

Finding supermassive black hole binary mergers in Pulsar Timing Array data

Sharon Mary Tomson^{✉,*}, Boris Goncharov, and Rutger van Haasteren

Max Planck Institute for Gravitational Physics (Albert Einstein Institute), 30167 Hannover, Germany
Leibniz Universität Hannover, 30167 Hannover, Germany



(Received 8 October 2025; accepted 31 March 2026; published 10 June 2026)

Galaxy observations suggest that mergers of supermassive black hole binaries (SMBHBs) are rare events on cosmological scales, with characteristic rates of order one per decade across the observable Universe. In this work, we present a framework to search for merging SMBHBs in pulsar timing array (PTA) data using a physically complete waveform model that includes the late inspiral, merger, ringdown, and gravitational-wave memory. This approach enables a unified treatment of continuous premerger emission and the nonoscillatory memory signal associated with coalescence. Using simulated PTA datasets, we demonstrate a proof of principle for parameter estimation of merging SMBHBs, focusing on representative systems with chirp masses of 10^8 and $10^{10}M_{\odot}$ at distances of 3 Mpc and 100 Mpc respectively. These signals are distinguishable for a PTA consisting of 25 pulsars observed over 13 years with 100 ns timing precision. For these sufficiently strong signals, we recover the simulated binaries with log Bayes factors exceeding 10, and show that chirp mass and luminosity distance can be jointly constrained with uncertainties governed by the characteristic mass-distance degeneracy of the signal. The sky position uncertainties are of the order of a few degrees, which could potentially enable electromagnetic follow-up and multimessenger observations of SMBHB mergers. We also show that the measurement uncertainties on the parameters of simulated merging binaries depend weakly on the presence of the gravitational wave background with Hellings-Downs correlations in our simulated data. We further demonstrate that commonly used memory burst approximations lead to biased strain amplitudes and inferred source parameters when compared to the full SMBHB waveform, even when optimally tuned. These results establish a pathway toward searching for SMBHB mergers and late-stage inspirals with PTAs using physically complete waveform models, complementing existing searches for continuous gravitational waves.

DOI: [10.1103/99r4-ks5s](https://doi.org/10.1103/99r4-ks5s)

I. INTRODUCTION

Pulsar Timing Arrays (PTAs) have the capability to detect gravitational waves (GWs) in the nanohertz frequency regime by precisely monitoring variations in the times of arrival (TOAs) of radio pulses emitted by millisecond pulsars [1]. These subtle fluctuations in TOAs are caused by GWs propagating through spacetime, perturbing the metric and inducing correlated timing deviations in widely separated pulsars. By analyzing TOAs, PTAs provide a unique means of probing low-frequency gravitational radiation.

Currently, several PTA collaborations are active: the North American Nanohertz Observatory for Gravitational

Waves (NANOGrav) [2], the Parkes Pulsar Timing Array (PPTA) [3], the European Pulsar Timing Array (EPTA) [4], the Indian Pulsar Timing Array [5], the Chinese PTA [6] and the MeerKAT PTA [7]. The International Pulsar Timing Array [8] is a consortium of PTAs.

It is expected that the stochastic gravitational wave background (GWB) will become the first gravitational wave signal observed by PTAs. Multiple lines of evidence for the background have been recently reported [9–18]. Such a background is expected to originate from a population of inspiraling supermassive black hole binaries (SMBHBs), although the early-Universe origins are also not ruled out [19,20].

In addition to the background, PTAs also search for deterministic signals which could be brighter and individually resolvable, such as the continuous gravitational waves (CGWs) from individual SMBHBs [21], burst events from cosmic string cusps [22,23], SMBHB mergers or hyperbolic binaries [24], and the gravitational wave memory [25–27]. Recently, PTAs have begun targeted CGW searches of active galactic nuclei candidates selected from

*Contact author: sharon.mary.tomson@aei.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

electromagnetic (EM) surveys, using EM-informed priors on sky location, distance/redshift, and CGW frequency [28]. Complementing these targeted searches, morphology-agnostic methods for generic burst signals have recently been developed for PTA datasets, enabling efficient, model-independent burst detection in the nanohertz band [29].

In this context, mergers of SMBHBs are of particular interest because they produce a component of the gravitational wave signal referred to as memory. The most prominent memory component, the *displacement memory*, manifests as the accumulating displacement of the freely falling test particles as a gravitational wave passes by. Theoretical work over the past few decades has classified memory, including displacement memory, into two main types: ordinary and null [25].

Ordinary memory—also called linear memory—occurs in systems where matter is ejected from the system: hyperbolic encounters of compact objects [24,30], asymmetric neutrino emission from supernovae [31,32], or relativistic jets from gamma-ray bursts [33]. Ordinary memory was first discovered in the 1970s by considering hyperbolic encounters of compact objects [34,35]. Search for linear memory from SMBHBs on hyperbolic trajectories has been performed in the NANOGrav 12.5-year dataset, finding no compelling evidence and placing limits on the existence of such systems [27]. Ordinary memory is negligible for SMBHB mergers.

Null memory, previously mostly referred to as nonlinear or Christodoulou [36] memory, is due to the contribution of gravitational radiation itself to the evolution of the system’s changing quadrupole and higher mass moments [37]. This implies that almost all GW sources inherently produce some level of memory. For quasicircular compact binary mergers, which are among the most significant and well-studied sources, the null memory contributes substantially to the time-domain waveform. In the post-Newtonian (PN) framework, the memory enters at the Newtonian (leading) order. This arises because the memory is a hereditary effect—it depends not just on the current configuration of the source, but on its entire past history. Furthermore, its nonoscillatory behavior makes it stand out in the waveform, providing a clear signature distinct from the oscillatory inspiral and merger phases. Because of these properties, null memory is not only a powerful probe of GW sources but also serves as a unique manifestation of general relativity’s nonlinear structure. This effect, first derived in full by Christodoulou [36], remains one of the most profound consequences of gravitational wave theory, highlighting how spacetime curvature can be permanently altered by the radiation it carries.

Current PTA searches for null memory largely rely on a signal model in which the memory-induced metric change is treated as an instantaneous, steplike shift [26]. This model, from now on referred to as a *memory burst*, models the GW memory as a ramp in timing residuals that begins at the time

of the burst and persists thereafter. The justification for this approximation arises from the fact that the memory grows over a timescale of a few days to weeks for SMBHBs with total mass $M \sim 10^8 M_\odot - 10^{10} M_\odot$. Compared to the decade-long observing span and typical biweekly cadence of PTA experiments, this growth appears instantaneous, and the detailed time structure of the signal becomes unresolved. This approximation has been used in several pioneering studies [26,38,39]. The detectability of such bursts using the memory burst model with 20 simulated pulsars timed at 100 ns precision over 10 years is investigated in van Haasteren and Levin [26]. The authors concluded that an SMBHB merger with $M = 10^8 M_\odot$ at $z = 0.1$ could be detected with 2σ confidence, whereas $M = 10^{10} M_\odot$ would be detectable throughout the Universe. Searches for gravitational wave memory that have been carried out on real PTA datasets include the frequentist approach using the memory burst model on the first PPTA dataset [40] and the Bayesian approach by NANOGrav on their 11-, 12.5-, and 15-year datasets [41–43]. Although no definitive detections have been reported, the 15-year analysis showed a mild preference (Bayes factor ≈ 3) for a model including a memory burst.

While searches for memory bursts apply to multiple physical sources, there are compelling reasons to move beyond this approximation. First, the burst template is derived under the objective to search for memory alone. This overlooks a key point: null memory arises as a consequence of the changing quadrupole moment brought out by the oscillatory gravitational waveform. It is not a stand alone phenomenon, but rather a byproduct of the same quadrupolar dynamics that generate the inspiral, merger, and ringdown phases. These oscillatory components, particularly for nearby or massive binaries, may themselves imprint detectable signatures in PTA timing residuals. Second, the burst model’s extreme simplicity makes it vulnerable to unmodeled noise.

In this work, we introduce the first complete signal model for SMBHB mergers that uses a full inspiral-merger-ringdown (IMR) numerical relativity surrogate which also includes null memory. This goes beyond the post-Newtonian, inspiral-only waveforms used in prior PTA continuous-wave searches. Using a high-fidelity numerical relativity (NR) waveform [44], we accurately model the full time-domain strain, including memory.

By simulating this signal in PTA datasets, we demonstrate how incorporating realistic waveform features enhances detection prospects and improves parameter estimation. Relying solely on a generic burst model discards the rich astrophysical information encoded in the full waveform. In contrast, a physically motivated model enables direct inference of key source properties such as chirp mass, mass ratio, luminosity distance, and inclination from the data.

The rest of the paper is organized as follows. In Sec. II, we outline the methodology of our search and introduce the

signal model and quantities we can infer. In Sec. III, we present our results on simulated datasets, sensitivity of PTA to SMBHB mergers and limitations of the memory burst model. In Sec. IV, we discuss improvements to the memory burst model, the new window opened onto SMBHBs for PTAs, future prospects of the SMBHB merger model and its potential for enabling EM follow-ups. We summarize the conclusions for the complete signal model in Sec. V.

II. METHODOLOGY

A. Memory signal from a binary merger

In this subsection, we provide background information about the calculation of gravitational wave memory strain. This information is useful for understanding the relation of the gravitational wave memory signal to properties of SMBHB, as well as to motivate our methodology.

Expressing the gravitational waveform time series $h_+ + ih_\times$ as the sum of the spherical harmonics with spin weight of -2 , the leading contribution to null displacement memory from quasicircular binaries arises from the amplitude of the dominant h_{22} mode of the oscillatory component of the signal [25]. Assuming a nonprecessing binary with component spins aligned with the orbital angular momentum, the memory can be computed analytically by integrating the square of the time derivative of this mode as

$$h_+^{(\text{mem})} \approx \frac{R}{192\pi} s_i^2 (17 + c_i^2) \int_{-\infty}^{T_R} |\dot{h}_{22}|^2 dt, \quad (2.1)$$

where R is the distance from the source to the observer, ι is the angle between the line of sight and the orbital angular momentum of the binary system, $c_i \equiv \cos \iota$, $s_i \equiv \sin \iota$, and T_R is the retarded time.

For compact binaries including SMBHBs, Eq. (2.1) evolves into a minimal waveform model (MWM) expression as explained by Favata [37] as

$$h_{\text{MWM}}^{(\text{mem})}(T) = \frac{8\pi M}{r(T)} \Theta(-T) + \Theta(T) \left\{ \frac{8\pi M}{r_m} + \frac{1}{\eta M} \times \sum_{n,n'=0}^{n_{\text{max}}} \frac{\sigma_{22n} \sigma_{22n'}^* A_{22n} A_{22n'}^*}{\sigma_{22n} + \sigma_{22n'}^*} [1 - e^{-(\sigma_{22n} + \sigma_{22n'}^*)T}] \right\}, \quad (2.2)$$

where $\Theta(t)$ is the Heaviside function; M the total mass; $r(T)$ the binary separation at time T ; η the symmetric mass ratio; A_{22n} the amplitudes of the quasinormal modes (QNMs) of the dominant $l = 2$, $m = 2$ GW mode during ringdown; σ_{22n} the complex QNM frequencies; and $r_m \equiv r(T = 0)$ the orbital separation at the end of the inspiral, where the inspiral and ringdown descriptions are matched (refer to [37] for more details). This resulting expression captures the gradual accumulation of memory during the inspiral and merger, analytically.

Favata also developed a more relativistic effective one body (EOB) waveform model calibrated to NR simulations shown in Fig. 1 of [37]. Although the EOB framework offers improvements over the MWM, its predictions still deviate from full NR-based calculations. For instance, the final memory amplitude from a hybrid PN/NR waveform was found to differ by 27% compared to the EOB estimate [37].

PTA searches have often relied on a much simplified representation of memory—a step function approximation in the metric perturbation. In the limit where the null memory saturates rapidly compared to the sampling interval, the detailed expression in Eq. (2.2) is approximated as a sudden change in the metric traveling through space [26],

$$h_+(\vec{r}, t) = h_0 \times \Theta[(t - t_0) - \vec{n} \cdot \vec{r}], \quad (2.3)$$

where h_0 is the amplitude of the step, t_0 is the time at which the step passes the Earth, $\vec{r}$ is the position vector of the observer, and $\vec{n}$ is the unit vector in the direction of wave propagation.

Even though the above approximations are well motivated and sound, they fall short in capturing the full physical accuracy of the SMBHB signal. As noted by [37], the step function and the EOB waveform model tend to overestimate the final memory saturation value compared to NR waveforms. Thus, to be able to confidently detect SMBHB mergers and perform a robust parameter estimation, full NR-informed waveforms that capture both the oscillatory and nonoscillatory contributions are necessary.

To address the above, we adopt the state-of-the-art numerical relativity surrogate waveform model NRHybSur3dq8_CCE of Yoo *et al.* [44]. It is built using the Cauchy-characteristic evolution (CCE), a method that evolves the gravitational field fully to future null infinity, thereby avoiding gauge ambiguities and capturing gravitational memory effects that are lost in more common extrapolation-based waveform extraction [45–47]. The model includes both the oscillatory (IMR) and null memory components of the strain, thereby providing a complete description of the expected waveform from SMBHBs.

In Yoo *et al.* [44], the CCE-extracted NR waveforms are first placed into the same asymptotic frame used by PN theory, known as the Bondi-Metzner-Sachs frame. This step removes residual gauge differences between the NR and PN descriptions. The early inspiral is modeled using analytic waveforms based on PN expansions informed by EOB dynamics. This single analytic inspiral model is then smoothly hybridized with the NR waveform, producing a continuous signal spanning inspiral, merger, and ringdown.

The model is parametrized in terms of chirp mass $\mathcal{M}$, mass ratio q , effective and antisymmetric spin χ_{eff} , luminosity distance D_L , inclination, polarization and sky location of the source. The NRHybSur3dq8_CCE model accurately

reproduces the CCE-extracted NR waveforms used for training, as demonstrated in Ref. [44]. To construct the GW signal model as observed by a PTA, we compute the two strain polarization components, $h_+(t)$ and $h_\times(t)$ using the NRHybSur3dq8_CCE waveform. In this analysis we include the oscillatory modes $(l, m) = (2 \pm 2), (2 \pm 1), (3 \pm 3)$ which describe the IMR part of the signal, together with the non-oscillatory $(2, 0)$ mode, which carries the gravitational-wave memory contribution (present only in the $h_+(t)$ polarization).

B. Projection in pulsar timing arrays

The response of a PTA to the shift in pulsar rotational frequency ν due to gravitational wave strain h with $+$ and $\times$ polarizations is given by van Haasteren and Levin [26]:

$$\frac{\delta\nu(t)}{\nu} = F(\hat{\Omega}, \hat{p}, \psi) \times [h_{+, \times}(t) - h_{+, \times}(t - \vec{n} \cdot \vec{r})], \quad (2.4)$$

where $F(\hat{\Omega}, \hat{p}, \psi)$ is the geometric projection factor which depends on $\hat{p}$, the unit vector pointing from the Earth to the pulsar, $\hat{\Omega}$, the unit vector pointing from the GW source to the Solar System Barycenter, and ψ the gravitational wave polarization angle. The angle between the source and pulsar directions is defined as $\cos \mu = -\hat{\Omega} \cdot \hat{p}$. The first strain term in Eq. (2.4) is the “Earth term,” and the second is the “pulsar term.” The Earth term is a coherent signal that appears simultaneously across all pulsars. The PTA response corresponding to the rotational frequency change is obtained by integrating Eq. (2.4):

$$\delta t(t) = \int_{t_0}^t \frac{\delta\nu(t')}{\nu} dt'. \quad (2.5)$$

It is also convenient to rewrite the Earth term of dimensionless gravitational wave strain $h(t)$ seen in Eq. (2.4) in the matrix form:

$$h(t) = \begin{bmatrix} F_+ & F_\times \end{bmatrix} \begin{bmatrix} \cos 2\psi & -\sin 2\psi \\ \sin 2\psi & \cos 2\psi \end{bmatrix} \begin{bmatrix} h_+(t) \\ h_\times(t) \end{bmatrix}, \quad (2.6)$$

where ψ is the polarization angle of the source and $h_{+, \times}$ denote the two independent polarization modes of the strain in the transverse-traceless gauge.

The functions $F_{+, \times}$ are the antenna pattern functions defined as

$$F_+(\hat{\Omega}) = \frac{1}{2} \frac{(\hat{m} \cdot \hat{p})^2 - (\hat{n} \cdot \hat{p})^2}{1 + \hat{\Omega} \cdot \hat{p}} \quad (2.7)$$

$$F_\times(\hat{\Omega}) = \frac{(\hat{m} \cdot \hat{p})(\hat{n} \cdot \hat{p})}{1 + \hat{\Omega} \cdot \hat{p}}; \quad (2.8)$$

$\hat{m}$ and $\hat{n}$ are orthonormal basis vectors spanning the plane transverse to the GW propagation direction $\hat{\Omega}$, defined as [48]

$$\hat{m} = (\sin \phi, -\cos \phi, 0) \quad (2.9)$$

$$\hat{n} = (-\cos \theta \cos \phi, -\cos \theta \sin \phi, \sin \theta), \quad (2.10)$$

with (θ, ϕ) being the sky location of the GW source.

The first two matrices in Eq. (2.6) are the expansion of $F(\hat{\Omega}, \hat{p}, \psi)$ in Eq. (2.4). The response on a pulsar $\delta t(t)$ is then expressed as

$$\delta t(t) = \begin{bmatrix} F_+ & F_\times \end{bmatrix} \begin{bmatrix} \cos 2\psi & -\sin 2\psi \\ \sin 2\psi & \cos 2\psi \end{bmatrix} \begin{bmatrix} \delta t_+(t) \\ \delta t_\times(t) \end{bmatrix}, \quad (2.11)$$

where functions $\delta t_+(t)$ and $\delta t_\times(t)$ represent the time-integrated gravitational wave strain as in Eq. (2.5):

$$\delta t_{+, \times}(t) = \int_{t_0}^t h_{+, \times}(t') dt'. \quad (2.12)$$

This formalism applies to any gravitational wave model, whether it describes a continuous gravitational wave or a memory burst. In this work we model only the Earth term, which is coherent across pulsars and provides the correlated signature used for detection and sky localization. The pulsar term depends on the uncertain pulsar distance, and for gravitational wave memory it is generally displaced by $L(1 + \hat{\Omega} \cdot \hat{p})$ placing the associated step outside a typical decade-long observing span for kpc-distance pulsars. When the pulsar term occurs within the data span, it yields at most a single step (jump) in the GW strain, localized to an individual pulsar, which can be difficult to distinguish from intrinsic pulsar noise processes such as pulsar glitches [49].

After computing the GW strain polarizations, $h_+(t)$ and $h_\times(t)$ using the surrogate waveform model, the timing residuals are then calculated using the formalism in Eq. (2.11). It is demonstrated in Fig. 1. Figure 1(a) shows the resulting strain waveform for various mode combinations: the dominant oscillatory modes without memory [in which the strain $h(t)$ decays to zero after the merger], the memory only $(2, 0)$ mode, and the full waveform including both the IMR and memory contributions. The corresponding response in the timing residuals for an arbitrary pulsar located at a right ascension (ra) of 258.4564° and a declination (dec) of 7.7937° is shown in Fig. 1(b). The SMBHB source is at ra of 0° and dec of 0° . The observation spans from MJD 54000 to 59000, covering approximately 13 years, with the merger occurring at MJD 58000. Figure 1(c) shows the postfit residuals obtained after the subtraction of the best-fit pulsar spin frequency and its first derivative.

To construct the signal model for use in PTA analysis, we generate $h(t)$ over an extended duration of 30 years, which is larger than the observation spans of the analyzed data. This ensures that the full signal, including the premerger oscillations and memory buildup, is captured regardless of

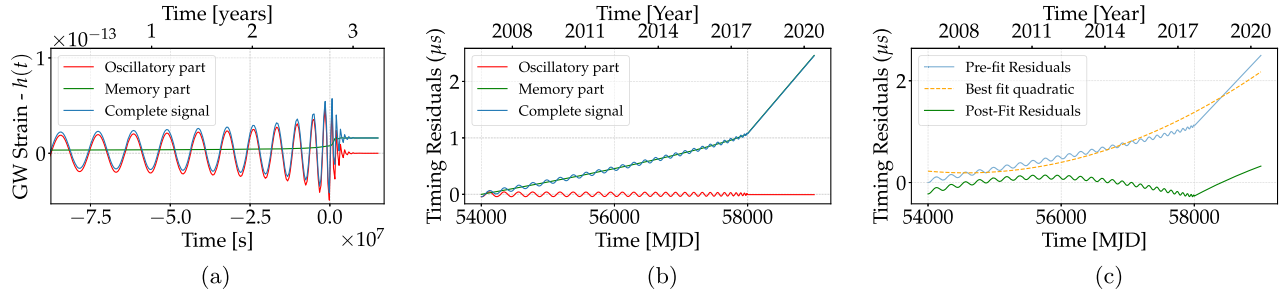


FIG. 1. Gravitational wave strain and timing residuals from a merger of a nonspinning SMBHB with parameters $\mathcal{M}_c = 10^{10} M_\odot$, $q = 1$, and $D_L = 1000$ Mpc. (a) GW strain waveform $h(t)$ for three different combinations of radiative modes: $(l, m) = (2 \pm 2)$, (2 ± 1) , (3 ± 3) without memory (red), the null memory-only $(2, 0)$ mode (green), and the complete signal including both oscillatory and memory contributions (blue). (b) The corresponding response induced by the GW source located at $(\text{ra}, \text{dec}) = (0^\circ, 0^\circ)$ on a pulsar at $(\text{ra}, \text{dec}) = (258.4564^\circ, 7.7937^\circ)$. Colors match those in (a), and the merger occurs at MJD 58000. (c) Postfit residuals after subtraction of the best-fit pulsar spin frequency and its derivatives. The full signal is shown in blue, the linear and quadratic spin-down model in orange, and the resulting postfit residuals in green.

when the binary coalescence occurs relative to the observation window. Such extrapolation is particularly important when searching for SMBHBs whose merger might occur outside the current PTA timespan. The waveform is computed starting from a frequency determined by the total binary mass and chosen duration (30 years before merger). The resulting waveform is interpolated to match the sampling time of the PTA data. To get the timing residual, we integrate the strain using a cumulative trapezoidal integration of $h(t)$. Postmerger, the memory component is extended as a constant offset based on the final saturation value of $h(t)$, similar to the memory burst model. This allows the signal to smoothly continue after the merger epoch. The implementation of the waveform generation and timing residual projection routines used in this work is publicly available in [50].

C. Data analysis

Throughout this study, we simulate pulse arrival time observations with an uncertainty on pulse arrival times of $\sigma_{\text{TOA}} = 100$ ns. Such low uncertainties are already achieved, and in some cases even surpassed, for several well-timed millisecond pulsars [51]. We simulate pulse arrival time observations at 1400 MHz, and, for simplicity, do not include radio-frequency-dependent noise in the simulations. In real PTA data, frequency dependent timing effects such as dispersion measure variations due to the interstellar medium and solar wind are modeled using wideband or multifrequency observations and corresponding chromatic noise terms in the timing model. When modeled consistently, we expect to achieve similar results. For simplicity, we also assume that the data are obtained using only one backend-receiver combination. Our simulations contain a single SMBHB merger signal, combined with a gravitational wave background, temporally uncorrelated noise referred to as “white,” and temporally correlated noise referred to as “red.” In this section, we provide the details of our analysis and noise models.

1. Pulsar noise

It is convenient to express pulsar noise in terms of the power spectral density of residuals, $P(f)$, where f is the Fourier frequency of the timing residuals. It also corresponds to the frequency of gravitational wave signals. White noise (WN) corresponds to the constant power spectral density across frequency, and it is assumed to be known in our simulations. Red noise (RN) is more prominent toward the lowest accessible frequency given by the inverse of the observation span, T_{obs} . We assume the power-law power spectral density of red noise [52–54],

$$P(f_j) = \frac{A^2}{12\pi^2} \left(\frac{f_j}{f_{\text{yr}}} \right)^{-\gamma} \text{yr}^3, \quad (2.13)$$

where the amplitude A is in the units of gravitational wave strain at the reference frequency $f_{\text{yr}} = 1, \text{yr}^{-1}$ and γ is the spectral index. We simulate and model red noise as a Gaussian process using a Fourier basis. In particular, the time series is a sum over N_f sine and cosine terms:

$$r_{\text{red}}(t) = \sum_{j=1}^{N_f} [a_j \sin(2\pi f_j t) + b_j \cos(2\pi f_j t)]. \quad (2.14)$$

Fourier amplitudes (a_j, b_j) are Gaussian random variables with zero mean and the standard deviation determined by $\sqrt{P(f)/T_{\text{obs}}}$, an approximation that has been shown to accurately reproduce the target Gaussian process [55–58]. In simulated PTA data with red noise, the hyperparameters $(A_{\text{RN}}, \gamma_{\text{RN}})$ for intrinsic per pulsar red noise are randomly drawn from uniform distributions over the ranges $[-15, -14]$ and $[1, 4]$, respectively.

2. Gravitational wave background

In some simulations, we test the robustness of our parameter estimation for SMBHB mergers to the presence

of an additional gravitational wave signal, namely, the stochastic GWB. It is modeled as a Gaussian process with the same power spectrum in all pulsars and additional interpulsar correlations which follow the Hellings-Downs function of an isotropic and unpolarized background [59],

$$\Gamma_{ab} = \frac{1}{2}\delta_{ab} + \frac{1}{2} - \frac{x_{ab}}{4} + \frac{3}{2}x_{ab} \ln x_{ab}, \quad (2.15)$$

where $x_{ab} = (1 - \cos \zeta_{ab})/2$, δ_{ab} is the Kronecker delta function, and ζ_{ab} is the sky separation angle for a given pair of pulsars. The power spectral density is

$$P_{\text{GWB}}(f) = \frac{A_{\text{GWB}}^2}{12\pi^2} \left(\frac{f}{f_{\text{yr}}} \right)^{-\gamma_{\text{GWB}}} \text{yr}^3. \quad (2.16)$$

Some of our simulations contain a GWB with $A_{\text{GWB}} = 10^{-15}$ and $\gamma = 13/3$. This choice of a spectral index corresponds to the GWB from a population of adiabatically inspiraling SMBHBs in circular orbits. For the purpose of detecting SMBHB mergers, GWB is a source of pulsar-correlated noise.

3. Likelihood

In this work, we perform an analysis of simulated PTA data using Bayesian inference. We construct a multivariate Gaussian likelihood. The likelihood of the vector of timing residuals $\delta \mathbf{t}$ is a multivariate Gaussian function [52,60,61],

$$\mathcal{L}(\delta \mathbf{t} | \boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi \mathbf{C})}} \exp \left[-\frac{1}{2} (\delta \mathbf{t} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\delta \mathbf{t} - \boldsymbol{\mu}) \right], \quad (2.17)$$

where $\boldsymbol{\mu}$ represents the time series model for an SMBHB merger with null memory. Our model parameters are $\boldsymbol{\theta}$.

The contribution of stochastic processes that include noise and the GWB is encoded in the covariance matrix

$$\mathbf{C} = \mathbf{N} + \mathbf{T} \mathbf{B} \mathbf{T}^T. \quad (2.18)$$

The diagonal white noise matrix $\mathbf{N}$ has elements σ . We simulate the values $\sigma^2 = \sigma_{\text{TOA}}^2$, and we assume them as fixed parameters. Component $\mathbf{T}$ is the design matrix, and $\mathbf{B}$ is the prior matrix. $\mathbf{T} = [\mathbf{M}, \mathbf{F}]$, where $\mathbf{M}$ is a matrix of partial derivatives of the TOAs with respect to each timing model parameter, and $\mathbf{F}$ is the Fourier design matrix consisting of Fourier basis functions that represent long term, time-correlated processes such as red noise and the gravitational wave background. Each component in $\mathbf{T}$ has a corresponding vector of coefficients, and together these coefficients form a combined vector $\mathbf{b} = [\boldsymbol{\epsilon}, \mathbf{a}]^T$ where $\boldsymbol{\epsilon}$ contains timing model corrections and $\mathbf{a}$ contains the sine and cosine amplitudes of the Fourier representation of correlated signals. These coefficients are treated as

nuisance parameters and are marginalized over in the likelihood. To perform this marginalization efficiently, each group of coefficients is assigned a Gaussian prior with a block-diagonal covariance matrix $\text{diag}(\mathbf{B}) = [\boldsymbol{\xi}, \boldsymbol{\varphi}]$, where $\boldsymbol{\xi}$ corresponds to the prior on the timing model parameters which is a broad diagonal prior (values of 10^{40}) and $\boldsymbol{\varphi}$ accounts for temporally correlated processes such as intrinsic spin noise and a stochastic gravitational wave background. The covariance matrix of $\mathbf{B}$ has elements

$$\Phi_{(ai),(bj)} = P_{ai} \delta_{ab} \delta_{ij} + \Gamma_{ab} P_i \delta_{ij}, \quad (2.19)$$

where a and b are pulsar indices, and i and j are frequency indices. P_{ai} is the power spectral density (PSD) of the red noise of pulsar a at a frequency bin i [Eq. (2.13)], P_i is the PSD of the gravitational wave background at the frequency i [Eq. (2.16)], and Γ_{ab} is the overlap reduction function that determines the degree of this correlation between pulsars a and b .

The posterior distribution over parameters is obtained using Bayes' theorem, which reads as

$$\mathcal{P}(\boldsymbol{\theta} | \delta \mathbf{t}) = \frac{\mathcal{L}(\delta \mathbf{t} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\mathcal{Z}}, \quad (2.20)$$

where $\pi(\boldsymbol{\theta})$ is the prior and $\mathcal{Z}$ is the marginal likelihood (evidence) such that

$$\mathcal{Z} = \int \mathcal{L}(\delta \mathbf{t} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (2.21)$$

Model comparison is performed using the Bayes factor, defined as the ratio of evidence of models. In cases where models are assigned equal prior odds, this is equivalent to the posterior odds ratio. In this work we do not evaluate the evidences $\mathcal{Z}$ via direct numerical integration. Instead, Bayes factors are computed using the product space (hypermodel) method [62] implemented in ENTERPRISE_EXTENSIONS [63], in which a model index is sampled jointly with the model parameters and posterior model probabilities are estimated from the relative fraction of samples assigned to each model.

III. RESULTS

A. Simulated datasets with SMBHB merger

We simulate a pulsar timing array consisting of 25 pulsars with 100-ns precision distributed uniformly across the sky. The observation spans over 13 years, from MJD 53000 to 58600, with 390 TOAs and observation gaps ranging from a few days to two weeks. A large, well-distributed pulsar array improves sky localization of the source and enhances the response to the GW signal. A noise-free set of TOA measurements is first generated and used as the baseline for constructing multiple datasets for testing signal recovery.

We simulate two different datasets containing signals from individual SMBHB mergers with observer-frame

chirp masses $\mathcal{M}_c = 10^8$ and $10^{10} M_\odot$ at luminosity distances of 3 and 100 Mpc, respectively. Both binaries are assumed to be equal in mass, circular and nonspinning. They are located at a polar and azimuthal angle of $(\theta, \phi) = (\pi/3, 2\pi/3)$ radians, with a wave polarization of 0 deg and an inclination of 90° , and a coalescence time of MJD 57000. The SMBHB merger model includes a continuous inspiral, which introduces quasisinusoidal features that can lead to degeneracies across parameters such as sky location, chirp mass, luminosity distance, and merger time. These combinations of degenerate parameters, in turn, can produce multiple local maxima in the likelihood. This situation is analogous to challenges encountered in continuous gravitational wave searches, where multimodal posteriors have been noted and addressed using specialized sampling algorithms [64]. To address this multimodal likelihood surface, we use a parallel tempering Markov chain Monte Carlo (PTMCMC) sampler [65].

In PTMCMC, multiple chains are run simultaneously at different temperatures, T characterized by the inverse temperature parameter $\beta = 1/T$. The posterior distribution at a given temperature is

$$P_\beta(\theta|\delta t) \propto \mathcal{L}(\delta t|\theta)^\beta \pi(\theta), \quad (3.1)$$

where $\mathcal{L}(\delta t|\theta)$ is the likelihood and $\pi(\theta)$ is the prior. As β goes from 0 to 1, we effectively move from the prior-only distribution ($\beta = 0$) to the posterior ($\beta = 1$). $\beta < 1$ flattens the posterior, reducing the relative heights of peaks and allowing chains to explore the parameter space more freely. This exchange between temperatures improves mixing and convergence. We use 20 geometrically spaced temperatures, which we find sufficient to recover posteriors robustly for signals with both oscillatory and memory components.

To test the robustness of parameter estimation to different noise conditions, we construct two types of simulated datasets:

- (i) Simulation `wn+rn+mem`: white noise, intrinsic per pulsar red noise whose hyperparameters ($A_{\text{RN}}, \gamma_{\text{RN}}$) are drawn from the uniform distributions $[-15, -14]$ and $[1, 4]$, respectively, and SMBHB merger.
- (ii) Simulation `wn+rn+gwb+mem`: white noise, intrinsic per pulsar red noise, and a stochastic gravitational wave background with Hellings-Downs correlations.

Each SMBHB merger is simulated on both noise realizations, producing a total of four datasets spanning both the masses and corresponding distances. After simulating the datasets, we perform parameter estimation. The parameters and priors used for this analysis are shown in Table I. We fix the white noise parameters to their simulated values to reduce dimensionality and focus the sampling on the recovery of the signal and red noise components. We adopt log-uniform priors on the chirp mass and luminosity distance. These choices correspond to scale-invariant,

TABLE I. Details of the prior distributions of various parameters used in the memory signal recovery analysis.

Parameter	Prior	Description
Intrinsic pulsar RN		
$\log_{10} A_{\text{RN}}$	$\mathcal{U}(-17, -12)$	Log-10 amplitude of red noise
γ_{RN}	$\mathcal{U}(1, 7)$	Spectral index of red noise
GWB		
$\log_{10} A_{\text{GWB}}$	$\mathcal{U}(-20, -6)$	Log-10 amplitude of GWB
γ_{GWB}	$\mathcal{U}(1, 7)$	Spectral index of GWB
SMBHB merger with null memory		
$\log_{10} \mathcal{M} [M_\odot]$	$\mathcal{U}(8, 12)$	Observer-frame chirp mass
$\log_{10} D_L [\text{Mpc}]$	$\mathcal{U}(0, 5)$	Luminosity distance
q	$\mathcal{U}(1, 7)$	Mass ratio
ψ	$\mathcal{U}(0, \pi)$	Polarization
$\cos \theta$	$\mathcal{U}(0, \pi)$	Cosine of polar angle
ϕ	$\mathcal{U}(0, 2\pi)$	Azimuthal angle
ι	$\mathcal{U}(0, \pi)$	Inclination angle
$t_{\text{merge}} [\text{MJD}]$	$\mathcal{U}(53000, 58600)$	Time of merger

weakly informative priors appropriate for parameters spanning multiple orders of magnitude. Our goal is to demonstrate the recovery performance of the waveform model without imposing strong astrophysical assumptions on the underlying population.

Figure 2 shows the joint posterior distributions of chirp mass and luminosity distance for the four different datasets. For the lower-mass system shown in Fig. 2(a), analyzed without a simulated stochastic gravitational-wave background, both the chirp mass and luminosity distance are recovered with the injected values lying within the 68% marginal credible intervals. When a stochastic background is included, the marginal posterior for the chirp mass shifts toward higher values, and the maximum-*a-posteriori* (MAP) estimate is displaced from the simulated mass. However, the injected parameters remain consistent with the data, lying within the joint $2 - 3\sigma$ credible region of the posterior. This behavior is a consequence of the strong mass-distance degeneracy characteristic of memory-dominated PTA signals since the GW memory strain $h_{\text{mem}} = \mathcal{M}_c/D_L$. The simulated chirp mass in Fig. 2(a) lies close to the lower prior bound, which was chosen based on PTA sensitivity considerations. Systems with smaller masses would produce memory signals that are not expected to be identifiable or constrained with the current PTA sensitivity. For this simulation, the memory strain amplitude is comparatively small, $h_{\text{mem}} \approx 4 \times 10^{-14}$, and dominated by low-frequency power. As a result, the inclusion of a correlated low-frequency process such as a stochastic gravitational-wave background increases the effective covariance at long timescales, broadening the posterior and allowing the MAP estimate to shift along the mass-distance degeneracy without excluding the true parameters. In contrast, for the higher-mass system shown

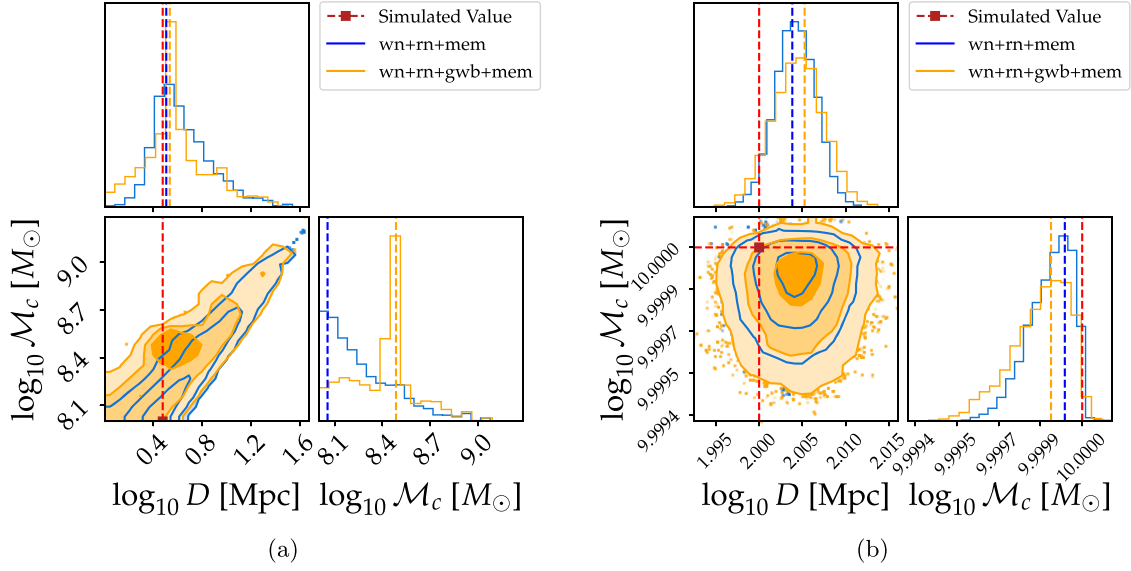


FIG. 2. The figure illustrates the posterior distributions from our simulation recovery studies with 25 pulsars. Simulation studies were conducted with two types of datasets, $wn+rn+mem$ and $wn+rn+gwb+mem$, and the contours are shown in blue and orange respectively. Three subfigures show a different SMBHB merger. (a) A merger with $\mathcal{M}_c = 10^8 M_\odot$ and $D_L = 3$ Mpc and (b) A merger with $\mathcal{M}_c = 10^{10} M_\odot$ and $D_L = 100$ Mpc. MAP values are indicated by dotted lines in the respective contour colors (blue and orange), and the true simulated values are shown as red dotted lines.

in Fig. 2(b) ($10^{10} M_\odot$), the projected memory amplitude is significantly larger ($h_{mem} \approx 1 \times 10^{-13}$), leading to tighter constraints. In this case, both the chirp mass and luminosity distance are recovered within the 68% marginal credible intervals in both noise configurations, and the MAP estimates remain close to the injected values. The full posterior distributions for all SMBHB merger parameters are shown in the Appendix Fig. 7 and 8, for the $\mathcal{M}_c =$

$10^8 M_\odot$ at $D_L = 3$ Mpc binary and in Fig. 9 and 10 for the $\mathcal{M}_c = 10^{10} M_\odot$ at $D_L = 100$ Mpc binary.

To illustrate the sky localization performance of our simulated SMBHB mergers, we projected the posterior samples onto HEALPix sky maps in equatorial coordinates (RA, Dec in degrees). This is shown in Fig. 3. The color scale in each map represents the relative posterior probability density, and the yellow cross marks the true

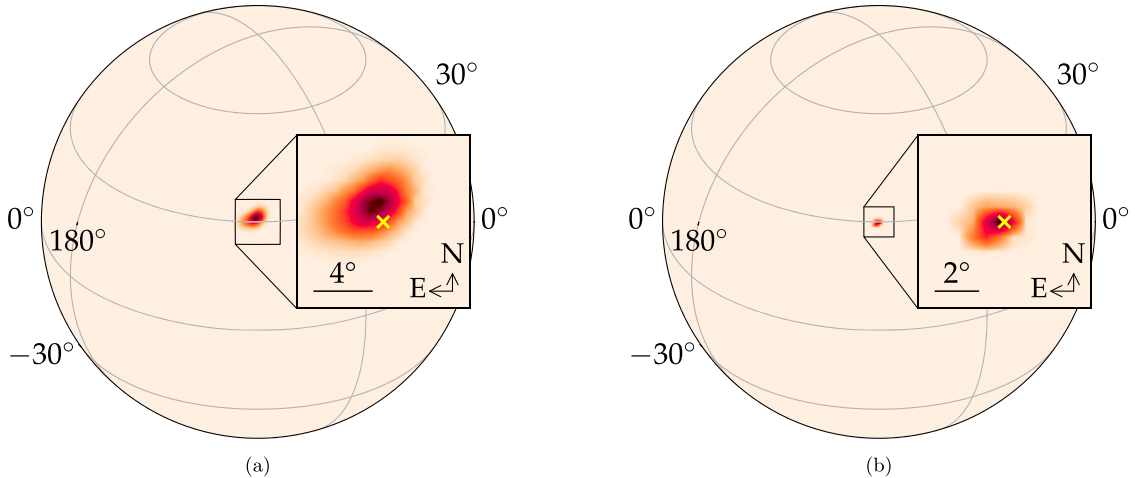


FIG. 3. Sky maps of the posterior distributions of simulated supermassive black hole binary mergers, shown in equatorial coordinates (RA, Dec in degrees). The color scale represents the relative posterior probability density derived from our HEALPix projections of the samples, while the yellow cross marks the true (simulated) sky location of the source. Each square projection shows an enlarged region around the source, with the scale bar indicating the angular width on the sky. The three subfigures correspond to mergers with increasing chirp mass and luminosity distance: (a) $\mathcal{M}_c = 10^8 M_\odot$, $D_L = 3$ Mpc; (b) $\mathcal{M}_c = 10^{10} M_\odot$, $D_L = 100$ Mpc. As the chirp mass increases, the posterior uncertainty in sky localization decreases, demonstrating improved measurement precision for more massive mergers.

(simulated) source location. Each map includes an enlarged inset that highlights the region around the source, with a scale bar indicating the angular width of the enlargement. As the chirp mass increases, the precision of sky localization improves noticeably. This trend reflects the stronger gravitational wave signal associated with more massive mergers, which allows the parameter estimation algorithms to constrain the source position more accurately. The insets also visually emphasize how the uncertainty region shrinks with mass, showing that even at larger distances, more massive binaries can be localized more precisely than lighter, closer systems.

Additionally, we perform a model selection using a product space method between the memory and the noise-only model to evaluate the detection confidence. In the case of both $wn+rn+mem$ and $wn+rn+gwb+mem$ datasets, all posterior samples favor the memory model, allowing us to place a conservative lower bound on the Bayes factor of $\ln \mathcal{B} > 10$. Throughout this work, model preference and detectability are quantified using Bayes factors rather than matched-filter signal-to-noise ratio.

B. Limitations of the memory burst approximation and its impact on observable signals

In Fig. 1 of Favata [37], a comparison is presented between the nonlinear memory waveform computed using an EOB and MWM. Favata reports that the EOB-based model underestimates the final memory strain offset by $\approx 27\%$ relative to the NR calculations, even after calibration. This discrepancy arises due to simplifications in the analytical model, particularly during the merger and ring-down phases, as well as the omission of higher-order radiative modes.

The analytic expression for memory strain used by Pshirkov *et al.* [39] is based on a further simplified approximation derived from the nonlinear memory formula

presented in Favata [37]. In particular, they approximate the radiated energy during the merger as a fixed mean value and assume an equal-mass, nonspinning binary on a circular orbit.

As a result, the predicted strain is systematically larger than what would be obtained from physically realistic SMBHB models. This leads to an overestimation of the memory amplitude or equivalently, an underestimation of the luminosity distance or overestimation of the source mass when interpreting the observed data using this model.

Due to these limitations, the memory burst model does not provide a reliable mapping between the observed signal and the underlying parameters of the SMBHB. Even if the final strain offset is calibrated correctly, the second limitation of the burst model, which is the absence of gradual memory growth, persists. This has important observational consequences, particularly in the context of PTA searches.

PTA analyses are not sensitive to the gravitational wave memory time series in full. Instead, they are sensitive to the *postfit residuals*, that is, the signal remaining after subtraction of the best-fit pulsar timing model. This subtraction typically includes a polynomial fit for the spin frequency and its time derivatives. Any signal that mimics these trends can be partially absorbed, reducing the amplitude of the observable signature.

The gravitational wave memory signal from a realistic SMBHB merger builds up gradually over the inspiral and coalescence, in contrast to the instantaneous step-function approximation assumed in the standard burst model. Even when both burst model and SMBHB merger model are calibrated to reach the same final strain offset, their evolution in time (particularly around the merger) leads to important differences in the observable residuals after subtraction of the pulsar timing model. For the comparison in Fig. 4, we adjust the final strain offset to be the same across both models for a SMBHB binary of $\mathcal{M}_c = 10^{10} M_\odot$

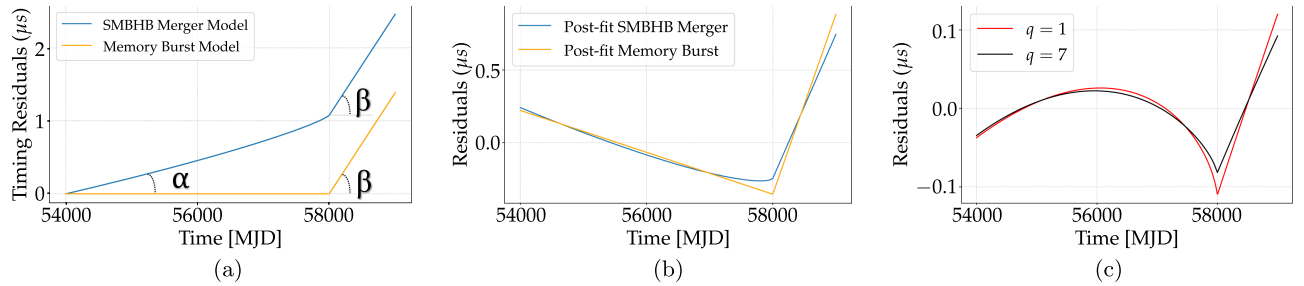


FIG. 4. Timing residuals from a pulsar located at $(ra, dec) = (258.4564^\circ, 7.7937^\circ)$ induced by a gravitational wave from a SMBHB merger with $\mathcal{M}_c = 10^{10} M_\odot$, $D_L = 100$ Mpc and $q = 1$ located at $(0^\circ, 0^\circ)$. (a) Prefit timing residuals for memory burst model (orange) and the SMBHB merger model (blue). Both models are adjusted to have the same final memory offset. The SMBHB model shows a gradual premerger rise at angle α , whereas the burst model shows an abrupt change at angle β . (b) Postfit residuals after subtraction of a linear spindown model. The observable residual bump for the SMBHB model (blue) is suppressed compared to the burst model (orange), due to the premerger curvature being partially absorbed by the fit. For the equal-offset case, the postfit RMS difference between the models is $0.399 \mu s$. (c) Difference in postfit timing residuals between the burst model and SMBHB model for varying mass ratios. The discrepancy is largest for the equal-mass case ($q = 1$) and decreases for more unequal-mass binaries. The RMS difference drops by $\sim 0.1 \mu s$ when the mass ratio increases from $q = 1$ to $q = 7$.

and $D_L = 100$ Mpc. This isolates the effect of the signal's shape rather than its amplitude.

Figure 4(a) compares the prefit residuals produced by the memory burst model (orange) and the full SMBHB merger model (blue) for the same simulated binary. Throughout this section, we treat the SMBHB waveform which includes the premerger memory buildup as the most accurate available representation of the true gravitational-wave memory signal. Both models exhibit a monotonic postmerger rise corresponding to the final memory offset, while the SMBHB waveform additionally shows a gradual premerger increase associated with the inspiral-driven buildup of memory. In contrast, the burst model approximates this growth as an instantaneous ramp at the merger epoch.

After subtracting the pulsar spin down and its derivatives, the postfit residual differs significantly between the two waveform models for the same SMBHB. This is depicted in Fig. 4(b). Since the SMBHB waveform contains a slow premerger rise extending over a long timescale, a significant fraction of its signal is absorbed by the timing-model fit, whereas the burst approximation concentrates the entire memory contribution near the merger time and is therefore less affected. As a result, even when both models are normalized to produce the same final memory offset, the observable postfit residuals differ.

To quantify this mismatch in the PTA observable, we compute the noise-agnostic root-mean-square (RMS) of the difference between the postfit residuals. For our fiducial source ($\mathcal{M}_c = 10^{10} M_\odot$, $D_L = 100$ Mpc, $q = 1$), enforcing the same final memory offset yields a postfit RMS difference of $0.399 \mu\text{s}$. Allowing the burst-model strain amplitude to vary in order to minimize this postfit RMS reduces the mismatch to $0.26 \mu\text{s}$. In this best-fit configuration, the computed burst model strain amplitude is larger by 11.8% compared to the amplitude from the SMBHB merger model. Interpreting this best-fit burst amplitude as a physical memory step leads to biased source parameters. Fixing the simulated luminosity distance and inferring the chirp mass from the adjusted burst amplitude [Eq. (3.2)] results in 37.5% discrepancy in the observer-frame chirp mass relative to the injected value; conversely, fixing the chirp mass and inferring the distance produces a 62.2% bias. Thus, even when optimally tuned to match the postfit residuals of the true SMBHB signal, the simplified burst approximation yields biased parameter estimates and leaves a residual mismatch that ultimately reduces sensitivity.

Figure 4(c) illustrates the difference between the postfit residuals of the two models. These differences are primarily driven by the gradual premerger memory buildup present in the SMBHB waveform but absent in the burst approximation, and therefore depend on the binary's mass, mass ratio, distance, and spin (we assume nonspinning binaries here). As the mass ratio becomes more unequal, the memory builds more steeply during the inspiral phase. We find that the mismatch is largest for equal-mass systems and

decreases for more unequal mass ratios. For example, the RMS difference between the residuals decreases by approximately $0.1 \mu\text{s}$ as the mass ratio increases from $q = 1$ to $q = 7$.

This bias is not a mere modeling artifact. It has real implications for data analysis. If a PTA search uses an idealized burst template while the true signal evolves gradually, as in realistic SMBHB mergers, the recovered signal may appear artificially strong. This leads to incorrectly inferred source parameters, such as an underestimated luminosity distance or overestimated strain. This highlights the importance of accurate waveform modeling in PTA-based memory searches.

C. Null hypothesis tests and upper limits

We simulate a dataset containing only pulsar-specific white and red noise to demonstrate our ability to correctly place upper limits and to ensure that we do not detect any spurious evidence for SMBHB mergers.

We use the same PTA configuration as in the strong signal simulation study, namely, a 25-pulsar array with uniform sky coverage, 13-year observing span but with 1000 TOAs and ~ 6 -day observation cadence. We then perform a model selection between the correct null hypothesis and the two signal hypotheses. Signal hypotheses additionally include the memory burst and the SMBHB merger signal, respectively. We find log Bayes factors of -0.375 for the memory burst model and 0.1 for the SMBHB merger model. These values indicate no significant evidence for a memory signal in either case, as expected.

Figure 5 summarizes the limits on the source frame chirp mass and the luminosity distance based on the SMBHB merger model and the memory burst model applied to the

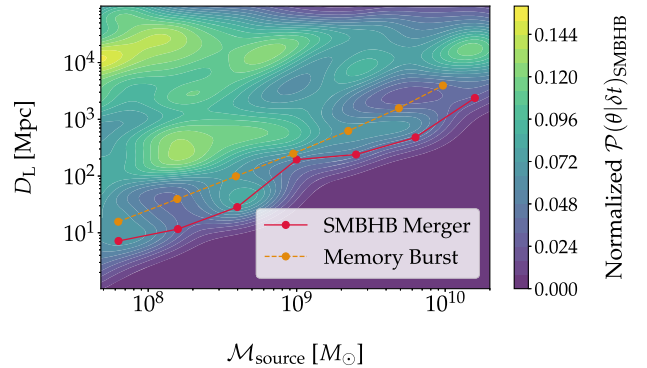


FIG. 5. Lower limits on luminosity distance as a function of chirp mass in the signal-free simulated data. The filled contours show the marginalized posterior density in the $\mathcal{M}$ - D_L plane using the full SMBHB merger waveform model. The red curve represents the exclusion boundary below which a source would have been detectable with 95% confidence. The orange curve shows the corresponding lower limit derived from the 95% upper limit on memory strain amplitude using the memory burst model, converted to distance via Eq. (3.2).

simulated data. For the SMBHB merger model, we find limits on both the chirp mass and luminosity distance. The filled contours represent the marginalized posterior density in the chirp mass-luminosity distance plane for the SMBHB model in the noise-only data. The red curve shows 95% lower limits on the luminosity distance for fixed bins in chirp mass, representing minimum distances to which our analysis would have been sensitive for a signal of a given chirp mass. For the memory burst model, we obtain a 95% upper limit on the strain amplitude of a gravitational-wave memory event. To translate this strain limit into a distance constraint, we adopt the analytic scaling of Pshirkov *et al.* [39], rewritten in terms of source-frame chirp mass and luminosity distance. For an equal-mass binary ($q = 1$) the memory strain can be expressed as

$$h_{\text{mem}} \approx 1.13 \times 10^{-15} \left(\frac{\mathcal{M}_{\text{source}}(1+z)}{10^8 M_{\odot}} \right) \left(\frac{1 \text{ Gpc}}{D_L} \right), \quad (3.2)$$

where D_L is the luminosity distance and z is the corresponding cosmological redshift assuming a flat Λ_{CDM} model. While the signal observed at Earth depends on the redshifted (observer-frame) chirp mass $\mathcal{M}_{\text{obs}} = \mathcal{M}_{\text{source}}(1+z)$, we report results in the source frame. This provides a straightforward mapping between the measured strain upper limit and a minimum allowed luminosity distance to a hypothetical source, assuming a mass ratio of 1. The orange curve shows the corresponding limit derived using the memory burst model via Eq. (3.2).

The difference in the exclusion boundaries between the two models reflects the modeling limitations discussed in the previous section. The memory burst model overestimates the final memory offset due to simplified assumptions such as EOB approximations and no gradual premerger buildup. These approximations cause the strain amplitude to appear larger than it would be in a realistic SMBHB merger scenario. As a result, the burst model places artificially stringent constraints on the luminosity distance, leading to higher D_L limits than the SMBHB merger model. This visual difference, shown in Fig. 5, illustrates that using the burst approximation can misrepresent the PTA's true sensitivity and potentially bias population-level interpretations based on upper limit studies.

IV. DISCUSSION

The method presented in this work represents a major step forward in the search for gravitational wave memory from SMBHBs using pulsar timing arrays. It provides several key advantages over earlier approaches that focused solely on memory burst detection. In what follows, we highlight the implications of our findings and outline the next steps.

a. Improvement upon the memory burst model. The SMBHB merger model reduces false positives and improves parameter estimation compared to the simplified memory burst model. Its detailed time structure also helps distinguish real signals from glitches or noise. The memory burst model, by assuming an instantaneous step and simplified energy estimates, tends to overestimate the strain amplitude of realistic binaries. This can lead to biased interpretations such as overestimating the mass of the binary and highlights the importance of using physically consistent waveforms in PTA memory searches to infer accurate astrophysical parameters from the observed signal.

b. Spinning SMBHBs. At present, our analysis is restricted to nonspinning, circular SMBHBs. The package NRHybSur3dq8_CCE supports aligned-spin binaries. Previous studies have shown that spin can significantly affect the memory amplitude. In particular, the final memory offset of strain for maximally aligned spins ($a_1 = a_2 = 1$) is by a factor of 2.5 larger than for a nonspinning binary Pollney and Reisswig [66]. This enhancement arises because higher-spin systems radiate more energy through gravitational waves, which in turn sources a larger null memory. Thus, considering spinning binaries may improve the odds of observing a merging SMBHB.

c. Eccentric SMBHBs. Eccentricity is another important factor not addressed in the present work. Favata [67] presents a derivation of the null memory from binaries on elliptical orbits and shows that the memory flux becomes sharply peaked near each pericenter passage as eccentricity increases. This differs from the smoother, continuous buildup of memory in circular binaries. While the total memory evolves toward the circular case as the orbit circularizes, systems that maintain eccentricity during their evolution may show a different memory accumulation profile. Figure 9 in Ref. [68] shows that a non-negligible fraction of SMBHBs may have significant eccentricities. Therefore, it is of interest to search for merging eccentric SMBHBs.

d. Prospects for electromagnetic counterparts enabled by the SMBHB merger model. EM counterparts to SMBHBs have been studied theoretically using both semi-analytic models and general relativistic simulations, predicting a range of possible signatures: precursors in accretion signatures, variable jets, circumbinary disk effects, or flares associated with the final coalescence [69,70]. These multiwavelength signatures offer rich insights into the binary environment, host galaxy, and physical parameters. When applied to future PTA experiments with increased sensitivity, the SMBHB merger model may allow EM follow-up and multimessenger observations. Current PTAs release data only once every few years, whereas electromagnetic follow-up of SMBHB mergers would require combining data and updating TOAs on timescales of months rather than years. This would allow the evolving significance of a candidate signal to be

monitored in near real time, enabling increased observing cadence and timely alerts to electromagnetic observatories as the source approaches merger. This may open up the possibility of detecting counterpart signatures such as disk shocks, jet reconfiguration, or transient EM flares, potentially observable across multiple bands [69,70]. It may be of interest to investigate this possibility further in future work.

e. New window onto SMBHBs. Our search for merging SMBHBs with the NR waveform model with null memory opens a new window onto SMBHBs for PTAs. The other two windows are the existing searches for stochastic background from SMBHBs and for continuous gravitational waves from SMBHB sources that stand out from the background.

We compare the parameter-space coverage of CGW searches and our merger search in Fig. 6.

Diagonal lines of equal time to merger are constructed based on the relation [72]

$$f^{-8/3}(t) = \frac{(8\pi)^{8/3}}{5} \left(\frac{GM}{c^3} \right)^{5/3} (t - t_c), \quad (4.1)$$

where G is Newton's constant, c is the speed of light, and t_c is the time of coalescence. This relation corresponds to the leading-order quadrupole approximation for quasicircular binaries; higher post-Newtonian corrections and orbital eccentricity would modify the detailed frequency evolution

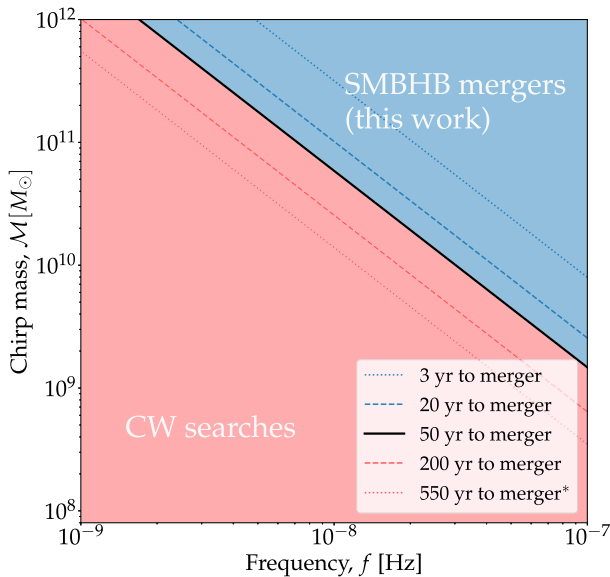


FIG. 6. $\mathcal{M}$ - f parameter space of (1) CGW searches for SMBHBs with PTAs and (2) the SMBHB merger searches that we introduce in this work. Diagonal lines indicate fixed times to the SMBHB merger. The line marked by an asterisk in the legend, corresponding to 550 years to merger, indicates a case where a signal will slowly drift out of the frequency bin on a timescale of PTA observations in the next decade [71].

but do not affect the qualitative parameter-space coverage illustrated in Fig. 6. Continuous gravitational wave searches typically assume a nonevolving frequency, which is valid up to about 3×10^{-8} Hz for a heavy SMBHB with $\mathcal{M} = 10^{10} M_\odot$. At higher frequencies, the signal will merge on a timescale comparable to that of the PTA observation. Moreover, even for an SMBHB that will merge in 550 years, illustrated by the bottom dotted line, the signal is expected to drift out of the frequency bin. This is shown in the targeted search for the SMBHB candidate 3C 66B [71]. If the candidate is a real SMBHB, it will also merge in about 550 years, and the signal may drift out of the frequency bin in the foreseeable future. In such regimes, conventional CGW searches lose sensitivity, whereas our method naturally captures both frequency evolution, the late inspiral, and the final merger. The gravitational-wave memory contribution is not strongly degenerate with CGWs, even though both are included in the model, as memory produces a monotonic, nonoscillatory signature in timing residuals, while CGWs generate quasisinusoidal signals that retain a distinct morphology after timing-model fitting. The presence of additional unmodeled CGW sources in the data would be effectively absorbed into the stochastic background, or be modeled explicitly if individually resolvable. In either case, such signals are expected to broaden posterior distributions rather than mimic or bias the memory parameters.

f. Impact of noise systematics and real PTA configurations. Chromatic noise processes such as dispersion measure variations, solar wind effects, and other frequency-dependent systematics are not degenerate with the gravitational wave memory signals, which produce an achromatic, nonoscillatory signature in timing residuals. When modeled consistently, such chromatic processes would not bias the recovered memory parameters. Similarly, nonideal PTA configurations with anisotropic sky distributions and pulsars spanning a wide range of timing precision are accommodated within the PTA likelihood framework and primarily affect sensitivity rather than qualitative inference. The feasibility of applying this method to real PTA data, including sub-banded TOAs, complex noise models, and large data volumes, has been demonstrated in our companion search for gravitational wave memory using EPTA and PPTA datasets with this waveform model [73]. While future improvements in computational efficiency and noise modeling will be important for next-generation datasets, these challenges do not limit the applicability of the method to current PTA observations.

V. CONCLUSIONS

In this work, we presented a physically complete waveform model for detecting gravitational-wave memory signals from SMBHBs in PTA data, extending beyond the conventional burst-with-memory search strategies. Our method models the full memory waveform inspired by

numerical relativity, allowing us to directly explore a broader astrophysical parameter space including binary mass, luminosity distance, and merger time. Unlike the simplified step-function memory model that tends to overestimate the final memory amplitude, our physically motivated waveforms allow more accurate signal recovery and parameter estimation. We showed the effectiveness of our pipeline in recovering memory signals under various noise realizations and configurations through a series of simulation studies across realistic datasets. Even in the presence of degeneracies and multimodal posteriors, the Bayesian framework correctly captures these correlations and allows statistically consistent recovery of the injected source parameters. Our analysis provides a foundation for more detailed and physically motivated searches for SMBHB memory in PTA data.

We made use of the following code. To generate the gravitational-wave strain signal from SMBHB mergers including the null memory component, we use the publicly available `NRHybSur3dq8_CCE` surrogate waveform model available as `GWSURROGATE` [74]. For simulating pulsar timing datasets with realistic observation spans, noise properties, and cadence, we use the `PTA_REPLICATOR` at [75] with an extension to include our SMBHB merger with memory waveform model. A tutorial and example scripts for simulating PTA datasets and recovering SMBHB merger signals with the waveform model used in this work are available at [76]. The Bayesian analysis is performed using the `ENTERPRISE` framework [77], the core data analysis software for pulsar timing arrays for computing

likelihoods and posteriors. We build the custom signal model representing the full SMBHB waveform within this framework. We employ `ENTERPRISE_WARP` at [78] which is a wrapper interface that facilitates streamlined integration between `ENTERPRISE` ([79]), `LIBSTEMPO`, `Tempo2`, `ENTERPRISE_EXTENSIONS` ([63]) and MCMC samplers. Posterior sampling for Bayesian inference is performed using the `PTMCMCSampler` package [65], which implements an efficient parallel tempering MCMC algorithm.

ACKNOWLEDGMENTS

We thank Abhimanyu Sushobhanan and Wang Wei Yu for insightful scientific discussions. This work is supported by the Max Planck Gesellschaft (MPG) and the ATLAS cluster computing team at AEI Hannover.

DATA AVAILABILITY

The simulated PTA datasets that support the findings of this article are openly available in [80].

APPENDIX: POSTERIOR DISTRIBUTIONS

The appendix presents the full posterior distributions for all simulation runs discussed in the main text, including the recovered sky locations, masses, distances, and other model parameters. These results provide a complete view of the parameter estimation performance of our SMBHB merger model across different simulated scenarios.

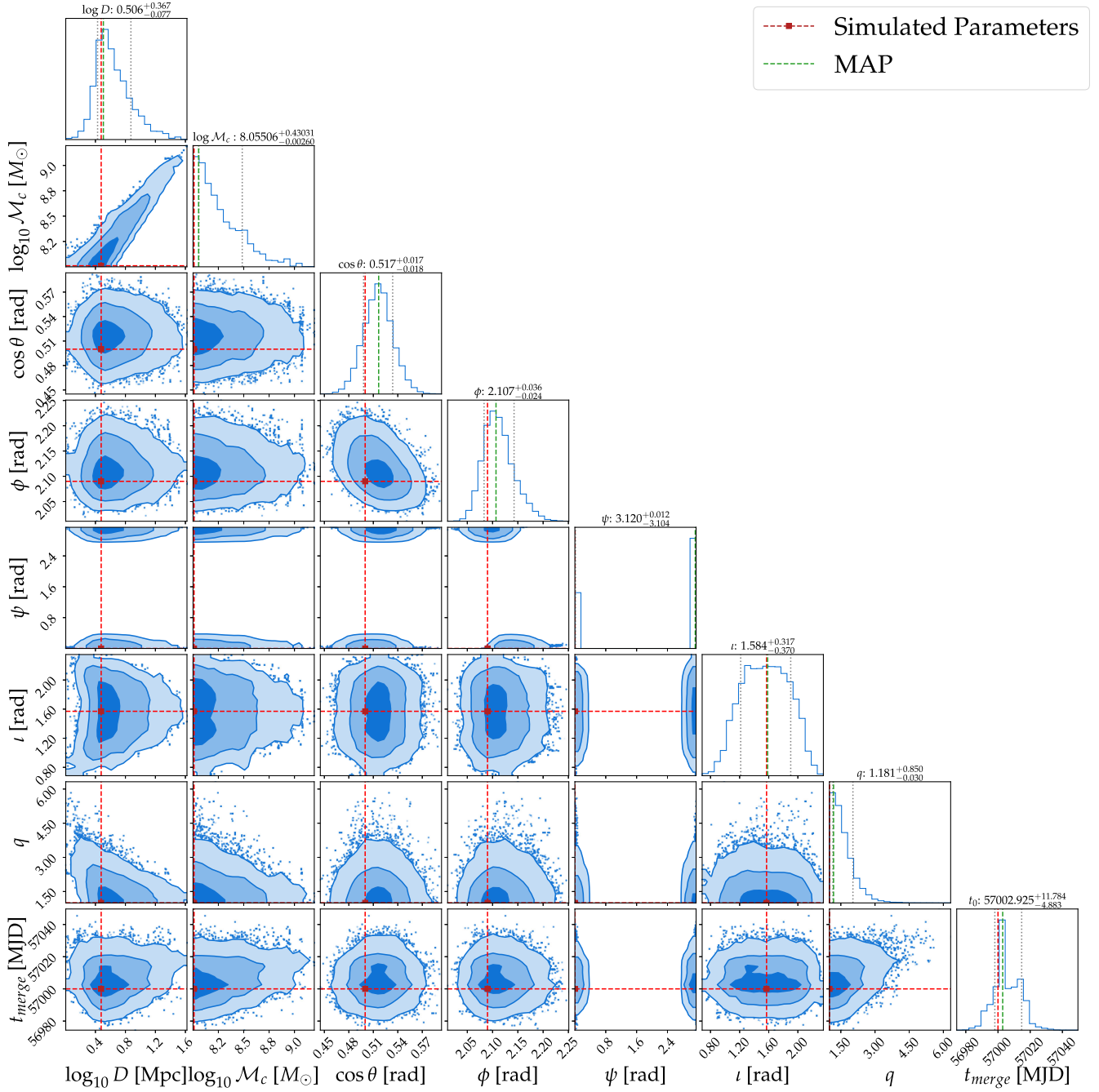


FIG. 7. Posterior distributions of SMBHB merger with null memory parameters from simulation studies on 25 pulsars. Parameters of the binary used for simulation are $\mathcal{M}_c = 10^8 M_\odot$ and $D_L = 3$ Mpc. Posteriors shown in blue are the recovery of the signal from the wn+mem+rn dataset. There are clear covariances between parameters especially the chirp mass and the mass ratio of the binary. Red lines are the simulated values of the SMBHB merger, and green lines are the MAP estimates from the recovered posterior. The gray dashed vertical lines on the diagonal panels denote the 68% credible interval (1σ) for each parameter.

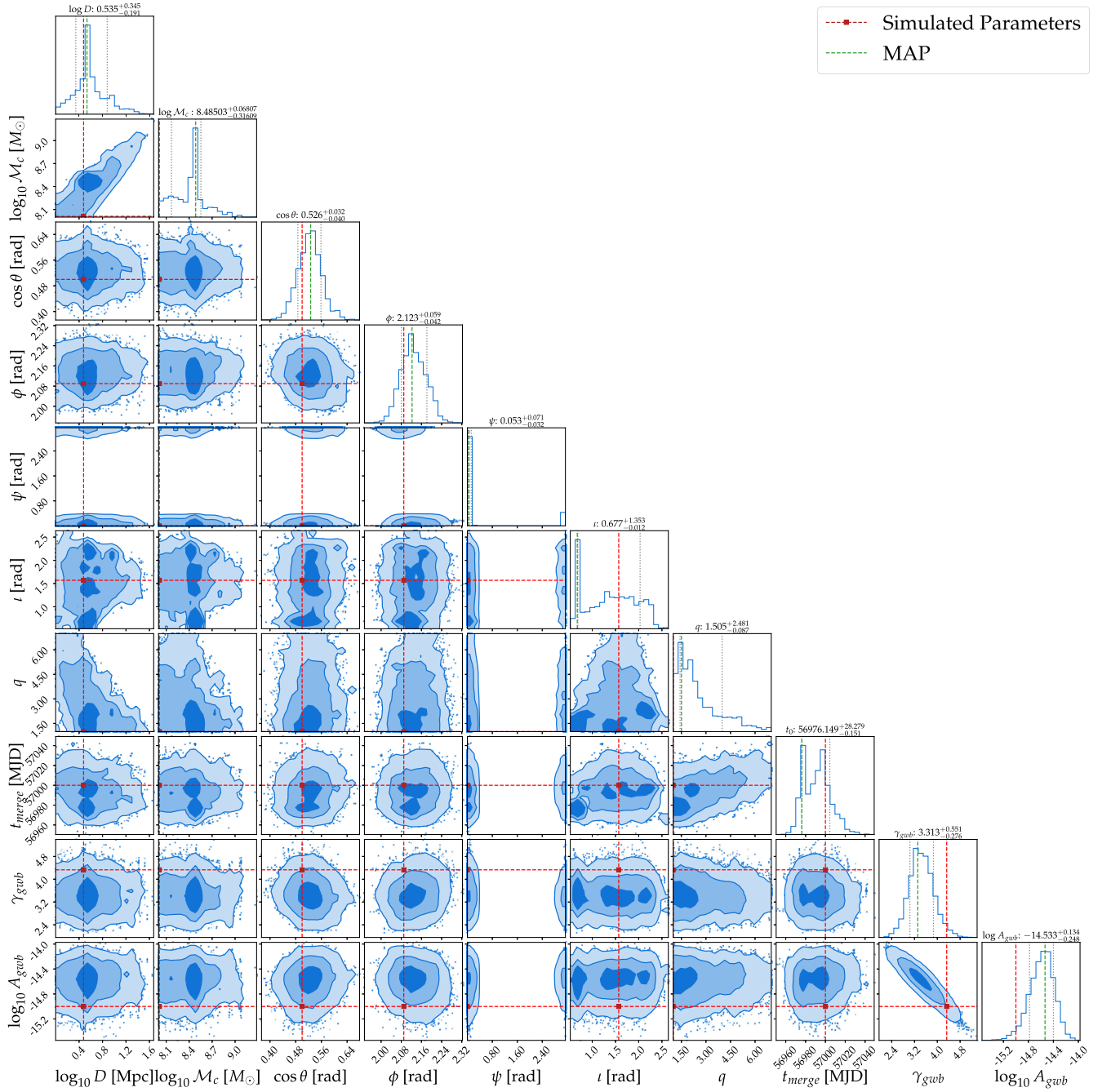


FIG. 8. Posterior distribution of the SMBHB merger from simulation studies on 25 pulsars. The dataset used is $wn+mem+rn+gwb$ as described in this paper. Parameters of the binary used for simulation are $M_c = 10^8 M_\odot$ and $D_L = 3$ Mpc. Overplotted red lines indicate the simulated values used in the simulated data, and green lines mark the MAP estimates from the recovered posterior. Gray dashed vertical lines on the denote the 68% credible interval (1σ) for each parameter. All recovered parameters agree with the simulated values within a percentage uncertainty of $\lesssim 0.1\%$.

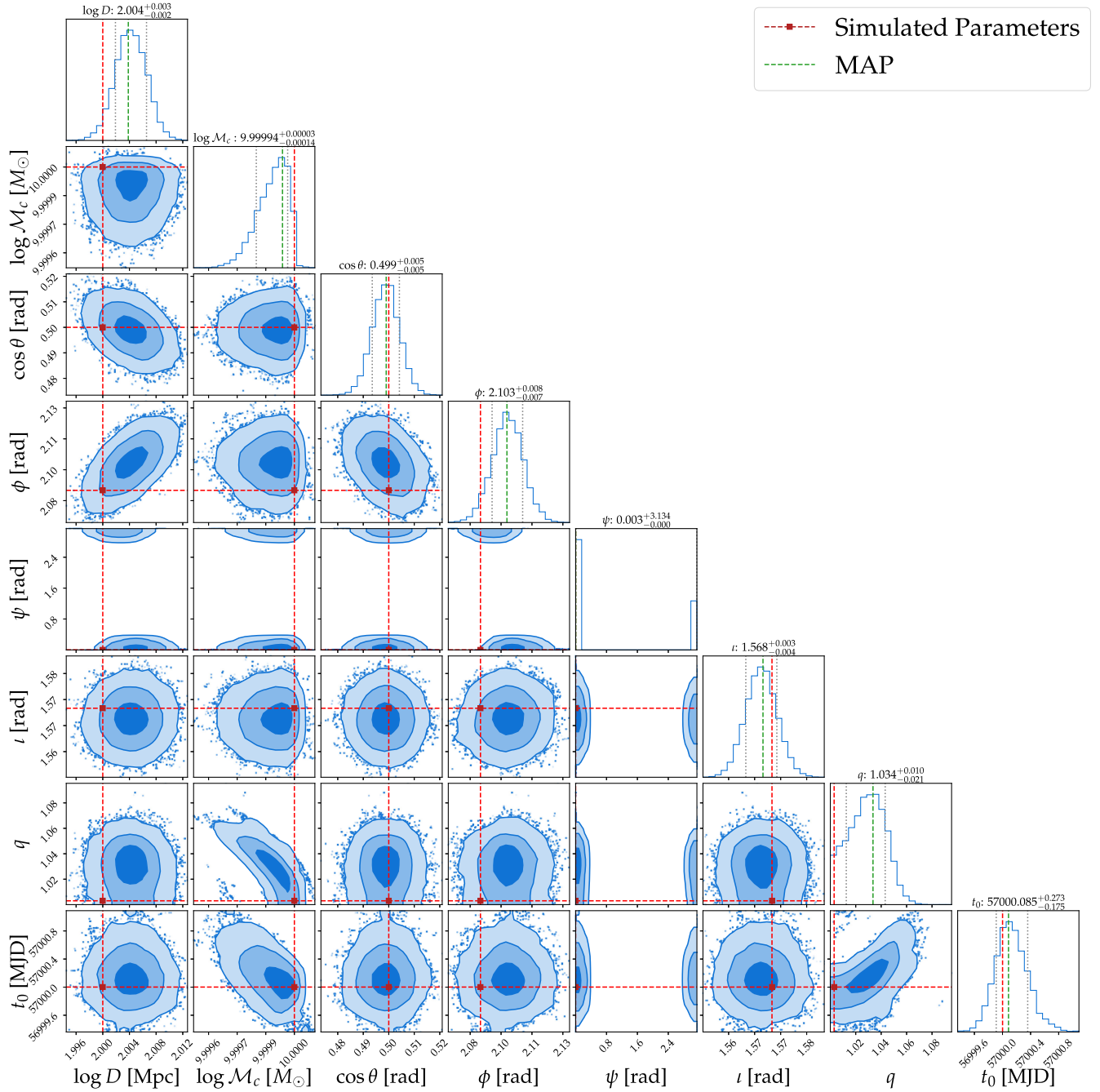


FIG. 9. Posterior distributions of SMBHB merger with null memory parameters from simulation studies on 25 pulsars. Parameters of the binary used for simulation are $\mathcal{M}_c = 10^{10} M_\odot$ and $D_L = 100$ Mpc. Posteriors shown in blue are the recovery of the signal from the wn+mem+rn dataset. Red lines are the simulated values of the SMBHB merger, and green lines are the MAP estimates from the recovered posterior.

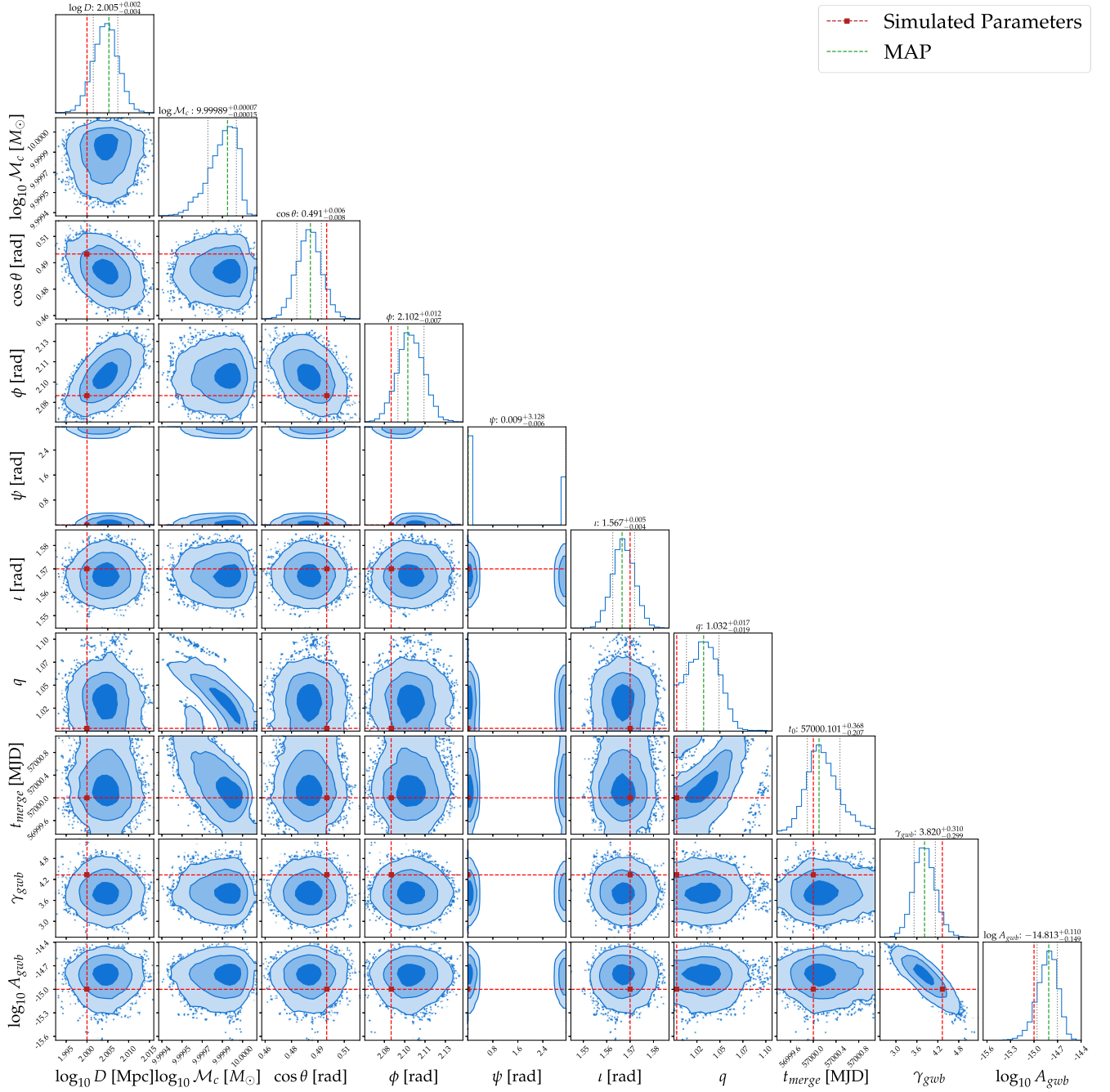


FIG. 10. Posterior distribution of the SMBHB merger from simulation studies on 25 pulsars. Parameters of the binary used for simulation are $\mathcal{M}_c = 10^{10} M_\odot$ and $D_L = 100$ Mpc. The dataset used is wn+mem+rn+gwb. Overplotted red lines indicate the simulated values used in the simulated data, and green lines mark the MAP estimates from the recovered posterior.

- [1] R. S. Foster and D. C. Backer, *Astrophys. J.* **361**, 300 (1990).
 [2] M. A. McLaughlin, *Classical Quantum Gravity* **30**, 224008 (2013).

- [3] R. N. Manchester *et al.*, *Pub. Astron. Soc. Aust.* **30**, e017 (2013).
 [4] G. Desvignes *et al.*, *Mon. Not. R. Astron. Soc.* **458**, 3341 (2016).

- [5] P. Tarafdar *et al.*, *Pub. Astron. Soc. Aust.* **39**, e053 (2022).
- [6] K. J. Lee, in *Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015*, Astronomical Society of the Pacific Conference Series Vol. 502, edited by L. Qain and D. Li (Astronomical Society of the Pacific, San Francisco, 2016), p. 19.
- [7] M. T. Miles, R. M. Shannon, M. Bailes, D. J. Reardon, M. J. Keith, A. D. Cameron, A. Parthasarathy, M. Shamohammadi, R. Spiewak, W. van Straten, S. Buchner, F. Camilo, M. Geyer, A. Karastergiou, M. Kramer, M. Serylak, G. Theureau, and V. Venkatraman Krishnan, *Mon. Not. R. Astron. Soc.* **519**, 3976 (2023).
- [8] J. P. W. Verbiest *et al.*, *Mon. Not. R. Astron. Soc.* **458**, 1267 (2016).
- [9] Z. Arzoumanian, P. T. Baker, H. Blumer, B. Bécsy, A. Brazier, P. R. Brook, S. Burke-Spolaor, S. Chatterjee, S. Chen, J. M. Cordes, N. J. Cornish, F. Crawford, H. T. Cromartie, M. E. Decesar, P. B. Demorest, T. Dolch, J. A. Ellis, E. C. Ferrara, W. t. Fiore (Nanograv Collaboration), *Astrophys. J. Lett.* **905**, L34 (2020).
- [10] B. Goncharov *et al.*, *Astrophys. J. Lett.* **917**, L19 (2021).
- [11] B. Goncharov, E. Thrane, R. M. Shannon, J. Harms, N. D. R. Bhat, G. Hobbs, M. Kerr, R. N. Manchester, D. J. Reardon, C. J. Russell, X.-J. Zhu, and A. Zic, *Astrophys. J. Lett.* **932**, L22 (2022).
- [12] S. Chen *et al.*, *Mon. Not. R. Astron. Soc.* **508**, 4970 (2021).
- [13] J. Antoniadis *et al.*, *Mon. Not. R. Astron. Soc.* **510**, 4873 (2022).
- [14] G. Agazie *et al.*, *Astrophys. J. Lett.* **951**, L8 (2023).
- [15] J. Antoniadis *et al.* (EPTA Collaboration and InPTA Collaboration), *Astron. Astrophys.* **678**, A50 (2023).
- [16] D. J. Reardon *et al.*, *Astrophys. J. Lett.* **951**, L6 (2023).
- [17] M. T. Miles *et al.*, *Mon. Not. R. Astron. Soc.* **536**, 1489 (2025).
- [18] H. Xu *et al.*, *Res. Astron. Astrophys.* **23**, 075024 (2023).
- [19] A. Afzal *et al.*, *Astrophys. J. Lett.* **951**, L11 (2023).
- [20] J. Antoniadis *et al.* (EPTA Collaboration and InPTA Collaboration), *Astron. Astrophys.* **685**, A94 (2024).
- [21] J. Ellis, F. Jenet, and X. Siemens, in *American Astronomical Society Meeting Abstracts #219*, American Astronomical Society Meeting Abstracts Vol. 219 (American Astronomical Society, 2012), p. 146.20.
- [22] T. Damour and A. Vilenkin, *Phys. Rev. Lett.* **85**, 3761 (2000).
- [23] Y. Xia, J. Wang, S. Kuroyanagi, W. Yan, Y. Wen, A. Kapur, J. Zou, Y. Feng, V. Di Marco, S. Mishra, C. J. Russell, S. Wang, D. Zhao, and X. Zhu, *Universe* **11**, 81 (2025).
- [24] M. Turner, *Astrophys. J.* **216**, 610 (1977).
- [25] M. Favata, *Classical Quantum Gravity* **27**, 084036 (2010).
- [26] R. van Haasteren and Y. Levin, *Mon. Not. R. Astron. Soc.* **401**, 2372 (2010).
- [27] S. Dandapat, A. Susobhanan, L. Dey, A. Gopakumar, P. T. Baker, and P. Jetzer, *Phys. Rev. D* **109**, 103018 (2024).
- [28] N. Agarwal *et al.*, *arXiv:2508.16534*.
- [29] J. A. Taylor, R. Burnette, B. Bécsy, and N. J. Cornish, *Phys. Rev. D* **111**, 022006 (2025).
- [30] S. Dandapat, M. Ebersold, A. Susobhanan, P. Rana, A. Gopakumar, S. Tiwari, M. Haney, H. M. Lee, and N. Kolhe, *Phys. Rev. D* **108**, 024013 (2023).
- [31] R. Epstein, *Astrophys. J.* **223**, 1037 (1978).
- [32] D. R. Madison, D. F. Chernoff, and J. M. Cordes, *Phys. Rev. D* **96**, 123016 (2017).
- [33] N. Sago, K. Ioka, T. Nakamura, and R. Yamazaki, *Phys. Rev. D* **70**, 104012 (2004).
- [34] Y. B. Zel'dovich and A. G. Polnarev, *Sov. Astron.* **18**, 17 (1974).
- [35] V. B. Braginsky and K. S. Thorne, *Nature (London)* **327**, 123 (1987).
- [36] D. Christodoulou, *Phys. Rev. Lett.* **67**, 1486 (1991).
- [37] M. Favata, *Astrophys. J. Lett.* **696**, L159 (2009).
- [38] N. Seto, *Mon. Not. R. Astron. Soc.* **400**, L38 (2009).
- [39] M. S. Pshirkov, D. Baskaran, and K. A. Postnov, *Mon. Not. R. Astron. Soc.* **402**, 417 (2010).
- [40] J. B. Wang *et al.*, *Mon. Not. R. Astron. Soc.* **446**, 1657 (2015).
- [41] K. Aggarwal *et al.*, *Astrophys. J.* **889**, 38 (2020).
- [42] G. Agazie *et al.* (The Nanograv Collaboration), *Astrophys. J.* **963**, 61 (2024).
- [43] G. Agazie *et al.*, *Astrophys. J.* **987**, 5 (2025).
- [44] J. Yoo, K. Mitman, V. Varma, M. Boyle, S. E. Field, N. Deppe, F. Hébert, L. E. Kidder, J. Moxon, H. P. Pfeiffer, M. A. Scheel, L. C. Stein, S. A. Teukolsky, W. Throwe, and N. L. Vu, *Phys. Rev. D* **108**, 064027 (2023).
- [45] N. Bishop, R. Gomez, L. Lehner, and J. Winicour, *arXiv:gr-qc/9705033*.
- [46] J. Moxon, M. A. Scheel, S. A. Teukolsky, N. Deppe, N. Vu, F. Hébert, L. E. Kidder, and W. Throwe, *Phys. Rev. D* **107**, 064013 (2023).
- [47] K. Mitman, J. Moxon, M. A. Scheel, S. A. Teukolsky, M. Boyle, N. Deppe, L. E. Kidder, and W. Throwe, *Phys. Rev. D* **102**, 104007 (2020).
- [48] J. A. Ellis, X. Siemens, and J. D. E. Creighton, *Astrophys. J.* **756**, 175 (2012).
- [49] J. W. McKee *et al.*, *Mon. Not. R. Astron. Soc.* **461**, 2809 (2016).
- [50] S. M. Tomson, Smbhb_merger_memory, https://github.com/sharontomson/SMBHB_Merger_Memory (2026), GitHub repository.
- [51] M. T. Miles *et al.*, *Mon. Not. R. Astron. Soc.* **536**, 1467 (2025).
- [52] R. van Haasteren, Y. Levin, P. McDonald, and T. Lu, *Mon. Not. R. Astron. Soc.* **395**, 1005 (2009).
- [53] W. Coles, G. Hobbs, D. J. Champion, R. N. Manchester, and J. P. W. Verbiest, *Mon. Not. R. Astron. Soc.* **418**, 561 (2011).
- [54] R. M. Shannon and J. M. Cordes, *Astrophys. J.* **725**, 1607 (2010).
- [55] L. Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren, *Phys. Rev. D* **87**, 104021 (2013).
- [56] R. van Haasteren and M. Vallisneri, *Mon. Not. R. Astron. Soc.* **446**, 1170 (2015).
- [57] B. Allen and J. D. Romano, *Phys. Rev. Lett.* **134**, 031401 (2025).
- [58] M. Crisostomi, R. van Haasteren, P. M. Meyers, and M. Vallisneri, *arXiv:2506.13866*.
- [59] R. W. Hellings and G. S. Downs, *Astrophys. J. Lett.* **265**, L39 (1983).
- [60] Z. Arzoumanian *et al.* (NANOGrav Collaboration), *Astrophys. J.* **821**, 13 (2016).

- [61] J. Antoniadis, Z. Arzoumanian, S. Babak, M. Bailes, A. S. Bak Nielsen, P. T. Baker, C. G. Bassa, B. Bécsey, A. Berthureau, M. Bonetti, A. Brazier, P. R. Brook, M. Burgay, S. Burke-Spolaor, R. N. Caballero, J. A. Casey-Clyde *et al.*, *Mon. Not. R. Astron. Soc.* **510**, 4873 (2022).
- [62] S. Hee, W. J. Handley, M. P. Hobson, and A. N. Lasenby, *Mon. Not. R. Astron. Soc.* **455**, 2461 (2016).
- [63] S. R. Taylor, P. T. Baker, J. S. Hazboun, J. Simon, and S. J. Vigeland, ENTERPRISE_EXTENSIONS (2021) https://github.com/nanograv/enterprise_extensions, v2.4.3.
- [64] A. Petiteau, S. Babak, A. Sesana, and M. de Araújo, *Phys. Rev. D* **87**, 064036 (2013).
- [65] J. Ellis and R. van Haasteren, PTMCMCSampler: Parallel tempering MCMC sampler package written in Python, Astrophysics Source Code Library, record ascl:1912.017 (2019) <https://ui.adsabs.harvard.edu/abs/2019ascl.soft12017E>.
- [66] D. Pollney and C. Reisswig, *Astrophys. J. Lett.* **732**, L13 (2011).
- [67] M. Favata, *Phys. Rev. D* **84**, 124013 (2011).
- [68] A. Sesana, *Astrophys. J.* **719**, 851 (2010).
- [69] E. M. Gutiérrez, L. Combi, S. C. Noble, M. Campanelli, J. H. Krolik, F. López Armengol, and F. García, *Astrophys. J.* **928**, 137 (2022).
- [70] T. Bogdanović, M. C. Miller, and L. Blecha, *Living Rev. Relativity* **25**, 3 (2022).
- [71] J. Cardinal Tremblay, B. Goncharov, R. van Haasteren, N. D. R. Bhat, Z.-C. Chen, V. Di Marco, S. Iguchi, A. Kapur, W. Ling, R. Mandow, S. Mishra, D. J. Reardon, R. M. Shannon, H. Sudou, J. Wang, S.-Y. Zhao, X.-J. Zhu, and A. Zic, *arXiv:2508.20007*.
- [72] B. P. Abbott *et al.*, *Ann. Phys. (Berlin)* **529**, 1600209 (2017).
- [73] S. M. Tomson *et al.*, *Astrophys. J. Lett.* **996**, L9 (2026).
- [74] S. Field, V. Varma, J. Blackman, B. Gadre, C. Galley, T. Islam, K. Mitman, M. Pürrer, A. Ravichandran, M. Scheel, L. Stein, and J. Yoo, *J. Open Source Software* **10**, 7073 (2025).
- [75] https://github.com/bencebecsy/pta_replicator.
- [76] https://github.com/sharontomson/SMBHB_Merger_Memory.
- [77] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker, ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust Pulsar Inference Suite (2020) 10.5281/zenodo.4059815.
- [78] https://github.com/bvgoncharov/enterprise_warp.
- [79] <https://github.com/nanograv/enterprise>.
- [80] S. M. Tomson, Simulated pulsar timing array data for smbhb merger searches with full waveform memory (2026) 10.5281/zenodo.19385397.