The Astrophysical Journal Letters, 996:L9 (14pp), 2026 January 1 https://doi.org/10.3847/2041-8213/ae2912

Search for Gravitational-wave Memory in PPTA and EPTA Data: A Complete Signal Model

Sharon Mary Tomson[1,2], Boris Goncharov[1,2], Rutger van Haasteren[1,2], Rahul Srinivasan[3,4], Enrico Barausse[3,4], Yirong Wen[5], Jingbo Wang[6], John Antoniadis[7,8], N. D. Ramesh Bhat[9], Zu-Cheng Chen[10,11], Ismael Cognard[12,13], Valentina Di Marco[14,15,16], Huanchen Hu[8], Gemma H. Janssen[17,18], Michael Kramer[8], Wenhua Ling[16], Kuo Liu[8,19,20], Saurav Mishra[15,16,21], Delphine Perrodin[22], Andrea Possenti[22], Christopher J. Russell[16], Ryan M. Shannon[21,15], Gilles Theureau[12,13], and Shuangqiang Wang[5,16]
[1] Max Planck Institute for Gravitational Physics (Albert Einstein Institute), 30167 Hannover, Germany; sharon.mary.tomson@aei.mpg.de, boris.goncharov@me.com
[2] Leibniz Universität Hannover, 30167 Hannover, Germany
[3] SISSA, Via Bonomea 265, 34136 Trieste, Italy & INFN Sezione di Trieste, Italy
[4] IFPU—Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy
[5] Xinjiang Astronomical Observatories, Chinese Academy of Sciences, Urumqi 830011, People's Republic of China
[6] Institute of Optoelectronic Technology, Lishui University, Lishui 323000, People's Republic of China
[7] Institute of Astrophysics, Foundation for Research & Technology—Hellas (FORTH), GR-70013 Heraklion, Greece
[8] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, DE-53121 Bonn, Germany
[9] International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia
[10] Department of Physics and Synergetic Innovation Center for Quantum Effects and Applications, Hunan Normal University, Changsha, Hunan 410081, People's Republic of China
[11] Institute of Interdisciplinary Studies, Hunan Normal University, Changsha, Hunan 410081, People's Republic of China
[12] LPC2E, OSUC, Université Orléans, CNRS, CNES, Observatoire de Paris, Université PSL, F-45071 Orléans, France
[13] Observatoire Radioastronomique de Nançay, Observatoire de Paris, Université PSL, Univ Orléeans, CNRS, 18330 Nançay, France
[14] School of Physics and Astronomy, Monash University, Clayton VIC 3800, Australia
[15] OzGrav: The ARC Center of Excellence for Gravitational Wave Discovery, Clayton VIC 3800, Australia
[16] CSIRO, Space and Astronomy, PO Box 76, Epping, NSW 1710, Australia
[17] ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, 7991 PD, Dwingeloo, The Netherlands
[18] Department of Astrophysics/IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[19] Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, People's Republic of China
[20] State Key Laboratory of Radio Astronomy and Technology, A20 Datun Road, Chaoyang District, Beijing, 100101, People's Republic of China
[21] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC, 3122, Australia
[22] INAF—Osservatorio Astronomico di Cagliari, via della Scienza 5, 09047, Selargius, Italy
Received 2025 September 2; revised 2025 December 5; accepted 2025 December 5; published 2025 December 24

Abstract

We perform searches for gravitational-wave memory in the data of two major Pulsar Timing Array (PTA) experiments located in Europe and Australia. Supermassive black hole binaries (SMBHBs) are the primary sources of gravitational waves in PTA experiments. We develop and carry out the first search for late inspirals and mergers of these sources based on full numerical relativity waveforms with null (nonlinear) gravitational-wave memory. Additionally, we search for generic bursts of null gravitational-wave memory, exploring the possibilities of reducing the computational costs of these searches through kernel density estimation and normalizing flow approximations of the posteriors. We rule out mergers of SMBHBs with a chirp mass of $10^{10}\,M_{\odot}$ up to 700 Mpc over 18 yr of observation at 95% credibility. We rule out the observation of generic displacement memory bursts with strain amplitudes $>10^{-14}$ in brief periods of observation time but across the sky, or over the whole observation time but for certain preferred sky positions, at 95% credibility.

Unified Astronomy Thesaurus concepts: Gravitational waves (678); Astronomy data analysis (1858); Pulsars (1306); Black holes (162)

1. Introduction

Pulsar Timing Array (PTA; R. S. Foster & D. C. Backer 1990) experiments have a primary goal of detecting nanohertz-frequency gravitational waves (GWs) through decade-long observations of pulse arrival times from a set of millisecond pulsars (M. V. Sazhin 1978; S. Detweiler 1979). This goal is now within reach, with various lines of evidence for the stochastic background in the data of all major PTAs, albeit with varying and sometimes debatable levels of statistical significance (Z. Arzoumanian et al. 2020; S. Chen et al. 2021; B. Goncharov et al. 2021b; J. Antoniadis et al. 2022; G. Agazie et al. 2023; EPTA Collaboration et al. 2023; D. J. Reardon et al. 2023a; H. Xu et al. 2023; M. T. Miles et al. 2025). GWs are sourced from a mass quadrupole moment, resulting in the primary oscillatory strain time series $h(t)$ observed at detectors. It is expected that the superposition of many such mass quadrupole sources, including supermassive black hole binaries (SMBHBs), constitutes the background. With this, general relativity (GR) predicts that the emission of GWs is always accompanied by an additional nonoscillatory contribution to the signal, known as the GW memory (M. Favata 2010); the standard position-shift effect is often termed *displacement memory* (A. Strominger & A. Zhiboedov 2014).

GW displacement memory is the net change in the GW strain caused by the propagation of the mass-energy from the GW source to infinity. The ordinary (linear) memory was discovered by Y. B. Zel'dovich & A. G. Polnarev (1974), when considering unbound masses (for example, unbound black holes in hyperbolic trajectories, gamma-ray bursts, or supernova events). The null (nonlinear)[23] memory was discovered by D. Christodoulou (1991) by considering compact binaries. In this case, the unbound energy is carried out by gravitational[24] radiation to null infinity, as suggested by K. S. Thorne (1992). Due to being sourced from the backreaction of GWs onto the stress-energy tensor, null memory is a nonlinear effect in GR.

In addition to the displacement memory, there are higher-order memory terms associated with higher-order temporal moments of the curve deviation near null infinity (A. M. Grant & K. Mitman 2024)—in particular, the spin memory (S. Pasterski et al. 2016) and the center-of-mass memory (D. A. Nichols 2018). It is worth emphasizing that the GW memory is tied to symmetries of spacetime through Noether's theorem. Other effects, previously termed "memory," unrelated to symmetries, such as "velocity memory," are now referred to as "persistent observables" (É. É. Flanagan et al. 2019). Fascinatingly, memory effects, spacetime symmetries, and soft theorems (S. Weinberg 1964) derived in different formalisms spanning GR and quantum field theory over decades of studies are found to represent the same physical phenomenon. This connection is known as the infrared triangle (A. Strominger & A. Zhiboedov 2014). Therefore, the detection of GW memory will enable studies of spacetime symmetries and tests of the nonlinear nature of GR (B. Goncharov et al. 2024).

Detecting a nonlinear displacement memory signal with PTAs depends on the occurrence of mergers of SMBHBs. While these systems are the dominant sources of the stochastic GW background (GWB) in the nanohertz band, individual mergers that can produce detectable memory are rare. Moreover, the rates of such mergers are not well constrained, introducing significant uncertainty on the expected number of detectable memory events. Observing the scattering of SMBHBs producing an ordinary memory is much less certain, as it is not clear how the scattered black hole would escape to asymptotic infinity in the vicinity of a galactic center. Luckily, observing spans of contemporary PTAs have already exceeded a few decades, while detector sensitivity continues to improve (G. Hobbs et al. 2020; R. Spiewak et al. 2022; S. Chen et al. 2025). Memory bursts were first studied theoretically in the PTA context by several early works. N. Seto (2009) outlined how PTAs could detect both inspiral and memory components from SMBHBs. Near-simultaneously, R. van Haasteren & Y. Levin (2010) developed more detailed signal modeling to explore the detection prospects of memory events, followed by M. S. Pshirkov et al. (2010), who assessed PTA sensitivity to such events. In all of these studies, the net strain displacement was modeled as a step function. We refer to this approximation as the *memory burst*. The detectability of this signal with an array of 20 pulsars observed with a precision of 100 ns with a cadence of approximately two weeks over a period of 10 yr (250 pulse arrival time measurements) was reported by R. van Haasteren & Y. Levin (2010). They found, with simplified noise assumptions, that the black hole merger with $\mathcal{M} = 10^8\,M_\odot$ at $z = 0.1$ (equivalently, $\mathcal{M} = 10^9\,M_\odot$ at $z = 1$) can be detected with 2σ confidence. M. Enoki et al. (2004) expect an event rate of SMBHB mergers with a total mass of $M \sim 10^8\,M_\odot$ at redshifts $z < 1$ to be 0.1 events a year. Overall, R. van Haasteren & Y. Levin (2010) estimate ~0.01–0.1 detectable memory bursts over the 10 yr observation time. J. M. Cordes & F. A. Jenet (2012) present the same estimate, suggesting a possible improvement when considering memory bursts at pulsar positions. More recent population studies find very low PTA rates for detectable memory. For example, V. Ravi et al. (2015) predict that over a 10 yr PTA span, only $\sim 10^{-5}$ bursts with memory amplitude $>5 \times 10^{-15}$ and $\sim 10^{-3}$ bursts with $>2 \times 10^{-15}$ amplitude would occur. K. Islo et al. (2019) forecast that memory bursts with a final GW strain offset of at least 5×10^{-16} occur only 8.7×10^{-4}–5.2×10^{-3} times per century with a more optimistic model.

PTAs have performed several searches for displacement memory. The Parkes Pulsar Timing Array (PPTA), using the frequentist approach, performed a search for memory bursts in their initial dataset, setting an upper limit on the rates of such events that could be detected (J. B. Wang et al. 2015). The North American Nanohertz Observatory for Gravitational Waves (NANOGrav) had conducted searches on their 11 yr (K. Aggarwal et al. 2020), 12.5 yr (G. Agazie et al. 2024), and 15 yr datasets (G. Agazie et al. 2025), using the Bayesian approach to data analysis. These analyses have found no statistically significant evidence for memory signals, with the 15 yr dataset yielding a Bayes factor of 3.1 ($\ln \mathcal{B} = 1.13$) in favor of a model including memory bursts against the null hypothesis including pulsar-intrinsic noise and a common time-correlated stochastic signal. Upper limits have been placed on the strain amplitude as a function of sky location and burst time. S. Dandapat et al. (2024) searched for ordinary displacement memory from hyperbolic SMBHB encounters using the NANOGrav 12.5 yr data.

To date, no search for GW memory bursts has been performed using the European Pulsar Timing Array (EPTA) data. In contrast J. B. Wang et al. (2015) conducted a search on PPTA data, but that was 10 yr ago. Thus, it is of interest to perform searches on the latest EPTA and PPTA data.

There is a motivation to improve searches for GW memory with PTAs. First, the memory burst approximation is suboptimal for SMBHBs—the primary sources of GWs at nanohertz frequencies. A full waveform model for an SMBHB merger consists of the inspiral, merger, ringdown, and memory component. Contemporary continuous GW searches from SMBHBs with PTAs only look for the inspiral components, assuming the signal to be monochromatic on the observational timescales of decades. The search for ordinary memory from hyperbolic SMBHBs by S. Dandapat et al. (2024) is beyond a simple burst, but a search for mergers with the most abundant bound SMBHBs producing null memory has not been performed yet. Second, searches for GW memory with PTAs are becoming computationally expensive, due to the ever-increasing complexity of data and noise, as well as the necessity of modeling the GWB.

J. Sun et al. (2023) introduced the factorized Bayesian approach, based on likelihood factorization and the use of lookup tables (LUTs). This method has been used in

[23] When referring to types of memory, "ordinary" and "null" are the modern terms, whereas "linear" and "nonlinear," respectively, are the legacy terms (K. Mitman et al. 2020, 2024; L. Bieri 2021).

[24] In the most general case, also electromagnetic radiation (L. Bieri & D. Garfinkle 2013).

G. Agazie et al. (2024, 2025). However, LUTs are a discrete approximation of the likelihood. Modern data analysis tools allow for the introduction of smooth and continuous approximations of the likelihood.

In this work, we address the above limitations and opportunities. We develop the first methodology to search for SMBHB mergers using the full waveform from numerical relativity (NR), including the inspiral phase, merger, ringdown, and memory (S. M. Tomson et al. 2025), in contrast to the inspiral-only waveforms employed in continuous wave searches. We also develop a methodology to search for GW memory using likelihood factorization with kernel density estimation (KDE) and normalizing flows (R. Srinivasan et al. 2024). Next, we perform searches for SMBHB mergers and generic GW memory bursts in the second data release (DR2) of the EPTA and PPTA DR3.

The rest of the Letter is organized as follows. In Section 2, we describe the properties of the data used in this work—EPTA DR2 and PPTA DR3. In Section 3, we outline the methodology of our searches. In Section 4, we present our results. In Section 5, we highlight important aspects of memory searches to keep in mind for future work. We summarize the conclusions for GW memory in Section 6.

2. The Datasets

PPTA DR3 (A. Zic et al. 2023) contains observations of 32 pulsars over 18 yr with a cadence of 3 weeks. The data were obtained using the 64 m Parkes "Murriyang" radio telescope. From its onset in 2004, this telescope has been continuously acquiring data from 48 hr of observations every 2–3 weeks. PPTA DR3 is a combination of the data from PPTA's previous release (data up to 2018) along with data acquired from an ultrawideband low-frequency (UWL) receiver (G. Hobbs et al. 2020). The UWL receiver observes in a continuous radio frequency bandwidth between 704 and 4032 MHz, enhancing the instantaneous sensitivity. Narrowband pulsar timing is implemented, and the arrival times are acquired using a frequency-dependent pulse portrait.

EPTA DR2 (J. Antoniadis et al. 2023) contains data from 25 pulsars that were collected from six European radio telescopes: the Effelsberg 100 m radio telescope, the 76 m Lovell Telescope, the radio telescope at Nançay Radio Observatory, the 64 m Sardinia Radio Telescope, and the Westerbork Synthesis Radio Telescope. Compared to the previous data release, which only included data from the legacy data recording system, EPTA DR2 adds data from next-generation coherent dedispersion recording systems covering a wider bandwidth and thus offering better sensitivity at each telescope. The time spans for pulsar datasets range from 14 to a maximum of 25 yr, with a broad frequency coverage of between about 300 MHz and 4 GHz.

We perform our analysis with two versions of EPTA DR2: the complete dataset and a shortened version that includes only the observations made with the next generation of backend receiver combinations. These newer backend systems use field-programmable gate array hardware to digitize the incoming electrical signals. The complete dataset, covering 24.7 yr of observations, will be referred to as the "EPTA 25 yr" data henceforth. The truncated version, including only the last 10.3 yr of observations, collected using newer wideband backends, will be referred to as the "EPTA 10 yr" data henceforth.

Our pulsar-specific noise models are as follows. For EPTA DR2, our models are based on the result of the analysis by B. Goncharov et al. (2024b), excluding the hierarchical modeling of red noise. B. Goncharov et al. (2024b) build their noise models on top of the original EPTA DR2 noise analysis by the EPTA Collaboration and InPTA Collaboration et al. (2023). For PPTA DR3, we adopt the noise models from D. J. Reardon et al. (2023b).

We model the contribution of the GWB as a temporally correlated and spatially uncorrelated (i.e., no Hellings–Downs or HD correlations) stochastic process, with the same power spectrum of temporal correlations across pulsars. This simplification enables significantly faster Bayesian inference, by reducing the complexity of the covariance matrix. The procedure is justified, because a null memory signal is deterministic and hence not intrinsically correlated with the GWB spatial covariance structure. However, in the case of the EPTA 10 yr dataset, we include HD correlations in the model, as preliminary analyses showed a spurious signal when spatial correlations were unaccounted for. Further details and discussion are provided in the results section.

3. Methodology

3.1. Bayesian Inference

We use the standard PTA Gaussian likelihood for the timing-residual data vector $\boldsymbol{\delta t}$:

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

where the vector $\boldsymbol{\mu}$, a function of $\boldsymbol{\theta}$, represents a prediction of deterministic contributions to pulse arrival times, including our models, as well as deterministic noise transients, such as exponential dips (B. Goncharov et al. 2021a), and $\boldsymbol{C}$ is built following established PTA practice (including analytic marginalization over the linearized timing model; R. van Haasteren et al. 2009; Z. Arzoumanian et al. 2016; J. Antoniadis et al. 2022). The time-correlated noise for each pulsar is modeled with a Fourier basis, with the number of frequencies per pulsar set from our single-pulsar noise analyses (EPTA Collaboration and InPTA Collaboration et al. 2023; D. J. Reardon et al. 2023b). We also include a common stochastic GWB component represented by 30 Fourier frequencies. The model selection uses the Bayesian evidence $\mathcal{Z}$; the Bayes factors $\mathcal{B}$ are the ratios of evidence between competing models.

3.2. Signal Propagation and Projection in PTAs

A passing GW induces a fractional pulsar frequency shift:

$$\frac{\delta\nu(t)}{\nu} = F_{+,\times}(\theta, \phi, \psi)\,[h_{+,\times}(t) - h_{+,\times}(t - \boldsymbol{n}\cdot\boldsymbol{r})], \quad (2)$$

where $F_{+,\times}(\theta, \phi, \psi)$ is the geometric (quadrupolar) antenna response for a given pulsar to a GW source at a sky location (θ, ϕ) relative to the pulsar's line of sight, and ψ is the GW polarization angle. The first term is the Earth term (common to all pulsars), and the second is the pulsar term (R. van Haasteren & Y. Levin 2010).

The projected timing residuals $\delta t(t)$ in each pulsar can be expressed in a compact format as

$$\delta t(t) = \begin{bmatrix} F_+ & F_\times \end{bmatrix} \begin{bmatrix} \cos 2\psi & -\sin 2\psi \\ \sin 2\psi & \cos 2\psi \end{bmatrix} \begin{bmatrix} \delta t_+(t) \\ \delta t_\times(t) \end{bmatrix}, \quad (3)$$

where $\delta t_{+\times}(t) = \int_{t_0}^{t} h_{+,\times}(t')\, dt'$, as the residuals can be obtained by integrating the GW strain. For full derivations and definitions, see Section B of our methods paper (S. M. Tomson et al. 2025).

PTAs measure the *postfit* response obtained after a simultaneous fit of the pulsar timing model (spin and spindown). Denoting the prefit Earth-term response as $\delta t(t)$, the postfit signal is

$$\delta t_{\rm post}(t) = \delta t(t) - \mathcal{P}_{\rm poly}[\delta t_{\rm E}](t), \quad (4)$$

where $\mathcal{P}_{\rm poly}$ is the least-squares projection onto $\{1, t, t^2\}$ over the data span. In what follows, we apply this common projection to each signal model.

In this work we consider two signals: (1) a generic burst-with-displacement-memory model; and (2) a complete SMBHB-merger waveform including null memory. Because PTA pulsars lie at ∼kiloparsec distances, the pulsar term yields at most a single jump (step) in the GW strain over a ∼decade data span. For robust detection, we focus exclusively on the Earth-term response from the signal models.

3.3. The Signal Model for Bursts with Displacement Memory

In the burst (instantaneous) limit, the strain time series can be approximated as a step function in the metric traveling through space (R. van Haasteren & Y. Levin 2010) and is linearly polarized. In a suitable polarization basis (a convention), the memory can be written entirely in the + mode:

$$h_+(\boldsymbol{r}, t) = h_0 \times \Theta[(t - t_0) - \boldsymbol{n} \cdot \boldsymbol{r}], \quad (5)$$

where h_0 is the amplitude of the step, t_0 is the time at which the burst passes the solar system barycenter (SSB), $\boldsymbol{r}$ is the location of the pulsar relative to the SSB, $\boldsymbol{n}$ is the unit vector in the direction of the GW propagation, and $h_\times(\boldsymbol{r}, t) = 0$. In M. S. Pshirkov et al. (2010), the amplitude of the step for SMBHBs is shown to be $h_0 \propto \mu/r$, where μ is the reduced mass of the binary and r is the comoving distance. See also Equation 4.5 in the more recent study by A. Elhashash & D. A. Nichols (2025).

The Earth-term strain induced by a GW memory burst, as observed in pulsar timing residuals, can be expressed as

$$h(t) = F(\theta, \phi, \psi) \cdot h_0 \cdot \Theta(t - t_0). \quad (6)$$

In pulsar timing analyses, however, the measured quantity is the timing residuals, meaning the time integral of $h(t)$. An instantaneous step in the strain therefore produces a linearly growing (ramp) signature in the residuals at times after the burst.

3.4. SMBHB Mergers with Null Memory: A Complete Model

The burst-with-memory approximation is source-agnostic. It models a displacement without assuming a specific origin. As such, it can capture signals produced by SMBHBs as well as by more exotic mechanisms (e.g., cosmic string cusps; A. C. Jenkins & M. Sakellariadou 2021), provided the memory growth is fast compared to the PTA cadence. However, it does not encode the accompanying oscillatory waveform, nor does it cover other memory effects (e.g., spin or center-of-mass memory). For sources with slower memory buildup or distinctive morphology, dedicated signal models are preferable.

While this enables a broad *source- and morphology-agnostic* search, it neglects the physical context and waveform features expected from realistic sources.

In this section, we refine our methodology by adopting a physically motivated model for mergers of SMBHBs, including the contribution of the null memory. There are several key advantages to using the full signal model for null memory searches:

1. *Physics-motivated*. SMBHBs are the primary expected source of memory in PTAs. The full waveform captures the complete time-domain structure of the GW signal, including the inspiral, merger, ringdown, and gradual buildup of null memory. This offers a dedicated representation of the waveform from SMBHB mergers.
2. *Robustness to noise*. Generic burst models, while flexible, are more prone to capturing unmodeled noise artifacts.
3. *Astrophysical interpretability*. The full waveform model is parameterized by the chirp mass $\mathcal{M}$, luminosity distance D_L, mass ratio q, and sky location (R.A., decl.) of the SMBHB. This allows for the direct inference of source properties from data, such as the expected rates and number density of SMBHB mergers. This may offer insights into the assembly history of SMBHBs and galaxy evolution.
4. *Observable postfit amplitude for SMBHBs*. The burst template assumes an instantaneous step at the merger, whereas the SMBHB memory builds up gradually during the late inspiral (M. Favata 2009) and is captured by our SMBHB-merger waveform. After the timing-model fit (spin and spindown), part of this slow rise is absorbed; at the merger epoch, the burst (step) template produces a larger postfit bump than the SMBHB waveform—i.e., it overestimates the observable (postfit) amplitude. This effect is quantified below and in our companion methods paper (S. M. Tomson et al. 2025).

We use the `NRHybSur3dq8_CCE` surrogate waveform model developed by J. Yoo et al. (2023), which combines NR and post-Newtonian and effective-one-body waveforms through hybridization. This model includes both the oscillatory inspiral–merger–ringdown (IMR) signal and the null memory contribution, capturing the full GW strain with high fidelity. The strain polarizations $h_+(t)$ and $h_\times(t)$ are computed from the dominant radiative modes, including memory in the non-oscillatory (2, 0) mode. The timing residuals are then obtained using Equation (3). The waveform is parameterized by physical quantities—such as the chirp mass, mass ratio, spin, luminosity distance, inclination, and sky location of the binary—enabling direct astrophysical inference from PTA signals.

Figure 1 shows the prefit PTA responses on a pulsar located at (R.A., decl.) = ($258\overset{\circ}{.}4564$, $7\overset{\circ}{.}7937$) for a fiducial equal-mass source ($\mathcal{M} = 10^{10}\, M_\odot$, $D_L = 100$ Mpc, (R.A., decl.) = (0°, 0°)), using both the memory-burst (ramp) template and our SMBHB-merger model. However, PTAs are not sensitive to this time series in full. After subtracting the timing model (spin and spindown), part of the SMBHB signal's slow premerger rise is absorbed by the fit, so the

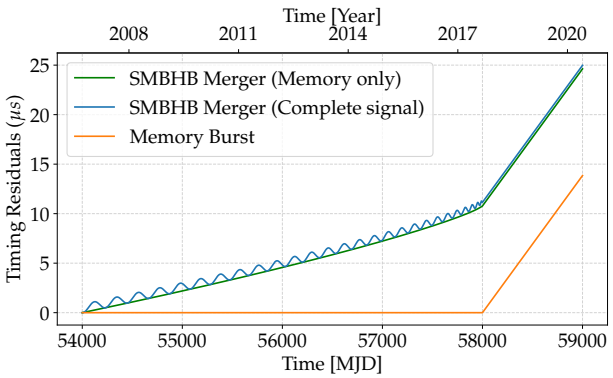


Figure 1. PTA timing residuals from a merger of a nonspinning SMBHB with parameters $\mathcal{M} = 10^{10}\,M_\odot$, $q = 1$, and $D_L = 100$ Mpc observed by a pulsar at (R.A., decl.) = (258.°4564, 7.°7937). The merger occurs at MJD = 58000. Green: SMBHB-merger model with only memory; blue: SMBHB-merger model (complete signal including IMR+memory); and orange: memory-burst model. All curves are normalized to the same final memory offset.

observable postfit bump is smaller than for a ramp with the same final memory offset. Quantitatively, the noise-agnostic rms difference between the postfit residuals of the two templates is 0.399 μs when their final memory offsets are matched. If the burst amplitude is allowed to float to minimize this mismatch, the rms drops to 0.26 μs, but the best-fit burst amplitude is 11.8% smaller than the merger model's offset (and the corresponding biases in $\mathcal{M}$ or D_L). The discrepancy shrinks by $\sim$0.1 μs when the mass ratio changes from $q = 1$ to $q = 7$. These results motivate the physically consistent IMR +memory waveform to avoid inflated signals and parameter estimation biases. The corresponding postfit residual panels and extended comparisons are shown in Figure 4 of S. M. Tomson et al. (2025).

3.5. Factorized Posterior Approximation

The factorized posterior (FP) approach is an accelerated Bayesian approach that constructs the full-PTA posterior from single-pulsar posteriors, provided the full posteriors can be represented as a product of individual posteriors. This is the case when neglecting interpulsar correlations for stochastic time-correlated signals. In this section, we review the application of FP to PTA data analyses. First, we explain the existing methodology of J. Sun et al. (2023). Second, we introduce extensions of this methodology, to allow for smooth and continuous posterior approximation, which are made possible by KDE and normalizing flows (R. Srinivasan et al. 2024).

Following J. Sun et al. (2023), the intrinsic (projected per-pulsar) amplitude of the memory h_i for the ith pulsar is related to the global (extrinsic) memory amplitude offset h_0 via the geometric projection factor as $h_i = h_0 F_i(\theta, \phi, \psi)$, where F_i is the geometric projection defined in Section 3.3 (Equation (6)). This factor determines the sign and magnitude of the intrinsic per-pulsar amplitude, indicating whether the timing residuals are advanced or delayed by the burst.

Once the intrinsic amplitudes h_i are computed, the FP approach combines the pulsar-term likelihoods:

$$p(\delta t|h_0, t_0, \theta, \phi, \psi) = \prod_{i=1}^{N} p_i(\delta t|h_i, t_0), \tag{7}$$

where t_0 is the burst epoch and N is the number of pulsars. Unlike the original implementation of J. Sun et al. (2023), which relies on precomputed LUTs over a grid of projected amplitudes, our approach uses smooth approximations of the single-pulsar posteriors using KDE and normalizing flows.

KDE is a nonparametric way of approximating a set of samples into a continuous (analytical) probability density function (B. W. Silverman 1986). Using KDEs within an FP framework for PTA analyses has been discussed previously (W. G. Lamb et al. 2023). It defines a kernel at every point in the parameter space and assigns a density to each point. We construct bounded KDEs from Markov Chain Monte Carlo (MCMC) samples over the 2D space of the intrinsic per-pulsar amplitude h_i and burst epoch t_0. While we start with a Gaussian kernel and use the "Scott" bandwidth selection rule as a baseline (D. W. Scott 2015), in practice we fine-tune the kernel width heuristically, based on the shape and structure of the posterior, to ensure adequate resolution and smoothness. We implement reflective padding to handle boundary effects that can lead to underestimated densities near the edges of the prior volume. This allows the KDE to correctly capture the probability mass near the limits without artificially truncating or distorting the distribution. As the memory signal can appear in pulsars with either positive or negative amplitude, depending on the sky geometry, we partition the MCMC samples into positive- and negative-amplitude subsets and build separate KDEs for each. The sign of the extrinsic amplitude h_0 during global inference selects which KDE to use for each pulsar's contribution. While KDEs provide a reasonable approximation of the 2D posteriors, they are computationally expensive during evaluation, since each likelihood call involves summing over all single-pulsar MCMC samples. This computational cost motivates the adoption of faster alternatives, like normalizing flows.

Normalizing flows are a more efficient approach, based on machine learning, to approximating a complex *target* per-pulsar posterior $p(x)$ (with parameters x—e.g., $x = (h_i, t_0)$) with a flexible *model* density $p_\theta(x)$. The model is built by transforming a simple *base* distribution $p_Z(z)$ (typically, $\mathcal{N}(0, I)$) through an invertible map $f_\theta: \mathbb{R}^d \to \mathbb{R}^d$ (a normalizing flow; D. J. Rezende & S. Mohamed 2015). Writing $x = f_\theta(z)$ and $z = f_\theta^{-1}(x)$, the change-of-variables formula gives the density

$$p_\theta(x) = p_Z\left(f_\theta^{-1}(x)\right) \left| \det \frac{\partial f_\theta^{-1}(x)}{\partial x} \right|. \tag{8}$$

We fit θ by maximizing the likelihood (equivalently, minimizing the negative log-likelihood) on samples from $p(x)$. After training, the sampling is fast (draw $z \sim p_Z$, map $x = f_\theta(z)$), and the density evaluation uses $z = f_\theta^{-1}(x)$ and the Jacobian. In this work, we adopt Masked Autoregressive Flows (G. Papamakarios et al. 2019), a widely used architecture for density estimation, and adopt the `floZ` package from R. Srinivasan et al. (2024). In our FP pipeline, flows are much faster than KDE factors, because the evaluation and sampling do not scale with the number of stored samples.

To ensure the robust modeling of the bounded parameters and preserve the posterior structure, we apply a parameter-wise logit transformation to the posterior samples before training. This maps each parameter from its prior box $[a, b]$ to the whole real line, eliminating hard cutoffs that otherwise cause discontinuities and spurious density outside the prior support.

Let $r = (x - a)/(b - a) \in (0, 1)$. We define

$$y = \mathrm{logit}(r) = \log \frac{r}{1 - r} = \log \frac{x - a}{b - x} \tag{9}$$

and $x = a + (b - a)\,\sigma(y)$, where $\sigma(y) = 1/(1 + e^{-y})$ is the logistic function (Equation (1.10) of A. Gelman et al. 2014). The 1D Jacobian is

$$\left| \frac{dy}{dx} \right| = \frac{1}{(b - a)} \cdot \frac{1}{r(1 - r)} = \frac{b - a}{(x - a)(b - x)}. \tag{10}$$

Because this transform is applied component-wise, the Jacobian is diagonal in n dimensions (n parameters), and the total log-Jacobian is the sum over components.

After the logit step, the samples are whitened via eigenvalue decomposition (A. Heavens et al. 2017). Namely, the samples are projected along the eigenvectors of the covariance matrix of the samples and rescaled by the square root of the eigenvalues. This ensures the flow operates in a decorrelated and standardized latent space, facilitating efficient training.

The normalizing flow is trained on the whitened logit-space samples $\tilde{y}$. During training, the log-Jacobian of this fixed preprocessing does not depend on the flow parameters and therefore does not affect optimization. At evaluation time, however, when reporting densities back in the original variables x, we apply the change-of-variables rule to account for all invertible steps:

$$\log p_\theta(x) = \log p_\theta(\tilde{y}) + \log \left| \det \frac{\partial \tilde{y}}{\partial y} \right| + \log \left| \det \frac{\partial y}{\partial x} \right|. \tag{11}$$

Here, $\log p_\theta(\tilde{y})$ is provided by the trained flow, and $\frac{\partial \tilde{y}}{\partial y}$ and $\frac{dy}{dx}$ are the Jacobians. To generate samples in the original parameter space after training, we invert the pipeline by first drawing the samples $\tilde{y}$ from the flow, applying inverse whitening and inverse logit transformations, respectively.

Training the flow model involves tuning key parameters, like the number of flow layers, hidden units, learning rate, and batch size, all of which impact performance and convergence. After the model construction, we generate samples from the model in the original parameter space by inverse transformation.

4. Results

4.1. Results for Generic Memory Bursts

We perform a search for generic memory bursts, defined as per the methodology of Section 3.3. We perform a simultaneous fit of the signal model—which includes our signal (memory bursts), a common-spectrum noise process attributed to the GWB, as well as the noise model—to the data. We find $\ln \mathcal{B} = 0.712$ for the signal component in the 10 yr EPTA DR2, $\ln \mathcal{B} = -0.309$ for the 25 yr EPTA DR2, and $\ln \mathcal{B} = 0.432$ for PPTA DR3.

For computational efficiency, we model the GWB as a common-spectrum red noise process without HD spatial correlations in our main analysis. However, for the 10 yr EPTA DR2 dataset, this approximation led to an outlier with a higher Bayes factor. To address this, we included the HD-correlated GWB model for this particular dataset, which yielded the Bayes factor we reported above. We find no evidence of GW memory bursts in these data. With no evidence for memory bursts, based on the results of the parameter estimation,[25] we compute upper limits.

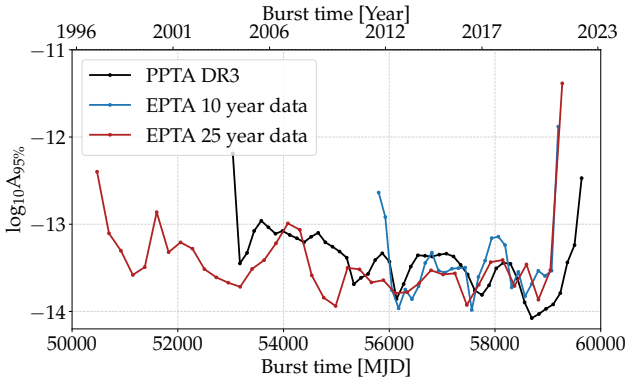


Figure 2. Upper limits on the strain amplitude h_0 of generic displacement memory bursts as a function of burst time at 95% credibility. Other burst parameters are marginalized over.

We follow the procedures outlined in G. Agazie et al. (2025) to compute the upper limits in two ways: (1) as a function of burst epoch, marginalizing over the sky location and polarization; and (2) as a function of sky position, marginalizing over the burst time and polarization. In Figure 2, we show the 95% upper limits on the memory-burst strain amplitude as a function of burst time. These limits vary, due to the inhomogeneous observational coverage across different years. The 25 yr EPTA DR2 extends to 1996, being our only source of information about memory bursts between 1996 and 2004. The PPTA has legacy observations dating back to 1994, but those epochs are not included in PPTA DR3 and are not used in this work. At about 2019, we achieve the most constraining limits with the PPTA data, ruling out strain amplitudes of $h > 10^{-14}$ between MJD 58600 and 58800. The sensitivity of the PPTA was enhanced during this period, due to the installation of the UWL receiver. It is visible that our sensitivity to memory bursts degrades toward the edges of the observation span. Memory bursts, which would have occurred beyond the observation span, would introduce pulsar timing delays that grow mostly linearly with time. Currently, PTAs cannot distinguish these signatures from errors in pulsar spin frequencies. Toward the middle of the observation span, the limits vary by up to 1 order of magnitude, due to the irregularity of the observations.

In Figure 3, we show the upper limits on the memory-burst strain amplitude as a function of the sky position. We place limits on the memory strain across 192 sky pixels of equal area. The burst epoch and polarization of the GW source are marginalized over. Due to the nonuniform distribution of pulsars in our PTA, the sensitivity to GW bursts varies significantly across the sky.

Thus, we perform a separate calculation of an upper limit for every sky location.

The most constraining limits for the PPTA DR3 are obtained at $(\theta, \phi) = (292.^\circ5, -66.^\circ444)$, and the least-constraining limits are obtained at $(\theta, \phi) = (56.^\circ25, 0)$. The most constraining limits for the 10 yr EPTA DR2 are obtained at $(\theta, \phi) = (281.^\circ25, 19.^\circ471)$, and the least-constraining limits are obtained at $(\theta, \phi) = (33.^\circ75, 19.^\circ471)$. The most constraining limits for the 25 yr EPTA DR2 are obtained at

[25] We show and discuss more broadly the results of the parameter estimation for all parameters of the memory-burst model in Appendix A.

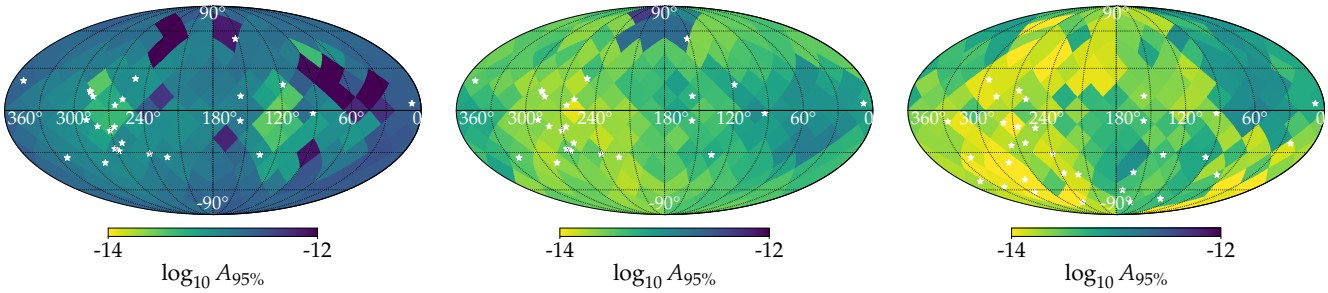


Figure 3. Upper limits on the strain amplitude h_0 of generic displacement memory bursts, shown in color, as a function of sky position (θ, ϕ) at 95% credibility. Other burst parameters are marginalized over. The three panels correspond to the EPTA 10 yr, EPTA 25 yr, and PPTA DR3 datasets, respectively (from left to right). The white stars mark the pulsar positions for each PTA.

(θ, ϕ) = (285°, −54.°341), and the least-constraining limits are obtained at (θ, ϕ) = (78.°75, −19.°471). The spatial variation in the upper limits reflects the distribution and sensitivity of pulsars across the sky in each array. As expected, the most-constrained upper limits are obtained in regions where the pulsar sky coverage is densest. For instance, the lowest upper limits in all three datasets—PPTA DR3, 10 yr EPTA DR2, and 25 yr EPTA DR2—are found in sky regions that have more pulsars. Notably, the most-constrained directions for the 10 and 25 yr EPTA datasets are broadly consistent, suggesting that extended timing baselines improve sensitivity without significantly shifting the sky region of maximum constraint. In contrast, the least-constraining limits tend to occur in regions with sparse or no nearby pulsars, demonstrating the inherent anisotropy in PTA sensitivity. The upper limits as a function of burst sky location from the 25 yr EPTA DR2 dataset are lower than the limits from the 10 yr dataset. Although some of the early EPTA data are noisier, the postfit memory response gains sensitivity from a longer observation time. This is also consistent with other long-timescale signal searches (e.g., ultralight dark matter), where extending the observation baseline tightens the constraints despite the mixed data quality (C. Smarra et al. 2023).

4.2. Search for SMBHB Mergers with Null Memory

We perform a search for SMBHB mergers in PPTA DR3 and 25 yr EPTA DR2 using the NRHybSur3dq8_CCE waveform, which includes strain time series for the inspiral and merger of SMBHBs, the ringdown of the remnant, and the null displacement memory. We find $\ln \mathcal{B} = -0.081$ for 10 yr EPTA DR2, $\ln \mathcal{B} = -0.769$ for 25 yr EPTA DR2, and $\ln \mathcal{B} = -0.679$ for PPTA DR3. Therefore, we find no evidence of a signal. Table 1 summarizes the Bayes factor values obtained from searches on different datasets using the models used in the Letter.

We present our constraints on the chirp masses $\mathcal{M}$ and luminosity distances D_L of merging SMBHBs in Figure 4. We show both the posterior density and the lower limits at 95% credibility on D_L up to which an SMBHB with a given $\mathcal{M}$ can be detected. For comparison, we also show limits based on the estimation of h_0 for the memory-burst model, assuming the burst strain amplitude corresponds to that of an SMBHB merger. In Section 3.3, we mention the scaling $h_0 \propto \mu/r$ (M. S. Pshirkov et al. 2010), where μ is the reduced mass and r is the comoving distance. For Figure 4, we reparameterize this scaling in terms of the source-frame chirp mass and luminosity distance. Using the definitions of (μ, r) and ($\mathcal{M}$, D_L), and assuming the mass ratio $q = 1$, we calculate $\mathcal{M} = (1+z)\mu(1+q)^{4/5}/(q^{2/5})$ and $D_L = (1+z)r$. Since the strain samples inferred from the search are in the observer frame, an additional $(1+z)$ factor is included, to convert the chirp mass to the source frame. Based on the cosmological model from Planck Collaboration et al. (2020), we convert the comoving distance r to the redshift z and then to the luminosity distance D_L. Using this mapping, we convert the upper limits on h_0 obtained from the burst search to lower limits on D_L for each value of $\mathcal{M}$.

Table 1
Bayes Factors

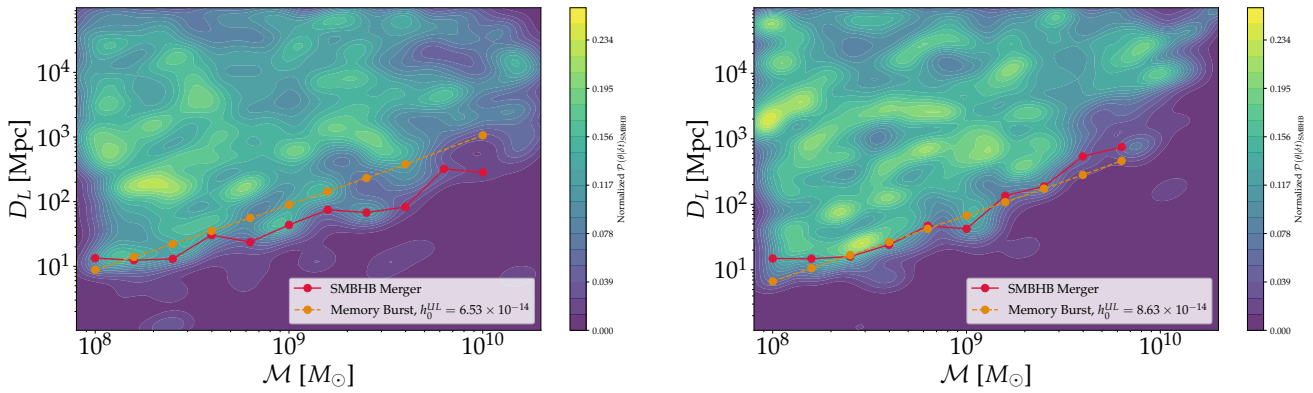
Dataset	$\ln \mathcal{B}$
Memory-burst Model	
EPTA 10 yr	0.712
EPTA 25 yr	−0.309
PPTA DR3	0.432
SMBHB-merger Model	
EPTA 10 yr	−0.081
EPTA 25 yr	−0.769
PPTA DR3	−0.679

Using the same SMBHB memory scaling $h_0 \propto \mu/r$ and the mapping described above, we also convert the posterior samples in ($\mathcal{M}$, D_L, q) obtained with the SMBHB-merger waveform back to memory strain amplitudes h_0. From these samples, we construct upper limits on the net null memory strain amplitude as a function of the SMBHB-merger (burst) time. The results are shown in Figure 5, where they are also compared to the results from the displacement-memory-burst search in Figure 2.

The limits for the NR model are slightly weaker (i.e., higher in strain and, correspondingly, lower in distance) than those from the generic burst model. This behavior arises because the burst model overestimates the observable residuals—it assumes an instant increase in the timing residuals, leading to artificially *constraining* upper limits. In contrast, the physically motivated NR waveform includes the gradual accumulation of memory during the inspiral. Consequently, the postfit residuals for the NR model are smaller than for the burst model, leading to slightly higher upper limits on the strain for the same dataset. Unlike the limits for the memory-burst model, the limits for the NR model do not degrade as quickly toward the edges of the observation time, because of the contribution of the inspiral component of the signal.

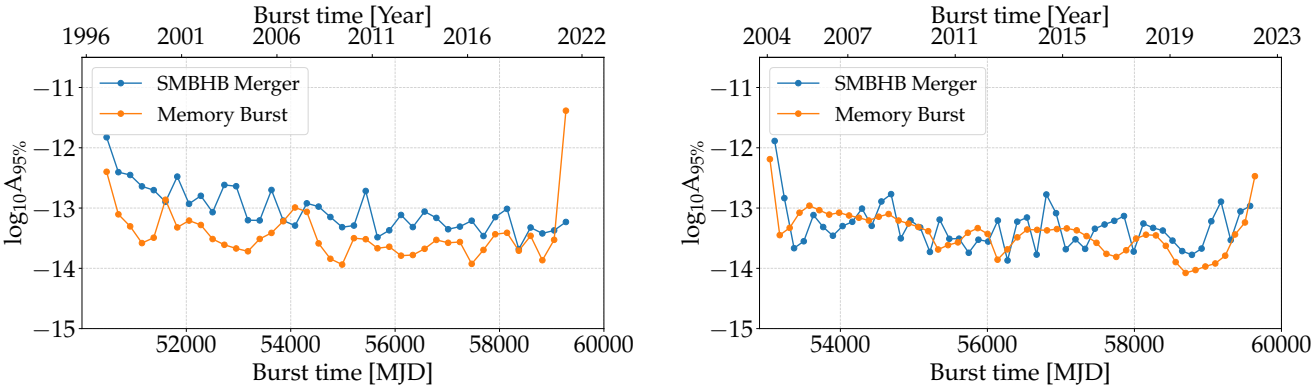
4.3. FP Comparison

We demonstrate the performance of the FP approach with the EPTA 10 yr dataset. We begin with the single-pulsar



(a) 25-year EPTA DR2 (b) PPTA DR3

Figure 4. Lower limits on the luminosity distance D_L as a function of the source-frame chirp mass $\mathcal{M}$. This comparison is valid for equal-mass binaries.



(a) 25-year EPTA DR2 (b) PPTA DR3

Figure 5. Upper limits on the strain amplitude h_0 of generic memory bursts (orange) compared to the SMBHB-merger model (blue) as a function of the burst epoch. The merger model includes the gradual memory buildup during the inspiral, so the postfit residuals are smaller than those of the burst model. As a result, the merger waveform yields slightly weaker upper limits compared to the generic burst model but provides a more accurate representation of the expected memory signal. Unlike the burst model, the merger model limits do not degrade as sharply near the observation edges, due to the contribution of the inspiral component.

analysis, using the burst-with-displacement-memory model to infer the intrinsic per-pulsar memory parameters. The resulting posteriors are then approximated using either KDEs or normalizing flows. The intrinsic memory parameters, such as the per-pulsar strain amplitude and sign, are subsequently mapped to the global parameters defined by the memory burst's strain amplitude, sky location, and polarization angle with Equation (7). The FP is then constructed by evaluating the approximated per-pulsar densities (via KDE or flows) at these projected values and multiplying them across all pulsars, yielding an efficient approximation to the full-PTA likelihood in the global parameter space. Figure 6 compares the posteriors for the memory-burst parameters obtained with four methods on the EPTA 10 yr data: the full-PTA likelihood (blue), the FP with LUTs (green; J. Sun et al. 2023), the FP with KDE (red), and the FP with normalizing flows (black).

To summarize the consistency between the FP approximations and the full-PTA posterior, we report two posterior agreement metrics. First, the *probability of superiority* (or common-language effect size) $P(A > B)$ for each parameter—i.e., the probability that a random draw from the FP posterior exceeds one from the full-PTA posterior (K. O. McGraw & S. Wong 1992; M. Hollander et al. 2014). Values of $P(A > B)$ near 0.5 indicate no systematic shift between the FP posterior and the full-PTA posterior. Values about 0 and 1 point to a decrease or an increase of the average FP posterior sample value, respectively, compared to the full-PTA posterior. Second, we introduce the *fractional overlap of credible intervals* (CIO) at 1σ credibility as a simple heuristic for interval agreement (H. F. Inman & E. L. B. Jr 1989). A larger CIO indicates a stronger interval concordance.

We find the following probability of superiority and CIO values for our three FP posteriors against the full-PTA posterior. For FP with KDE, we obtain a mean probability of superiority $\overline{P(A > B)} = 0.498$ (range 0.418–0.529 across parameters). For FP with normalizing flows, we obtain $\overline{P(A > B)} = 0.498$ (range 0.404–0.6 across parameters). For FP with LUTs, we obtain $\overline{P(A > B)} = 0.499$ (range 0.442–0.595 across parameters). As for the CIO, we find values in the range (0.95–0.99) for all FP approaches and all parameters, except $\log_{10} A_{\rm B}$, indicating an excellent agreement. For $\log_{10} A_{\rm B}$, we find the CIO in the range (0.3–0.4), suggesting a significant difference in the 1σ credible levels obtained with the full-PTA analysis and FP approach. We attribute this to difficulties in the approximation of the tail of the marginal posterior for $\log_{10} A_{\rm B} < -13.5$. However, the peak of the marginalized full-PTA posterior matches well with those obtained with FP approaches. When calculating the CIO for the 3σ credible level for $\log_{10} A_{\rm B}$, we find values of (0.92–0.95). These numbers indicate the close agreement of posterior approximations with the full-PTA result.

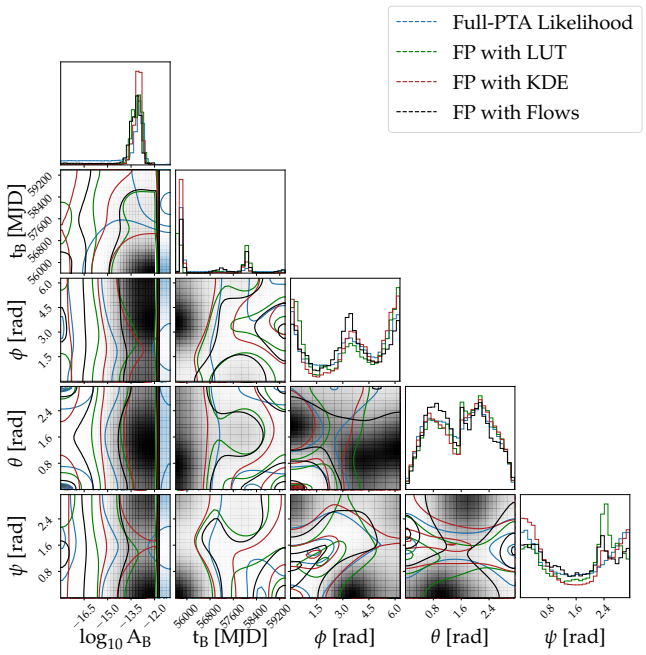


Figure 6. Comparison of posterior distributions for the global memory-burst parameters (amplitude, burst epoch, sky position, and polarization) using three inference methods applied to the 10 yr EPTA dataset. The full-PTA likelihood results are shown in blue, FP with LUTs in green, FP with the KDE approximation in red, and FP with normalizing flows in black.

While the full-PTA likelihood provides the most accurate inference, it comes with a high computational cost. The average time per likelihood evaluation is 2.1984 ± 0.0646 s for the full-PTA likelihood, 1.40 ± 0.076 s for FP with KDE ($\sim$1.6 times faster), 0.323 ± 0.007 s for FP with flows ($\sim$6.8 times faster), and 0.046 ± 0.005 s for FP with LUTs. Such improvements are crucial for enabling scalable Bayesian inference in high-dimensional PTA analyses. To reach this accuracy, LUTs require substantially more single-pulsar samples (i.e., a denser grid), to avoid interpolation artifacts and boundary bias. In contrast, the flows and KDE yield continuous density approximations that remain accurate with fewer samples. LUTs offer extremely fast and deterministic evaluations but are discrete and memory-intensive. The flows and KDE provide smooth differentiable densities that better capture curved posteriors and are easier to compose.

5. Discussion

Sensitivity to merging SMBHBs. The analysis using the SMBHB-merger model builds on the framework established in the methods paper (S. M. Tomson et al. 2025), which demonstrates that memory-inclusive SMBHB-merger waveforms enable PTAs to probe a previously inaccessible region of the binary parameter space—beyond the reach of existing continuous and memory-burst models.

False positives. During the FP analyses, we have realized the importance of modeling the common-spectrum time-correlated stochastic process. The inclusion of this term is not straightforward for single-pulsar analyses from which the FP posterior is derived, because two power-law terms are degenerate in single-pulsar data. The workaround is to enforce a fixed spectral index in the common process term, as is often done in searches for the GWB. Otherwise, we find false positives in our results. The same approach is used in G. Agazie et al. (2025). They have further adopted the model of pulsar-intrinsic time-correlated noise, where, instead of a power-law noise spectrum, the noise power at every frequency is a free parameter. However, we find that it does not rule out the false positives.

Posterior approximation. We note two general forms of mismatch that can arise in approximate posteriors: under-constraining, which typically leads to conservative but acceptable inference, and overconstraining, which can bias results and introduce false positives. In our analysis, the KDE approximation showed mild signs of overconstraining in some parameters, whereas the flow-based approximation remained robust. Overall, we find that both FP-based approximations reproduce the full-PTA likelihood posteriors reasonably well, with the flow model providing the closest match. Nevertheless, we recommend verifying the accuracy of the approximation in future analyses. Additionally, for the EPTA 10 yr dataset, using the SMBHB-merger memory model, modeling HD correlations between pulsars is required. FP approaches, which neglect these correlations, are therefore not suitable for searches that include correlations between the pulsars. Variational inference with normalizing flows, in contrast, is capable of capturing the full correlated posterior structure (M. Vallisneri et al. 2025).

The KDE-based approximation required substantial empirical tuning to achieve comparable results. Due to the structured and sometimes multimodal nature of the posterior in memory-burst searches, naive KDE settings led to either oversmoothed distributions or spurious modes. Through iterative refinement, we found that a relatively small kernel bandwidth of 0.05 provided the best compromise, accurately capturing sharp features while minimizing artifacts. This highlights a limitation of KDE in approximating complex posteriors.

6. Conclusion

We have performed a search for a number of GW displacement memory signals. First, we developed a search for SMBHB mergers, including the signal model with null memory. We performed the search on the PPTA DR3 and 10 and 25 yr EPTA DR2 datasets. Second, we searched for generic bursts of displacement memory in all three datasets. We placed upper limits on the strain amplitude of memory bursts, ruling out the observation of these signals with strain amplitude $h_0 > 10^{-13}$ over a period of 22 yr at 95% credibility. We rule out the existence of generic memory bursts with amplitudes $h_0 > 10^{-14}$ in brief periods of observation time. With the full NR model for SMBHB mergers, we rule out the existence of null memory bursts with $h_0 > 10^{-14}$ over a longer period of about 7 yr. We have also placed lower limits on the luminosity distance D_L for SMBHBs with $\mathcal{M}$ between 10^8 and $10^{10}\,M_\odot$ at 95% credibility. With 25 yr EPTA DR2, we rule out mergers of SMBHBs with $\mathcal{M} = 10^{10}\,M_\odot$ up to 280 Mpc. With PPTA DR3 spanning 18 yr, we rule out mergers of SMBHBs with $\mathcal{M} = 10^{10}\,M_\odot$ up to 700 Mpc.

Acknowledgments

We are grateful to Abhimanyu Sushobhanan and Wang Wei Yu for their insightful scientific discussions and support with data analysis throughout the course of this project. We also thank Bruce Allen for helpful comments. This work was

supported by the Max Planck Gesellschaft (MPG) and the ATLAS cluster computing team at AEI Hannover.

We acknowledge the use of data from the Parkes Pulsar Timing Array (PPTA) and support from the Australian Research Council Centre of Excellence for Gravitational-Wave Discovery (OzGrav; grants CE170100004, CE170100007, and CE23010016). We also extend our gratitude to the European Pulsar Timing Array (EPTA) and the facilities that provide its data: the 100 m radio telescope in Effelsberg operated by the Max-Planck-Institut für Radioastronomie (MPIfR), the Westerbork Synthesis Radio Telescope (operated by ASTRON, with support from the Netherlands Foundation for Scientific Research or NWO), the Nançay Radio Observatory (operated by the Paris Observatory under CNRS and Université d'Orléans, supported through PNCG and PNHE), and the Sardinia Radio Telescope (funded by MIUR, ASI, and RAS; operated by INAF). We acknowledge support from GRAvity from Astrophysical to Microscopic Scales (GRAMS; ERC Consolidator Grant No. 815673), the ERC Synergy Grant "GWSky" (grant No. 101167314), the PRIN 2022 project GUVIRP, and the Marie Sklodowska-Curie Fellowship No. 101007855.

Software: PTMCMCSAMPLER (J. Ellis & R. van Haasteren 2019), ENTERPRISE (J. A. Ellis et al. 2020), FLOZ (R. Srinivasan et al. 2024; github.com/Rahul-Srinivasan/floZ), GWSURROGATE (S. Field et al. 2025), ENTERPRISE_WARP (github.com/bvgoncharov/enterprise_warp).

Appendix A
The Results of Parameter Estimation

In this section, we show the results of the estimation of all parameters for the models and the data we discuss in this work.

Table 2
Priors Used in Our Analysis

Parameter	Prior	Description
	Memory-burst Model	
$\log_{10} h_0$	$\mathcal{U}(-17,\ -10)$	Log-10 Amplitude
	SMBHB-Merger Model	
$\log_{10} \mathcal{M}$ [$M_\odot$]	$\mathcal{U}(8,\ 12)$	Log-10 Chirp Mass
$\log_{10} D_{\rm L}$ [Mpc]	$\mathcal{U}(0,\ 5)$	Log-10 Luminosity Distance
q	$\mathcal{U}(1,\ 7)$	Mass Ratio
	All Models	
ψ	$\mathcal{U}(0,\ \pi)$	Polarization
$\cos\theta$	$\mathcal{U}(-1,\ 1)$	Cosine of Polar Angle
ϕ	$\mathcal{U}(0,\ 2\pi)$	Azimuthal Angle
t_0 [MJD]	$\mathcal{U}(55611,\ 59385)$	10 yr EPTA DR2
	$\mathcal{U}(50360,\ 59385)$	25 yr EPTA DR2
	$\mathcal{U}(50340,\ 59640)$	PPTA DR3

Figure 7 depicts the posteriors of the generic memory-burst search on the EPTA 10 yr and 25 yr datasets and the PPTA DR3 dataset. In Figures 8 and 9, we show the posteriors for all parameters of the NR waveform model in a search for SMBHB mergers with null memory. The priors used for this analysis are shown in Table 2.

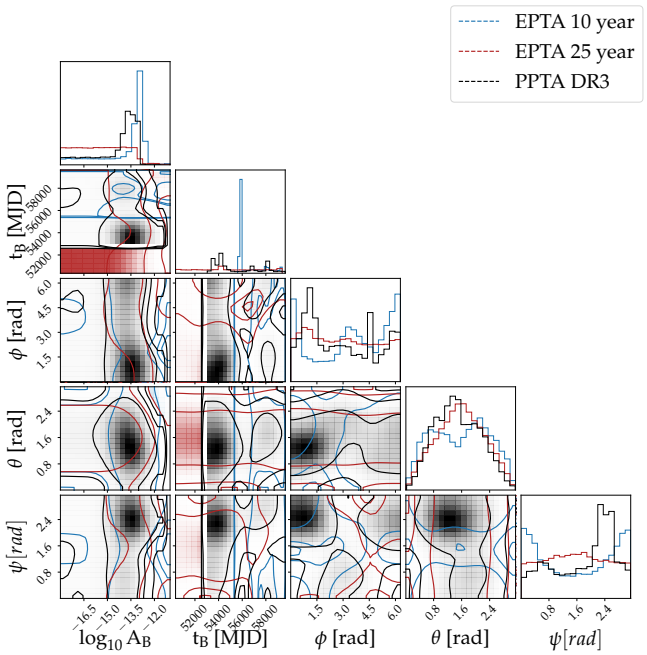
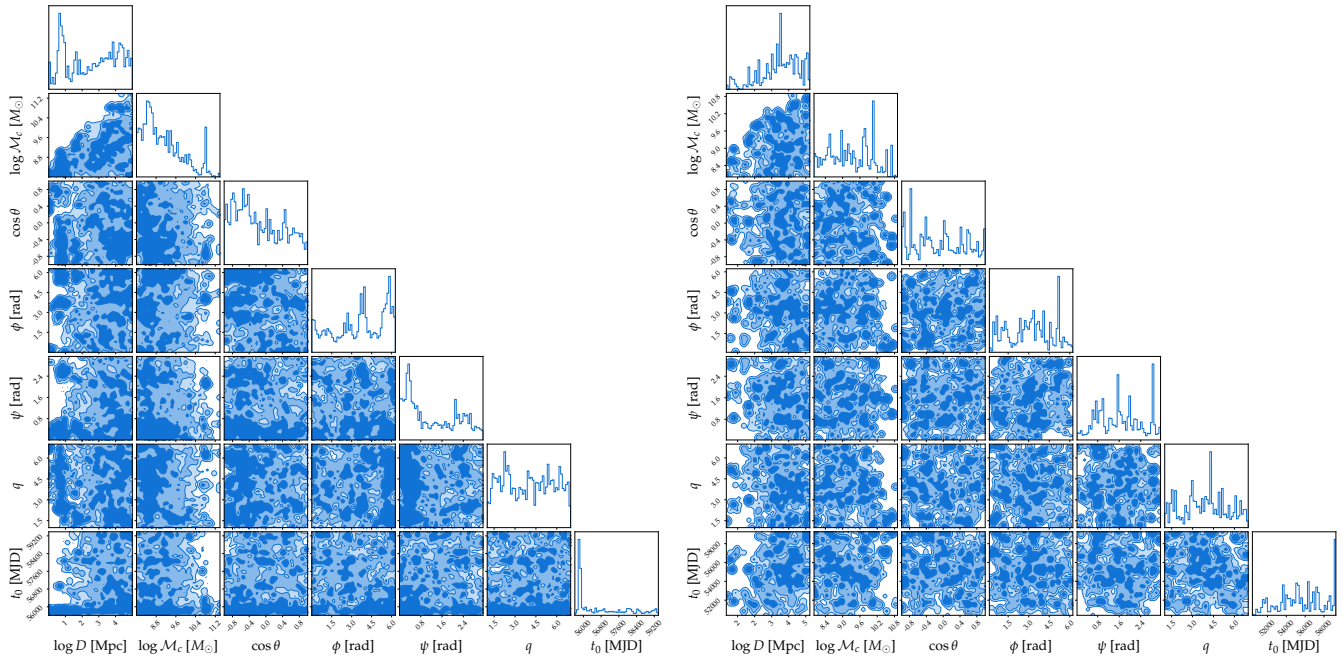


Figure 7. Posterior distributions of GW memory signal parameters from generic memory-burst searches on the 10 yr EPTA (blue), 25 yr EPTA (red), and PPTA DR3 (black) datasets. The prior ranges on the burst epoch t_0[MJD] for each dataset are: $\mathcal{U}(55611,\ 59385)$ for 10 yr EPTA, $\mathcal{U}(50360,\ 59385)$ for 25 yr EPTA, and $\mathcal{U}(50340,\ 59640)$ for PPTA DR3.



(a) 10-year EPTA DR2 (b) 25-year EPTA DR2

Figure 8. Posterior distributions for searches for SMBHB mergers with null memory. The parameters are the chirp mass $\mathcal{M}$, mass ratio q, luminosity distance D_L, burst epoch t_B, R.A., decl., and polarization angle ψ of the expected SMBHB mergers. Panel (a) shows the results for the 10 yr EPTA DR2 dataset, while panel (b) shows those for the 25 yr EPTA DR2 dataset.

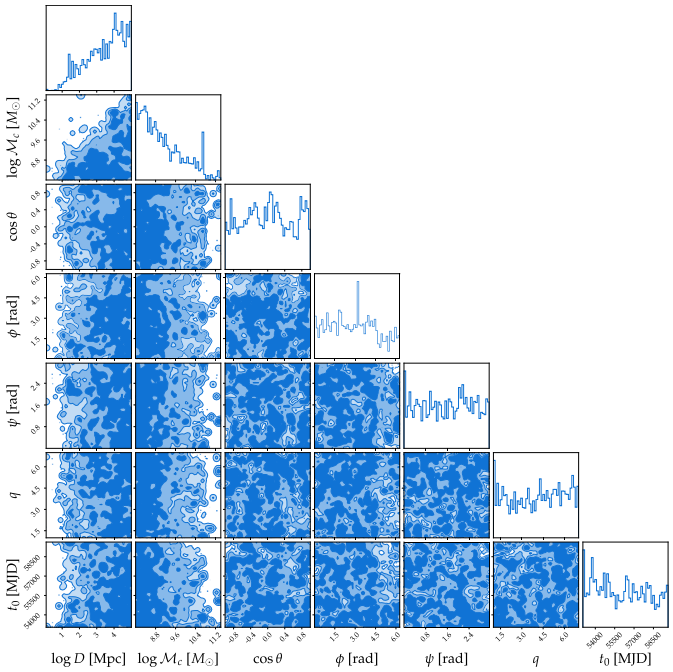


Figure 9. Posterior distributions for searches for SMBHB mergers with null memory on the PPTA DR3 dataset.

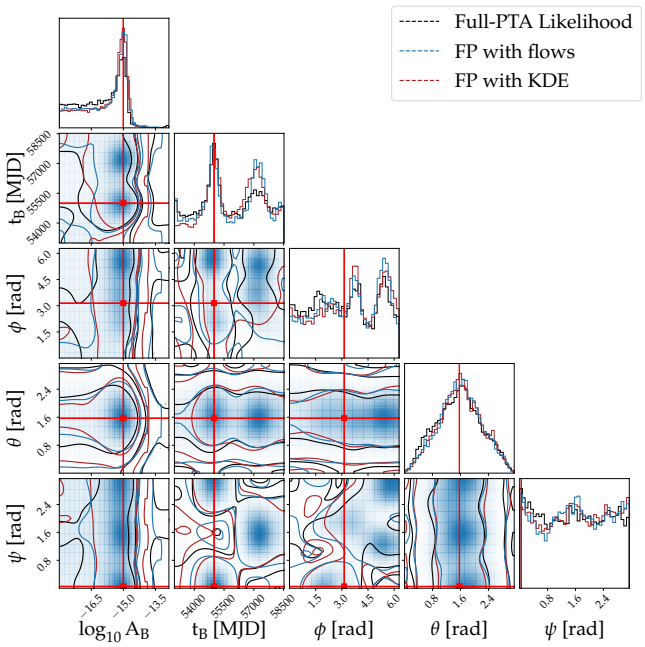


Figure 10. Comparison of posterior distributions for a simulated memory-burst signal obtained using three inference approaches: (1) the full-PTA likelihood computed with the standard enterprise framework (black); (2) the FP approach with KDE (red); and (3) the FP approach with normalizing flows (blue). The vertical and horizontal red lines show the simulated parameter values.

Appendix B
Comparison of Posteriors across Methods

In this section, we compare the posterior distributions obtained using three different approaches: the full-PTA likelihood, the FP approach with KDE, and FP with normalizing flows. Figure 10 depicts a comparison of the distributions for a simulated dataset consisting of 10 pulsars with a simulated generic burst memory. The simulated (true) burst parameters are depicted as the vertical and horizontal lines in the figure. As shown in Figure 10, the results from the flow-based and KDE approximations are generally consistent with the full-PTA result, capturing the overall shape and covariance of the distribution.

The choice of bandwidth in KDE plays a crucial role in balancing bias and variance—an overly large bandwidth can oversmooth features, while a small one may lead to noise amplification. Similarly, for flow-based models, the number of flow layers, hidden units, and coupling-block structure significantly affect the expressiveness and convergence of the model. Careful selection and tuning of these parameters is necessary to ensure accurate density approximation, particularly in multimodal distributions. Flow-based models are particularly well suited to capturing multimodal distributions due to their ability to learn complex nonlinear transformations of the posterior.

Appendix C
Single-pulsar Limits

We show the single-pulsar limits on the memory-burst amplitude for the 10 yr EPTA DR2, 25 yr EPTA DR2, and PPTA DR3 datasets in Figure 11.

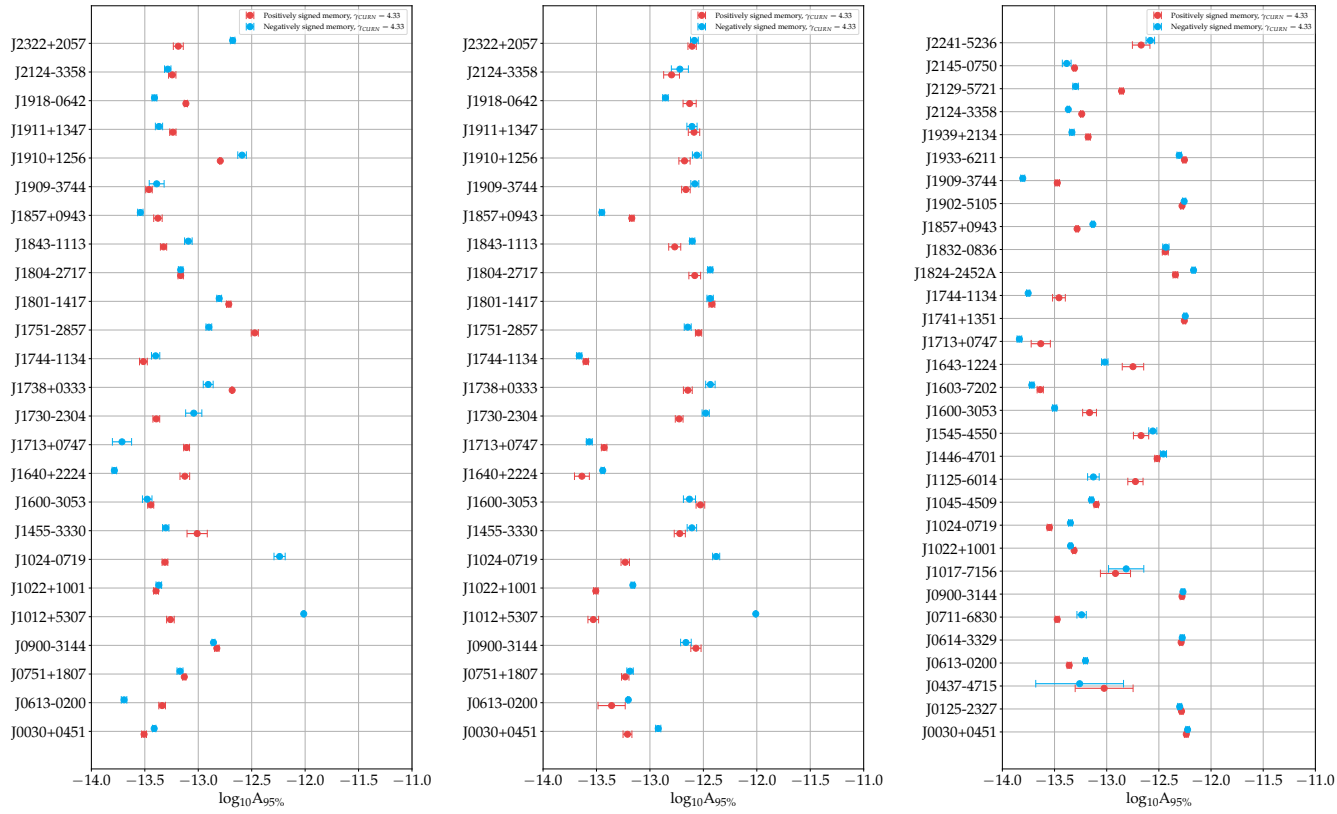


Figure 11. Pulsar-term upper limits on the strain amplitude h_0, marginalizing over the burst epoch as well as each pulsar's red- and white-noise parameters. Panels (a)–(c) show the results for the EPTA 10 yr, EPTA 25 yr, and PPTA DR3 datasets, respectively (from left to right).

ORCID iDs

Sharon Mary Tomson https://orcid.org/0000-0001-7603-1637
Boris Goncharov https://orcid.org/0000-0003-3189-5807
Rutger van Haasteren https://orcid.org/0000-0002-6428-2620
Rahul Srinivasan https://orcid.org/0000-0002-7176-6690
Enrico Barausse https://orcid.org/0000-0001-6499-6263
Yirong Wen https://orcid.org/0009-0006-4212-3801
Jingbo Wang https://orcid.org/0000-0001-9782-1603
John Antoniadis https://orcid.org/0000-0003-4453-3776
N. D. Ramesh Bhat https://orcid.org/0000-0002-8383-5059
Zu-Cheng Chen https://orcid.org/0000-0001-7016-9934
Ismael Cognard https://orcid.org/0000-0002-1775-9692
Valentina Di Marco https://orcid.org/0000-0003-3432-0494
Huanchen Hu https://orcid.org/0000-0002-3407-8071
Gemma H. Janssen https://orcid.org/0000-0003-3068-3677
Michael Kramer https://orcid.org/0000-0002-4175-2271
Wenhua Ling https://orcid.org/0009-0009-9142-6608
Kuo Liu https://orcid.org/0000-0002-2953-7376
Saurav Mishra https://orcid.org/0009-0001-5633-3512
Delphine Perrodin https://orcid.org/0000-0002-1806-2483
Andrea Possenti https://orcid.org/0000-0001-5902-3731
Christopher J. Russell https://orcid.org/0000-0002-1942-7296
Ryan M. Shannon https://orcid.org/0000-0002-7285-6348
Gilles Theureau https://orcid.org/0000-0002-3649-276X
Shuangqiang Wang https://orcid.org/0000-0003-4498-6070

References

Agazie, G., Anumarlapudi, A., Archibald, A. M., et al. 2023, ApJL, 951, L8
Agazie, G., Anumarlapudi, A., Archibald, A. M., et al. 2025, ApJ, 987, 5
Agazie, G., Arzoumanian, Z., Baker, P. T., et al. 2024, ApJ, 963, 61
Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2020, ApJ, 889, 38
Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, MNRAS, 510, 4873
Antoniadis, J., Babak, S., Bak Nielsen, A.-S., et al. 2023, A&A, 678, A48
Arzoumanian, Z., Baker, P. T., Blumer, H., et al. 2020, ApJL, 905, L34
Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2016, ApJ, 821, 13
Bieri, L. 2021, PhRvD, 103, 024043
Bieri, L., & Garfinkle, D. 2013, CQGra, 30, 195009
Chen, S., Caballero, R. N., Guo, Y. J., et al. 2021, MNRAS, 508, 4970
Chen, S., Xu, H., Guo, Y., et al. 2025, A&A, 699, A165
Christodoulou, D. 1991, PhRvL, 67, 1486
Cordes, J. M., & Jenet, F. A. 2012, ApJ, 752, 54
Dandapat, S., Susobhanan, A., Dey, L., et al. 2024, PhRvD, 109, 103018
Detweiler, S. 1979, ApJ, 234, 1100
Elhashash, A., & Nichols, D. A. 2025, PhRvD, 111, 044052
Ellis, J., & van Haasteren, R., 2019 PTMCMCSampler: Parallel tempering MCMC sampler package written in Python, Astrophysics Source Code Library, ascl:1912.017
Ellis, J. A., Vallisneri, M., Taylor, S. R., & Baker, P. T., 2020 ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE, v3.0.0, Zenodo, doi:10.5281/zenodo.4059815
Enoki, M., Inoue, K. T., Nagashima, M., & Sugiyama, N. 2004, ApJ, 615, 19
EPTA Collaboration, InPTA Collaboration, Antoniadis, J., et al. 2023, A&A, 678, A50
EPTA Collaboration and InPTA Collaboration, Antoniadis, J., Arumugam, P., et al. 2023, A&A, 678, A49
Favata, M. 2009, ApJL, 696, L159
Favata, M. 2010, CQGra, 27, 084036
Field, S., Varma, V., Blackman, J., et al. 2025, JOSS, 10, 7073
Flanagan, É. É., Grant, A. M., Harte, A. I., & Nichols, D. A. 2019, PhRvD, 99, 084044
Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300

Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, Bayesian Data Analysis (3rd ed.) (Chapman & Hall/CRC)
Goncharov, B., Donnay, L., & Harms, J. 2024, PhRvL, 132, 241401
Goncharov, B., Reardon, D. J., Shannon, R. M., et al. 2021a, MNRAS, 502, 478
Goncharov, B., Sardana, S., Sesana, A., Tomson, S. M., et al. 2025, Nat. Commun., 16, 9692
Goncharov, B., Shannon, R. M., Reardon, D. J., et al. 2021b, ApJL, 917, L19
Grant, A. M., & Mitman, K. 2024, CQGra, 41, 175003
Heavens, A., Fantaye, Y., Mootoovaloo, A., et al. 2017, arXiv:1704.03472
Hobbs, G., Manchester, R. N., Dunning, A., et al. 2020, PASA, 37, e012
Hollander, M., Wolfe, D., & Chicken, E. 2014, Nonparametric Statistical Methods (Wiley)
Inman, H. F., & Jr, E. L. B. 1989, CSTM, 18, 3851
Islo, K., Simon, J., Burke-Spolaor, S., & Siemens, X. 2019, arXiv:1906.11936
Jenkins, A. C., & Sakellariadou, M. 2021, CQGra, 38, 165004
Rezende, D. J., & Mohamed, S. 2015, arXiv:1505.05770
Lamb, W. G., Taylor, S. R., & van Haasteren, R. 2023, PhRvD, 108, 103019
McGraw, K. O., & Wong, S. 1992, PsycB, 111, 361
Miles, M. T., Shannon, R. M., Reardon, D. J., et al. 2025, MNRAS, 536, 1489
Mitman, K., Boyle, M., Stein, L. C., et al. 2024, CQGra, 41, 223001
Mitman, K., Moxon, J., Scheel, M. A., et al. 2020, PhRvD, 102, 104007
Nichols, D. A. 2018, PhRvD, 98, 064032
Papamakarios, G., Nalisnick, E., Jimenez Rezende, D., Mohamed, S., & Lakshminarayanan, B. 2021, J. Mach. Learn. Res., 22, 6457
Pasterski, S., Strominger, A., & Zhiboedov, A. 2016, JHEP, 2016, 53
Planck Collaboration, Aghanim, N., & Akrami, Y. 2020, A&A, 641, A6
Pshirkov, M. S., Baskaran, D., & Postnov, K. A. 2010, MNRAS, 402, 417
Ravi, V., Wyithe, J. S. B., Shannon, R. M., & Hobbs, G. 2015, MNRAS, 447, 2772
Reardon, D. J., Zic, A., Shannon, R. M., et al. 2023a, ApJL, 951, L6
Reardon, D. J., Zic, A., Shannon, R. M., et al. 2023b, ApJL, 951, L7
Sazhin, M. V. 1978, SvA, 22, 36
Scott, D. W. 2015, Multivariate Density Estimation: Theory, Practice, and Visualization (John Wiley and Sons)
Seto, N. 2009, MNRAS, 400, L38
Silverman, B. W. 1986, Density Estimation for Statistics and Data Analysis (Chapman and Hall)
Smarra, C., Goncharov, B., Barausse, E., et al. 2023, PhRvL, 131, 171001
Spiewak, R., Bailes, M., Miles, M. T., et al. 2022, PASA, 39, e027
Srinivasan, R., Crisostomi, M., Trotta, R., Barausse, E., & Breschi, M. 2024, PhRvD, 110, 123007
Strominger, A., & Zhiboedov, A. 2016, J. High Energ. Phys., 2016, 86
Sun, J., Baker, P. T., Johnson, A. D., Madison, D. R., & Siemens, X. 2023, ApJ, 951, 121
Thorne, K. S. 1992, PhRvD, 45, 520
Tomson, S. M., Goncharov, B., & van Haasteren, R. 2025, arXiv:2510.04537
Vallisneri, M., Crisostomi, M., Johnson, A. D., & Meyers, P. M. 2025, PhRvL, 135, 071401
van Haasteren, R., & Levin, Y. 2010, MNRAS, 401, 2372
van Haasteren, R., Levin, Y., McDonald, P., & Lu, T. 2009, MNRAS, 395, 1005
Wang, J. B., Hobbs, G., Coles, W., et al. 2015, MNRAS, 446, 1657
Weinberg, S. 1964, PhRv, 135, 1049
Xu, H., Chen, S., Guo, Y., et al. 2023, RAA, 23, 075024
Yoo, J., Mitman, K., Varma, V., et al. 2023, PhRvD, 108, 064027
Zel'dovich, Y. B., & Polnarev, A. G. 1974, SvA, 18, 17
Zic, A., Reardon, D. J., Kapur, A., et al. 2023, PASA, 40, e049