122004 (2021).
[57] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), Phys. Rev. D **104**, 102001 (2021).
[58] M. J. Szczepańczyk *et al.*, Phys. Rev. D **107**, 062002 (2023).
[59] L. S. Finn and A. N. Lommen, Astrophys. J. **718**, 1400 (2010).
[60] X. Deng, Phys. Rev. D **90**, 024020 (2014).
[61] X.-J. Zhu, L. Wen, G. Hobbs, Y. Zhang, Y. Wang, D. R. Madison, R. N. Manchester, M. Kerr, P. A. Rosado, and J.-B. Wang, Mon. Not. R. Astron. Soc. **449**, 1650 (2015).
[62] D. R. Madison *et al.*, Mon. Not. R. Astron. Soc. **455**, 3662 (2016).
[63] B. Bécsy and N. J. Cornish, Classical Quantum Gravity **38**, 095012 (2021).
[64] B. Bécsy, R. Burnette, and J. Taylor (to be published).
[65] F. B. Estabrook and H. D. Wahlquist, Gen. Relativ. Gravit. **6**, 439 (1975).
[66] R. van Haasteren, Y. Levin, P. McDonald, and T. Lu, Mon. Not. R. Astron. Soc. **395**, 1005 (2009).
[67] L. Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren, Phys. Rev. D **87**, 104021 (2013).

[68] R. van Haasteren and M. Vallisneri, Phys. Rev. D **90**, 104012 (2014).
[69] S. R. Taylor, *Nanohertz Gravitational Wave Astronomy* (CRC Press, Boca Raton, FL, 2021).
[70] E. Phinney, arXiv:astro-ph/0108028.
[71] R. van Haasteren and Y. Levin, Mon. Not. R. Astron. Soc. **428**, 1147 (2013).
[72] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite, Zenodo (2020).
[73] J. M. Dickey, Ann. Math. Stat. **42**, 204 (1971).
[74] J. Skilling, AIP Conf. Proc. **735**, 395 (2004).
[75] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (OUP, Oxford, 2006).
[76] R. Shaw, M. Bridges, and M. P. Hobson, Mon. Not. R. Astron. Soc. **378**, 1365 (2007).
[77] P. Mukherjee, D. Parkinson, and A. R. Liddle, Astrophys. J. Lett. **638**, L51 (2006).
[78] F. Feroz, M. P. Hobson, and M. Bridges, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).
[79] J. Antoniadis *et al.* (EPTA Collaboration), arXiv:2306.16226.