

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society © 2022 The Author(s) Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

McGrath, Casey, Daniel J D’Orazio and Jolien Creighton. "Measuring the Hubble constant with double gravitational wave sources in pulsar timing" Monthly Notices of the Royal Astronomical Society 517, no. 1 (November 2022): 1242–1263. <https://doi.org/10.1093/mnras/stac2593>.

<https://doi.org/10.1093/mnras/stac2593>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it’s important to you. Thank you.

Measuring the Hubble constant with double gravitational wave sources in pulsar timing

Casey McGrath¹,^{1,2,3}★ Daniel J. D’Orazio¹⁴★ and Jolien Creighton¹⁵★

¹Center for Space Sciences and Technology, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

²Gravitational Astrophysics Lab, NASA/GSFC, Greenbelt, MD 20771, USA

³Center for Research and Exploration in Space Science and Technology II, NASA/GSFC, Greenbelt, MD 20771, USA

⁴Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

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

Accepted 2022 September 9. Received 2022 September 9; in original form 2022 August 12

ABSTRACT

Pulsar timing arrays (PTAs) are searching for gravitational waves from supermassive black hole binaries (SMBHBs). Here we show how future PTAs could use a detection of gravitational waves from individually resolved SMBHB sources to produce a purely gravitational wave-based measurement of the Hubble constant. This is achieved by measuring two separate distances to the same source from the gravitational wave signal in the timing residual: the luminosity distance D_L through frequency evolution effects, and the parallax distance D_{par} through wavefront curvature (Fresnel) effects. We present a generalized timing residual model including these effects in an expanding universe. Of these two distances, D_{par} is challenging to measure due to the pulsar distance wrapping problem, a degeneracy in the Earth-pulsar distance and gravitational wave source parameters that requires highly precise, sub-parsec level, pulsar distance measurements to overcome. However, in this paper we demonstrate that combining the knowledge of two SMBHB sources in the timing residual largely removes the wrapping cycle degeneracy. Two sources simultaneously calibrate the PTA by identifying the distances to the pulsars, which is useful in its own right, and allow recovery of the source luminosity and parallax distances which results in a measurement of the Hubble constant. We find that, with optimistic PTAs in the era of the Square Kilometre Array, two fortuitous SMBHB sources within a few hundred Mpc could be used to measure the Hubble constant with a relative uncertainty on the order of 10 per cent.

Key words: gravitational waves – quasars: supermassive black holes – pulsars: general – cosmological parameters.

1 INTRODUCTION

Pulsar timing arrays (PTAs) are currently searching for gravitational waves with frequencies $\mathcal{O}(1 \text{ to } 100 \text{ nHz})$ produced by supermassive black hole binaries (SMBHBs) at the centres of coalescing galaxies. These experiments time highly regular millisecond pulsars across our Galaxy and look for irregularities with the arrival times of the pulsar’s pulses. A continuous gravitational wave source is expected to cause these arrival times to periodically drift in and out of synchronization with a reference clock. While the first gravitational wave detection using this experimental method may be a stochastic background of unresolved sources (see, however, Kelley et al. 2018), these experiments are expected to become more sensitive to the point where loud individual sources will be resolved and parameter estimation will be able to extract the source parameters from the data.

Of notable interest is the ability of a gravitational wave-based experiment to measure cosmological parameters such as the Hubble constant H_0 . These new gravitational wave-based measurements are important in helping us to resolve the current tension between experimental measurements of H_0 (The LIGO Scientific Collaboration, the Virgo Collaboration, the 1M2H Collaboration, the Dark Energy Camera GW-EM Collaboration & the DES Collaboration 2017; The LIGO Scientific Collaboration, the Virgo Collaboration & the KAGRA Collaboration 2021; Feeney et al. 2019, and references therein). In this study we are motivated by the recent work of D’Orazio & Loeb (2021) and McGrath & Creighton (2021) (hereafter DL21 and MC21). DL21 demonstrated how a measurement of the Hubble constant could be made from a purely gravitational wave-based method, by measuring the source’s luminosity distance D_L and it’s comoving distance D_c . In this approach D_L is recovered from frequency evolution in the timing residual signal, and D_c comes from probing the curvature of the wavefront across the Earth-pulsar baseline of the PTA experiment. In this paper, we show that more generally, the distance measured from the curvature of the wavefront is the ‘parallax distance’ D_{par} , which is equivalent to D_c in a flat universe. MC21 generalized the current gravitational wave timing residual models by classifying these wavefront curvature effects into the ‘Fresnel regime,’ and they studied how well this new distance parameter

* E-mail: casey.d.mcgrath@nasa.gov (CM); daniel.dorazio@nbi.ku.dk (DJD); jolien@uwm.edu (JC)

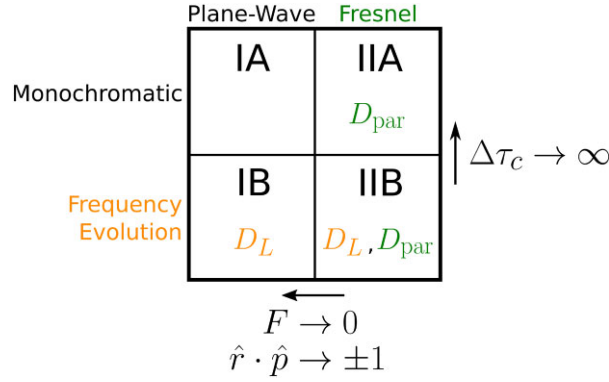


Figure 1. A classification summary of the four gravitational wave regimes, increasing in generality from left-to-right and top-to-bottom. The frequency evolution regime reduces to the monochromatic regime in the large coalescence time limit ($\Delta\tau_c \rightarrow \infty$), and the Fresnel regime reduces to plane-wave regime in either the small Fresnel number limit ($F \rightarrow 0$) or the natural plane-wave limit ($\hat{r} \cdot \hat{p} \rightarrow \pm 1$). Importantly, in a cosmologically expanding universe the frequency evolution regimes allow for the direct measurement of the source luminosity distance D_L , and the Fresnel regimes allow for the direct measurement of the source parallax distance D_{par} . In the fully general IIB model, both of these distances can be independently measured, and therefore through equation (7) a value of H_0 can be inferred.

could be recovered for different PTA constraints and source parameters. Both of these studies were strongly motivated by and synthesized the previous work of Deng & Finn (2011) and Corbin & Cornish (2010).

For comparison, the ‘standard sirens’ approach to measuring H_0 is a hybrid technique requiring a source observable via both gravitational wave and electromagnetic messengers (Schutz 1986; Holz & Hughes 2005). Here the luminosity distance D_L is measured via gravitational waves from a chirping source, while the redshift z is measured from the host galaxy through electromagnetic observations. It has been showed that future PTAs may also be able to make measurements of H_0 through this technique (Wang et al. 2022). Combining these measurements of D_L and z allows an inference of H_0 to be made. As an example of a purely gravitational wave-based measurement of H_0 , gravitational wave signals from binary neutron star, or neutron star black hole mergers could be used to obtain both measurements of the source’s luminosity distance D_L and redshift z (Messenger & Read 2012; Ghosh, Biswas & Bose 2022; Shiralilou et al. 2022). This would require well constrained knowledge of the neutron star equation of state from numerous detections, which may be possible with future Einstein Telescope and Cosmic Explorer-era gravitational wave observatories. The approach presented in this paper is also purely gravitational wave-based, but here we trade the redshift measurement for a second distance measurement made from the Fresnel wavefront curvature effects, and we do not require any intrinsic knowledge of rest-frame source properties.

In this study, we take the models developed in MC21 and further generalize them to a cosmologically expanding universe. We apply the same Bayesian framework and methods described in that study, in order to predict how well future PTA experiments may be able to measure the Hubble constant. When detecting a single SMBHB source, we find that in order to recover the parallax distance parameter (and hence measure H_0), we require highly accurate measurements of the distances to the pulsars in our array. But crucially, we find that when detecting two SMBHB sources simultaneously, no prior knowledge on the pulsar distances is required in order to recover D_{par} and H_0 . Multiple simultaneous continuous wave SMBHBs have not been widely investigated within PTA research (some studies include Babak & Sesana 2012; Petiteau et al. 2013; Qian, Mohanty & Wang 2022), therefore this work helps us to motivate a particular insight that multiple sources can provide over just a single source. The results presented in this work also demonstrate how the same methods used to measure H_0 do so through improved measurements of the pulsar distances, which is an important result in its own right. In this paper, we focus on the H_0 measurement, but an additional paper (McGrath, D’Orazio, & Creighton, in preparation) will provide greater details on the pulsar distance recovery.

This paper is outlined as follows. Section 2.1 presents the theoretical background and generalization of our timing residual models, which shows how the Hubble constant enters the pulsar timing model. Section 3 explains the methods used to estimate system parameters and H_0 from mock observations. Section 4 presents the primary obstacle to a practical H_0 measurement, the pulsar distance wrapping problem, and our solution. Section 5 presents the main results, simulated measurements of the Hubble constant for different mock PTAs. Finally, conclusions and future directions are presented in Section 6.

2 THEORETICAL BACKGROUND

2.1 Fresnel and frequency evolving regimes

To detect and extract information from gravitational waves using PTAs, one must model the pulse-time-of-arrival deviations induced by a gravitational wave passing through the array. Such gravitational wave-induced timing residual models can be classified based on two physical qualities/assumptions of the model: the shape of the incoming gravitational wavefront, and its frequency evolution. Following MC21, Fig. 1 categorizes four distinct model regimes. Models that assume no wavefront curvature over the Earth-pulsar baseline make the ‘plane-wave’ assumption (I), while models which account for the first-order curvature terms are classified as ‘Fresnel’ models (II). Then, if the frequency

of the gravitational wave is assumed constant over the thousands of years it takes a photon to travel from the pulsar to the Earth, the model is ‘monochromatic’ (A); otherwise it is in a ‘frequency evolving’ (B) regime.

The result is four distinct timing residual models, increasing in generality from left-to-right, and top-to-bottom. While simpler models are mathematically and computationally simpler and less expensive to calculate, they may produce inaccurate predictions of the timing residual if the source is producing gravitational waves which fall into a more complicated regime. The appropriate limits wherein one model approaches another are governed by the coalescence time $\Delta\tau_c$ (equation 10) of the source, the Fresnel number F (equation 5), and the relative source-pulsar-Earth orientation ($\hat{r} \cdot \hat{p}$; see equation 11). Therefore considering the relative size of these quantities is crucial in deciding when one model’s prediction will fail in comparison to another.

All four models are provided and studied explicitly in MC21, which critically assumed a flat, static universe. However, DL21 show that in a cosmologically expanding universe, two distinct distances appear in the gravitational wave-induced pulsar timing residual model – the luminosity distance D_L and the comoving coordinate distance D_c . Accounting for frequency evolution effects (row B, Fig. 1) in the timing model allows the direct measurement of D_L , while accounting for Fresnel effects (column II, Fig. 1) allows the direct measurement of the parallax distance D_{par} , which reduces to D_c in a flat universe. The goal of this paper is to remove the assumption that the universe is static on cosmological scales, re-derive the formulae of MC21 in a cosmologically expanding universe, and study implications for recovering H_0 . We provide a general list of assumptions made in this work in Appendix C.

2.2 Incorporating cosmological expansion and the Hubble constant

In a flat static universe we have eight gravitational wave source parameters for a circular binary SMBHB system in the Newtonian regime, $\{R, \theta, \phi, \iota, \psi, \theta_0, \mathcal{M}, \omega_0\}$. These are the Earth-source comoving coordinate distance (for a flat static universe), sky angles, orientation Euler angles (inclination and polarization), initial phase (which also can be interpreted as the third orientation Euler angle), system chirp mass $\mathcal{M} \equiv (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$ (for component masses m_1 and m_2), and initial (angular) orbital frequency. The pulsar parameters are $\{L, \theta_p, \phi_p\}$, which are the Earth-pulsar distance, and pulsar sky angles.

In our cosmological framework we divide space–time into two types of frames: a ‘global cosmological frame’ where the background metric is Friedmann–Lemaître–Robertson–Walker (FLRW), and ‘local frames’ where the background metric is Minkowski (see Appendix C, assumption (iii)). The gravitational waves (as described by the metric perturbation below) are generated in the local source frame, propagate over the cosmological frame, and reach the Milky Way Galaxy where we assume they are in the local observer frame. Therefore the cosmological effects of an expanding universe only need to be considered during the gravitational wave’s propagation between the local source and observer frames.

We begin by assuming that the FLRW metric describes our global space–time background between the Earth and our gravitational wave source:

$$\begin{aligned} ds^2 &= -c^2 dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right], \\ &= a^2(\eta) \left[-c^2 d\eta^2 + d\chi^2 + r(\chi)^2 d\Omega^2 \right], \end{aligned} \quad (1)$$

where $d\Omega^2 \equiv d\theta^2 + \sin^2(\theta)d\phi^2$, a is the universe scale factor, and k is the space–time curvature constant. The coordinates (t, r, θ, ϕ) are the comoving coordinates, while the version shown in the second line is made through the change to conformal time $d\eta \equiv \frac{dt}{a}$ (i.e. $\eta \equiv \int^t \frac{dt'}{a}$) and the spatial coordinate transformation:

$$r(\chi) \equiv \begin{cases} \frac{1}{\sqrt{k}} \sin(\sqrt{k}\chi) & , \quad k > 0 \\ \chi & , \quad k = 0 \\ \frac{1}{\sqrt{|k|}} \sinh(\sqrt{|k|}\chi) & , \quad k < 0. \end{cases} \quad (2)$$

Here we are using the convention that the curvature parameter carries units of $[\text{length}^{-2}]$, where $k > 0$, $k = 0$, or $k < 0$ for closed, flat, and open spatial geometries, respectively. This implies that our time-dependent scale factor a is unitless. We also choose to normalize $a(t_0) \equiv 1$ at the present day time t_0 , but we will often write it explicitly in our derivations here. Therefore, the Gaussian curvature of a spatial slice of our universe at the present time is k .

In this framework, without an independent measurement of the source’s redshift (either from electromagnetic observations, or as we will explain, from effects of the curvature of the wavefront itself), the redshift parameter z cannot be disentangled from the source rest frame R (now the transverse comoving distance; see Appendix A), \mathcal{M} , and ω_0 parameters (see for example, Maggiore 2008). The result is that we can only observe the luminosity distance $D_L = (1 + z)R$, a redshifted system chirp mass $\mathcal{M}_{\text{obs}} \equiv (1 + z)\mathcal{M}$, and a redshifted orbital frequency parameter $\omega_{0,\text{obs}} = \frac{\omega_0}{(1+z)}$.

As the gravitational wave propagates through the FLRW universe, the wavefront is traced by the retarded time. Therefore in order to understand time delay effects of the wave arriving at our pulsar compared to the time delay effects of the wave arriving at the Earth, we must calculate the expression for the retarded time of the wave at a field position \vec{x} for a source at the location \vec{x}' (see for example, Caldwell 1993). Crucially, if we make the assumption that the a does not evolve appreciably over the time it takes a photon to travel from our pulsar to the

Earth (see assumption (ii)), then $\eta_{\text{ret}} - \eta \approx \frac{1}{a(t_0)} (t_{\text{ret}} - t)$ and we can write:

$$\eta_{\text{ret}} = \eta - \frac{|\vec{x}' - \vec{x}|}{c} \\ \therefore t_{\text{ret}} \approx t - a(t_0) \frac{|\vec{x}' - \vec{x}|}{c}. \quad (3)$$

Next we must calculate $|\vec{x}' - \vec{x}|$ in the three possible spatial geometries of our universe. The details of this are shown in Appendix B, and the result is that regardless of the background curvature of the spatial slices of our cosmology, the generalized expression for the retarded time evaluated with the field position set to the pulsar's location is:

$$t_{\text{ret}}(t, \vec{x} = L\hat{p}) \approx t - \underbrace{\frac{D_c}{c}}_{\text{"Far Field"}} + \underbrace{(\hat{r} \cdot \hat{p}) a(t_0) \frac{L}{c}}_{\text{"Plane-Wave"}} - \underbrace{\frac{1}{2} (1 - (\hat{r} \cdot \hat{p})^2) a^2(t_0) \frac{L}{c} \frac{L}{D_{\text{par}}}}_{\text{"Fresnel"}} + \dots \quad (4)$$

The Fresnel term gives the first-order curvature of the physical wavefront, and introduces an additional time delay to the arrival time of the wavefront as predicted by the plane-wave approximation. A useful proxy for determining when this term in the expansion will contribute a significant correction to the plane-wave approximation is the Fresnel number (see MC21 and DL21 for more discussion of this quantity):

$$F \equiv \frac{L^2}{\lambda_{\text{gw,obs}} D_{\text{par}}} \approx 0.0003 \left(\frac{\omega_{0,\text{obs}}}{1 \text{ nHz}} \right)^2 \left(\frac{L}{1 \text{ kpc}} \right)^2 \left(\frac{100 \text{ Mpc}}{D_{\text{par}}} \right), \quad (5)$$

$$\text{and } \lambda_{\text{gw,obs}} \equiv \frac{\pi c}{\omega_{0,\text{obs}}} \approx (30.5 \text{ pc}) \left(\frac{1 \text{ nHz}}{\omega_{0,\text{obs}}} \right). \quad (6)$$

There are two notable results that come out of the expression in equation (4). First, we find that two cosmological distance measurements to the same source appear in the 'Fresnel' term in the expansion – the line-of-sight comoving distance D_c , and the parallax distance D_{par} . In practice, however, we can only measure D_{par} from our timing residual models. This is because in addition to $\frac{D_c}{c}$ appearing as the first term in the expression for the retarded time measured at the pulsar, it also appears in the expression for the retarded time measured at the Earth (for the Earth, just set $L = 0$ in equation 4). Therefore, when the rest of our timing residual model is worked out (see for example the IIB model in equations 17 and 18), we can simply choose the fiducial time $t_{0,\text{obs}} = -\frac{D_c}{c}$ and all dependence on D_c vanishes (this point is discussed in MC21). Second, is that in principle we could also choose to include the cosmological curvature constant k as a parameter to attempt to directly measure it from the gravitational wave signal. Equation (A10) provides the connection between D_{par} and D_L , and therefore would introduce both z and k as additional model parameters. However, since the Fresnel corrections are already smaller order corrections and k is currently understood to be very close to zero, we will restrict our attention in this work to a geometrically flat universe and assume $k = 0$.

Therefore, working under the assumption of a geometrically flat universe, the relationships given in Appendix A provide us the connection to the Hubble constant:

$$k = 0 \quad \longrightarrow \quad \begin{cases} D_L = (1 + z) D_{\text{par}}, \\ H_0 = \frac{c}{R} \int_0^z \frac{dz'}{E(z')} = \frac{c}{D_{\text{par}}} \int_0^{D_L/D_{\text{par}}-1} \frac{dz'}{E(z')}. \end{cases} \quad (7)$$

By procuring a measurement of both the distances D_L and D_{par} from our pulsar timing model (see Section 2.3.1), we can directly measure the source's redshift. By then assuming values of the cosmological density parameters which appear in the Hubble function E (see equation A2), or by using the small redshift approximation (see the discussion around equation 20), we can directly measure the Hubble constant H_0 . In this work we simulate a flat Λ CDM universe using the density parameters values $\Omega_r = 0$, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $\Omega_k = 0$, and a Hubble constant value of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2.3 Constructing the gravitational wave timing residual

We now derive the response of a PTA to the passing of a gravitational wave in the most general regime (IIB of Fig. 1), in an expanding universe. Throughout, the notation and conventions used in this paper mirror those in MC21. The main pedantic distinction is the 'obs' subscript on many of the quantities, simply to remind the reader that unlike in MC21, here there is a difference between source frame quantities and 'observer' frame quantities.

The metric perturbation produced by a circular binary system can be written as (Maggiore 2008; Creighton & Anderson 2011):

$$\begin{cases} h_+(t_{\text{obs}}) & \equiv -h(t_{\text{obs}}) \cos(2\Theta(t_{\text{obs}})), \\ h_\times(t_{\text{obs}}) & \equiv -h(t_{\text{obs}}) \sin(2\Theta(t_{\text{obs}})), \\ h(t_{\text{obs}}) & \equiv \frac{4(G\mathcal{M}_{\text{obs}})^{5/3}}{c^4 D_L} \omega(t_{\text{obs}})^{2/3}, \\ h_{0,\text{obs}} & \equiv \frac{4(G\mathcal{M}_{\text{obs}})^{5/3}}{c^4 D_L} \omega_{0,\text{obs}}^{2/3}, \end{cases} \quad (8)$$

for 'plus' and 'cross' polarizations, and where t_{obs} denotes the observer's clock time. The angular phase and frequency functions $\Theta(t_{\text{obs}})$ and $\omega(t_{\text{obs}})$ are defined by the monochromatic regime (A) or the frequency evolving regime (B), depending on which model we desire:

$$\text{Monochromaticism}_{(A)} \quad \begin{cases} \omega(t_{\text{obs}}) = \omega_{0,\text{obs}}, \\ \Theta(t_{\text{obs}}) = \theta_0 + \omega_{0,\text{obs}}(t_{\text{obs}} - t_{0,\text{obs}}), \end{cases} \quad (9)$$

$$\text{Frequency Evolution}_{(B)} \begin{cases} \omega(t_{\text{obs}}) = \omega_{0,\text{obs}} \left[1 - \frac{t_{\text{obs}} - t_{0,\text{obs}}}{\Delta\tau_{c,\text{obs}}} \right]^{-3/8}, \\ \Theta(t_{\text{obs}}) = \theta_0 + \theta_c \left[1 - \left(\frac{\omega(t_{\text{obs}})}{\omega_{0,\text{obs}}} \right)^{-5/3} \right], \\ \Delta\tau_{c,\text{obs}} \equiv \frac{5}{256} \left(\frac{c^3}{G\mathcal{M}_{\text{obs}}} \right)^{5/3} \frac{1}{\omega_{0,\text{obs}}^{8/3}} = (1+z)\Delta\tau_c, \\ \theta_c = \frac{1}{32} \left(\frac{c^3}{G\mathcal{M}_{\text{obs}}\omega_{0,\text{obs}}} \right)^{5/3} \equiv \frac{8}{5} \Delta\tau_{c,\text{obs}} \omega_{0,\text{obs}} = \frac{8}{5} \Delta\tau_c \omega_0, \end{cases} \quad (10)$$

where $t_{0,\text{obs}}$ here (and below) denotes the fiducial time for the model. The frequency evolution regime is governed by the physically significant quantities $\Delta\tau_{c,\text{obs}}$ which is the ‘observed coalescence time,’ and θ_c which is the ‘coalescence angle’ (the total angle swept out in the source’s orbital plane before the system coalesces). A more in-depth discussion of these quantities and the physical assumptions in these regimes is given in [MC21](#).

The strength of the gravitational wave as it reaches the path that the pulsar photons take between the pulsar and the Earth also depends on the relative Earth-pulsar-source geometrical alignment. With the Earth at the centre of the coordinate system, the source frame orientation vectors and the Earth-to-pulsar vector are defined by:

$$\begin{cases} \hat{r} \equiv [\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)], & \text{(Earth – to – gravitational wave source unit vector)} \\ \hat{\theta} \equiv [\cos(\theta)\cos(\phi), \cos(\theta)\sin(\phi), \sin(\theta)], & \text{(transverse plane basis vector)} \\ \hat{\phi} \equiv [-\sin(\phi), \cos(\phi), 0], & \text{(transverse plane basis vector)} \\ \hat{p} \equiv [\sin(\theta_p)\cos(\phi_p), \sin(\theta_p)\sin(\phi_p), \cos(\theta_p)], & \text{(Earth – to – pulsar unit vector)} \end{cases} \quad (11)$$

from which we define the following generalized polarization tensors:

$$\begin{cases} e_{ij}^{f+} \equiv \hat{\theta}_i \hat{\theta}_j - \hat{\phi}_i \hat{\phi}_j, \\ e_{ij}^{f\times} \equiv \hat{\phi}_i \hat{\theta}_j + \hat{\theta}_i \hat{\phi}_j, \\ E_{ij}^{f+} \equiv \frac{1}{2} (1 + \cos^2(\iota)) [\cos(2\psi)e_{ij}^{f+} + \sin(2\psi)e_{ij}^{f\times}], \\ E_{ij}^{f\times} \equiv \cos(\iota) [-\sin(2\psi)e_{ij}^{f+} + \cos(2\psi)e_{ij}^{f\times}]. \end{cases} \quad (12)$$

This notation groups all of the geometrical orientation and location angles into the definition of the polarization tensor, in order to keep them from mixing the plus and cross metric perturbations. With these we then define the following antenna patterns which decide the detector’s sensitivity to the gravitational wave based on the Earth-pulsar-source geometrical alignment:

$$\begin{cases} f^+ \equiv \frac{\hat{p}^i \hat{p}^j e_{ij}^{f+}}{(1-\hat{r}\cdot\hat{p})} = \frac{\hat{p}^i \hat{p}^j \hat{\theta}_i \hat{\theta}_j - \hat{p}^i \hat{p}^j \hat{\phi}_i \hat{\phi}_j}{(1-\hat{r}\cdot\hat{p})} = \frac{(\hat{p}\cdot\hat{\theta})^2 - (\hat{p}\cdot\hat{\phi})^2}{(1-\hat{r}\cdot\hat{p})}, \\ f^\times \equiv \frac{\hat{p}^i \hat{p}^j e_{ij}^{f\times}}{(1-\hat{r}\cdot\hat{p})} = \frac{\hat{p}^i \hat{p}^j \hat{\phi}_i \hat{\theta}_j + \hat{p}^i \hat{p}^j \hat{\theta}_i \hat{\phi}_j}{(1-\hat{r}\cdot\hat{p})} = \frac{2(\hat{p}\cdot\hat{\theta})(\hat{p}\cdot\hat{\phi})}{(1-\hat{r}\cdot\hat{p})}, \\ F^+ \equiv \frac{\hat{p}^i \hat{p}^j E_{ij}^{f+}}{(1-\hat{r}\cdot\hat{p})} = \frac{1}{2} (1 + \cos^2(\iota)) [\cos(2\psi)f^+ + \sin(2\psi)f^\times], \\ F^\times \equiv \frac{\hat{p}^i \hat{p}^j E_{ij}^{f\times}}{(1-\hat{r}\cdot\hat{p})} = \cos(\iota) [-\sin(2\psi)f^+ + \cos(2\psi)f^\times]. \end{cases} \quad (13)$$

The final gravitational waveform in the transverse-traceless gauge along the \hat{r} -axis can now be expressed as:

$$h_{ij}^{TT\hat{r}} = E_{ij}^{f+} h_+ + E_{ij}^{f\times} h_\times = E_{ij}^{\hat{r}A} h_A \quad \text{for } A \in [+, \times]. \quad (14)$$

In the local observer frame the gravitational waves will affect the timing of local pulsars. The derivation at this point remains unchanged to what was shown in [MC21](#) (see, for example, Section A2 of that paper), and we again compute the effect that the gravitational waves have on the timed signals of these pulsars. The gravitational wave-induced fractional shift of the pulsar’s period T is:

$$\frac{\Delta T^{\text{GW}}}{T}(t_{\text{obs}}) \approx \frac{1}{2} \hat{p}^i \hat{p}^j E_{ij}^{\hat{r}A} \int_{t_{\text{obs}} - \frac{L}{c}}^{t_{\text{obs}}} \frac{\partial h_A(t_{\text{obs,ret}}, t'_{\text{obs}}, \vec{x})}{\partial t'_{\text{obs}}} \Big|_{\vec{x}=\vec{x}_0(t'_{\text{obs}})} dt'_{\text{obs}}, \quad (15)$$

where t_{obs} is the time a pulsar’s photon is observed arriving at Earth, $t_{\text{obs,ret}}$ is the retarded time of the gravitational wave, and $\vec{x}_0(t_{\text{obs}})$ is the spatial path of the photon between the pulsar and the Earth. Finally the gravitational wave-induced timing residual is the integrated fractional period shift due to the gravitational wave over the observation time. Conceptually this is the difference between the observed and expected time-of-arrival of a pulsar’s pulse ([Creighton & Anderson 2011](#); [Schneider 2015](#); [Maggiore 2018](#)):

$$\text{Res}^{\text{GW}}(t_{\text{obs}}) = \int \frac{\Delta T^{\text{GW}}}{T}(t'_{\text{obs}}) dt'_{\text{obs}} = \int \frac{T_{\text{obs}}(t'_{\text{obs}}) - T}{T} dt'_{\text{obs}} = \text{Obs}(t_{\text{obs}}) - \text{Exp}(t_{\text{obs}}). \quad (16)$$

2.3.1 The Fresnel, frequency evolution model (regime IIB)

Section 3 of [MC21](#) details the four gravitational wave-induced timing residual model regimes IA-IIB, characterized by frequency evolution and curvature effects (see Fig. 1). Generalizing to a flat cosmologically expanding universe does not change the derivation behind those four models, but it does change the interpretation of some of the model parameters.

Specifically, cosmological redshift causes the luminosity distance D_L to become the parameter that appears in the amplitude of the metric perturbation (equation 8), while similarly, we now recognize that the chirp mass and orbital frequency parameters are measured in the observer frame of reference, and are no longer equivalent to their values in the source frame (although the source frame values can be obtained if the source redshift z is recovered through parameter estimation and equation 7). Importantly, the parallax distance D_{par} (i.e. the transverse comoving distance D_M , or comoving distance, D_c , in our flat universe; see equation (A10)) now enters via the retarded time calculated in a flat expanding universe (equation 4).

Hence, the most general IIB model becomes:

$$\text{Res}^{\text{GW}}(t_{\text{obs}}) = \frac{F^A}{4} \left[\frac{h_A(\omega_{0E}, \Theta_E - \frac{\pi}{4})}{\omega_{0E}} - \frac{h_A(\bar{\omega}_{0P}, \bar{\Theta}_P - \frac{\pi}{4})}{\bar{\omega}_{0P}} \right] \quad \text{for } A \in [+, \times], \quad (17)$$

$$\text{where } \begin{cases} \omega_{0E} & \equiv \omega_{0,\text{obs}}, \\ \bar{\omega}_{0P} & \equiv \omega_{0,\text{obs}} \left[1 + \frac{(1-\hat{r}\cdot\hat{p})\frac{L}{c} + \frac{1}{2}(1-\hat{r}\cdot\hat{p})^2\frac{L}{c}\frac{L}{D_{\text{par}}}}{\Delta\tau_{c,\text{obs}}} \right]^{-3/8}, \\ \theta_{0E} & = \theta_0, \\ \bar{\theta}_{0P} & = \theta_0 + \theta_c \left(1 - \left[1 + \frac{(1-\hat{r}\cdot\hat{p})\frac{L}{c} + \frac{1}{2}(1-\hat{r}\cdot\hat{p})^2\frac{L}{c}\frac{L}{D_{\text{par}}}}{\Delta\tau_{c,\text{obs}}} \right]^{5/8} \right), \\ \Theta_E & = \theta_{0E} + \omega_{0E} t_{\text{obs}}, \\ \bar{\Theta}_P & = \bar{\theta}_{0P} + \bar{\omega}_{0P} t_{\text{obs}}, \end{cases} \quad (18)$$

along with equations (8), (10), (11), (12), and (13). Written this way the timing residual is the sum of an ‘Earth term’ (subscripted with an ‘E’) and a ‘pulsar term’ (subscripted with a ‘P’). Note that the overbar notation here is used for the same distinguishing purpose as in MC21, to remind the reader that this model has not been analytically derived, but rather proposed. MC21 explains why this model is physically motivated.

3 METHODS

3.1 Model parametrization

In this cosmological framework, the gravitational wave timing residual in the full IIB regime is governed by the source parameters: $\vec{s} = \{D_L, D_{\text{par}}, \theta, \phi, \iota, \psi, \theta_0, \mathcal{M}_{\text{obs}}, \omega_{0,\text{obs}}\}$. Following the explanation given in MC21 we swap the D_L parameter in our model with the ‘observed Earth term timing residual amplitude’ parameter, defined as:

$$A_{E,\text{obs}} \equiv \frac{h_{0,\text{obs}}}{4\omega_{0,\text{obs}}} = \frac{(G\mathcal{M}_{\text{obs}})^{5/3}}{c^4 D_L \omega_{0,\text{obs}}^{1/3}} \approx (140 \text{ ns}) \left(\frac{\mathcal{M}_{\text{obs}}}{10^9 M_\odot} \right)^{5/3} \left(\frac{100 \text{ Mpc}}{D_L} \right) \left(\frac{1 \text{ nHz}}{\omega_{0,\text{obs}}} \right)^{1/3}. \quad (19)$$

Because present day capabilities on measuring the distances to pulsars often cannot constrain them to better than the order of 100 pc (Arzoumanian et al. 2018; Deller et al. 2019), which is larger than our gravitational wave wavelengths (see equation 6), we include all of the pulsar distances as free parameters in our model (Corbin & Cornish 2010; Lee et al. 2011; MC21). Therefore we can divide the model parameters into source parameters and pulsar distance parameters $\vec{X} = [\vec{s}, \vec{L}]$.

To measure H_0 , one measures D_{par} , and also $A_{E,\text{obs}}, \mathcal{M}_{\text{obs}}, \omega_{0,\text{obs}}$, which together give D_L via equation (19). Then combining D_L and D_{par} in equation (7) gives H_0 . Note that this does require knowledge of the density parameters that go into H_0 (equation A1). In practice we take a simplified approach by using the small redshift approximation to replace the parallax distance parameter D_{par} with H_0 . This does not qualitatively change the calculation, but allows computational simplicity for this proof-of-principle study and is relevant for the source distances for which Fresnel effects are prominent (DL21, MC21). For a flat $k = 0$ universe and for $z \ll 1$, the Hubble Law is $z \approx \frac{H_0}{c} D_L$, so we can combine equations (A10) and (19) to write:

$$D_{\text{par}} \approx \left[\frac{c^4 A_{E,\text{obs}} \omega_{0,\text{obs}}^{1/3}}{(G\mathcal{M}_{\text{obs}})^{5/3}} + \frac{H_0}{c} \right]^{-1} \quad (\text{for } k = 0, z \ll 1). \quad (20)$$

Note that if one were to use the equally valid small redshift approximation $z \approx \frac{H_0}{c} D_{\text{par}}$, then our result for H_0 would differ by $\mathcal{O}(z^2)$ (\lesssim a few per cent for the fiducial cases in this study).

This approach is particularly useful when working with the two source problem described in Section 4.2, as equation (20) will reduce the dimensionality of the model by one ($D_{\text{par},1}$ and $D_{\text{par},2}$ are both replaced by H_0), therefore giving us the direct joint posterior on H_0 from the two sources. In the full calculation, one must make an estimate of the joint posterior on H_0 (for example, with a kernel density estimate) using the measured parameters from both sources. The low-redshift approach builds into our model the additional knowledge that H_0 is a constant irrespective of the source we are detecting. But again the approximation is only valid for $z \ll 1$.

As a final note, unless otherwise stated, for the studies presented in this work we chose to inject the following default values for the gravitational wave source parameters: $D_{\text{par}} = 100 \text{ Mpc}$, $D_L = 102.35 \text{ Mpc}$, $\mathcal{M}_{\text{obs}} = 10^9 M_\odot$, $\omega_{0,\text{obs}} = 30 \text{ nHz}$, and angular parameters $\{\theta, \phi, \iota, \psi, \theta_0\} = \{\frac{3\pi}{7}, \frac{5\pi}{3}, \frac{2\pi}{5}, \frac{\pi}{7}, 1\}$.

3.2 Likelihood and priors

Following MC21, for this work we assume that the timing residual data is the sum of the underlying gravitational wave-induced residual plus some random noise, that is $\vec{\text{Res}} = \vec{\text{Res}}^{\text{GW}} + \vec{N}$. For simplicity the noise we model is white noise, and each data point collected has some uncertainty uncorrelated between observations/pulsars, that is the timing covariance matrix $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2)$. Therefore the likelihood function we propose and use in this work is:

$$\begin{aligned} \mathcal{L}(\vec{\text{Res}} | \vec{X}) &= \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp \left[-\frac{1}{2} \left(\vec{\text{Res}} - \vec{\text{Res}}^{\text{GW}}(\vec{X}) \right)^T \Sigma^{-1} \left(\vec{\text{Res}} - \vec{\text{Res}}^{\text{GW}}(\vec{X}) \right) \right], \\ &= \frac{1}{\sqrt{(2\pi)^d \sigma^d}} \prod_{a=1}^d \exp \left[-\frac{1}{2} \frac{1}{\sigma_a^2} \left(\text{Res}_a - \text{Res}_a^{\text{GW}}(\vec{X}) \right)^2 \right]. \end{aligned} \quad (21)$$

Here $\vec{\text{Res}}$ is the d -dimensional timing residual data (with dimension equal to the number of pulsars times the number of observations per pulsar, indexed by a), $\vec{\text{Res}}^{\text{GW}}$ is our residual model (for the IIB model, see Section 2.3.1) for every data point, and the model parameters are contained in the vector \vec{X} . Log-parameters are used for the non-angular parameters $\{\log_{10}(A_{E,\text{obs}}/A_{E,\text{obs,true}}), \log_{10}(D_{\text{par}}/D_{\text{par,true}}), \log_{10}(\mathcal{M}_{\text{obs}}/\mathcal{M}_{\text{obs,true}}), \log_{10}(\omega_{0,\text{obs}}/\omega_{0,\text{obs,true}})\}$, as well as $\log_{10}(H_0/H_{0,\text{true}})$. For our priors, we required all of the physical parameters be non-negative ($A_{E,\text{obs}}, \mathcal{M}_{\text{obs}}, \omega_{0,\text{obs}}, D_L, D_{\text{par}}, L, H_0$), and that $D_{\text{par}} \leq D_L$ (which assumes $z > 0$ in equation A10). For the angular parameters we placed the general boundaries: $0 < \theta, \iota, \theta_0 < \pi$, and $0 < \phi, \psi < 2\pi$.

3.3 Mock observations

In order to create the mock timing residual data for this study, we generate different mock PTAs, namely ‘fiducial PTAs’ and ‘Square Kilometre Array (SKA)-era PTAs.’ The details for constructing these arrays follow below, but each serves a distinct purpose in this study. While the fiducial PTA is physically motivated, it is a simpler construction, and is primarily used throughout this work to look for various trends and scaling laws in our predicted measurement capabilities. On the other hand, the various SKA PTAs are meant to represent more realistic hypothetical future timing arrays, and are therefore used to simulate more realistic results.

Note that for simplicity, in this work we do not add noise \vec{N} on top of our gravitational wave-induced timing residuals. Therefore the simulated timing residuals are purely the gravitational wave timing residual component, which is why in the MCMC results shown (Figs 2, 4, and D3) the posteriors are centred on the true injected parameters. Finally, for all of our simulations in this paper, we use an observation time of 10 yr, and a timing cadence of 30 observations per year.

3.3.1 A fiducial PTA

To gauge dependence of parameter recovery on pulsar number and timing precision we choose a simple fiducial PTA. The fiducial PTA is realistically motivated by using a simple Milky Way Galaxy structure model to create a density distribution of pulsars n_{psr} (Schneider 2015) to sample from,

$$n_{\text{psr}} \propto \exp \left(-\frac{r}{h_R} \right) \left[\exp \left(-\frac{|z|}{h_{\text{thin}}} \right) + 0.02 \exp \left(-\frac{|z|}{h_{\text{thick}}} \right) \right] \times \exp \left(-\frac{1}{2} \left[\left(\frac{x}{R} \right)^2 + \left(\frac{y}{R} \right)^2 + \left(\frac{z}{R} \right)^2 \right] \right) H(r - r_{\text{min}}), \quad (22)$$

where the centre of the coordinate system is at the Earth’s position, the Galactic centre is at $x_{\text{MW}} \approx 8$ kpc, and $r \equiv \sqrt{(x - x_{\text{MW}})^2 + y^2}$. Here h_R is the scale length of the Galactic disc ($h_R \approx 3.5$ kpc), h_{thin} is the scale height of the thin disc ($h_{\text{thin}} \approx 0.325$ kpc), and h_{thick} is the scale height of the thick disc ($h_{\text{thick}} \sim 1.5$ kpc). The first term in this expression creates the distribution of stars in the Milky Way, while the second term simply places preference on stars within a Gaussian ball centred on Earth. The motivation for including this second term is simply the idea that we will likely be more sensitive to timing pulsars within some scale distance R from the Earth. We also placed a minimum distance r_{min} on this sphere (hence the Heaviside function). For our fiducial PTA we chose $R = 5$ kpc and $r_{\text{min}} = 0.1$ kpc. An example of 1000 pulsars generated from this distribution is shown in Fig. D1.

3.3.2 Simulated SKA-era PTAs

In addition to the fiducial PTA, we use the `PSRPOP` package (Bates et al. 2014) to generate a population of millisecond pulsars (MSPs) that could be discovered and used for timing in the SKA era. We begin by generating a population of 3×10^4 MSPs in the galaxy (motivated by Smits et al. 2009) and simulate an SKA-like survey which detects 8893 of these with signal-to-noise (SNR) above 9. We separate the detected MSPs into 20 radial bins out to the furthest detected pulsars at ~ 18 kpc (with the closest at ~ 0.2 kpc). To simulate which MSPs will be best for high-precision timing, we sort the MSPs by SNR in each radial bin and retain the top p per cent for the final PTA. In the outer three bins which contain < 10 MSPs we take only the highest SNR pulsar, each of which have SNRs ≥ 20 . Depending on the value of p , the final PTAs contain between ~ 40 and 1800 pulsars (See Appendix D for further details).

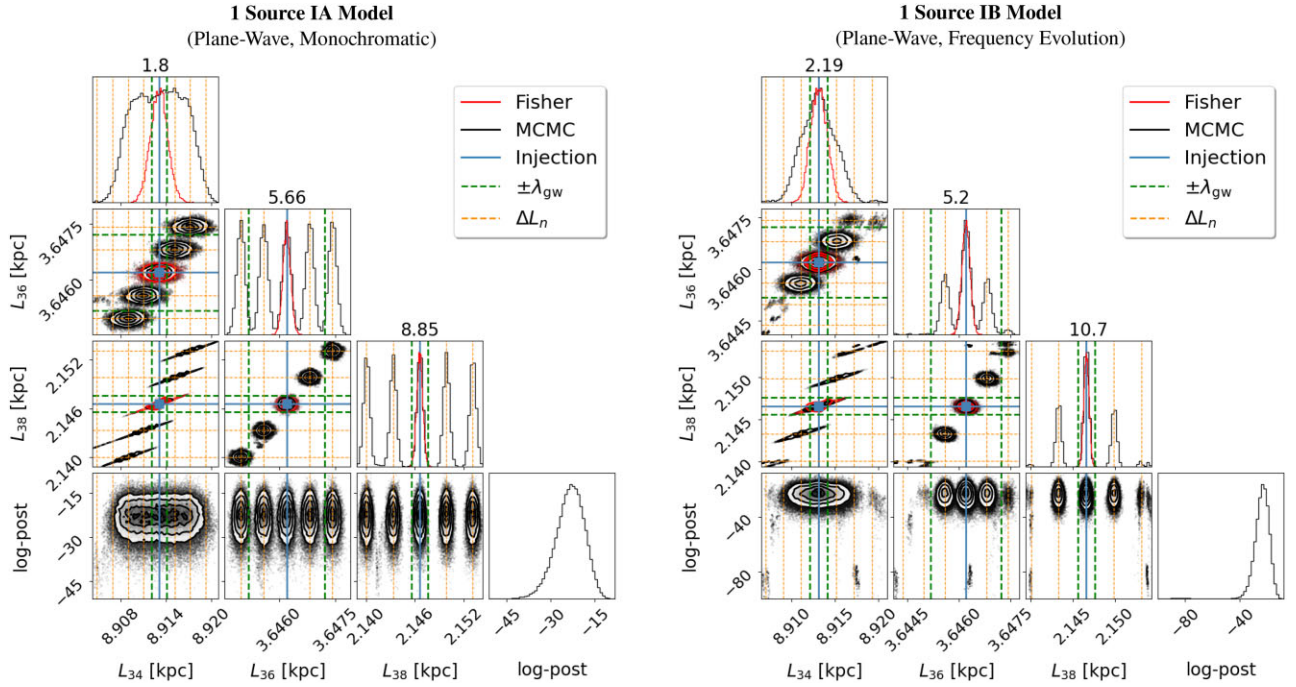


Figure 2. Corner plots demonstrating the pulsar distance error envelope from an MCMC simulation of a PTA with a single gravitational wave source (default parameters). The results are compared using the (non-Fresnel regime) IA versus IB models (MC21) for the exact same PTA and source. The full PTA contained 40 pulsars from the fiducial PTA (see Fig. D3), with a uniform timing uncertainty of $\sigma = 0.2$ ns. This corner plot singles out three pulsars and displays the modal overlap criterion ratios from equation (28) above each 1D histogram. The bottom row shows the (un-normalized) log-posterior values. **Left-hand panel:** For the monochromatic IA model, each mode receives the same support in the likelihood function at every wrapping cycle. **Right-hand panel:** For the IB model, we see how frequency evolution helps us to weaken the wrapping cycle degeneracy by reducing support in the likelihood at secondary wrapping cycles modes. Note the samples in this panel have been cleaned, such that all samples with log-posterior ≤ -100 have been removed from the data set (see the discussion of cleaning in Section 4.3).

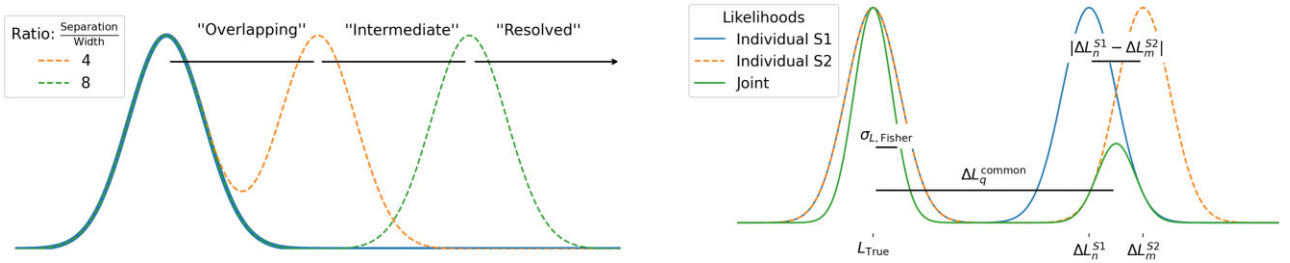


Figure 3. Here we visualize the modal overlap criterion used to analyse the pulsar wrapping cycle degeneracy (equations 28–30). **Left-hand panel:** The ‘modal overlap’ criterion between the primary and secondary modes in the pulsar distance wrapping problem is defined as the ratio of the separation of the primary and secondary modes (ΔL_1 for a single source, and $\Delta L_1^{\text{common}} = |\Delta L_n^{S1} + \Delta L_m^{S2}|/2$ for two sources) to the width of the modes ($\sigma_{L, \text{Fisher}}$). We classify values of this ratio < 4 as ‘overlapping’, between 4 and 8 as ‘intermediate’, and > 8 as ‘resolved.’ **Right-hand panel:** Individual likelihoods for two sources S1 and S2 will agree on the location of the primary mode in the pulsar distance, but as long as the sources are located on different parts of the sky and/or have different frequencies, they will have different sets of pulsar distance wrapping cycles (equation 27). Therefore the joint likelihood for both sources will find support at the distances where there is some common modal overlap between ΔL_n^{S1} and ΔL_m^{S2} .

Because our SKA simulation allows us to estimate detection SNRs and spin periods for the simulated SKA-era population of MSPs, we can additionally estimate a timing uncertainty for the a th MSP in the array. To do so, we assume that the pulse time-of-arrival uncertainty is proportional to the radiometer noise (e.g. Verbiest & Shaifullah 2018),

$$\sigma_a \propto \frac{P_a}{\sqrt{\Delta T_a \text{SNR}_a}}, \quad (23)$$

for integration time ΔT , pulse period P , and detection signal-to-noise SNR. While this represents only part of the noise budget, it allows a study where pulsars have a more realistic spread of timing uncertainties than in the fiducial PTAs. We then consider two scenarios, one where the integration time is constant for all pulsars ($\Delta T_a = \text{cst.}$), and one where the observation strategy is intelligently chosen to boost the signal

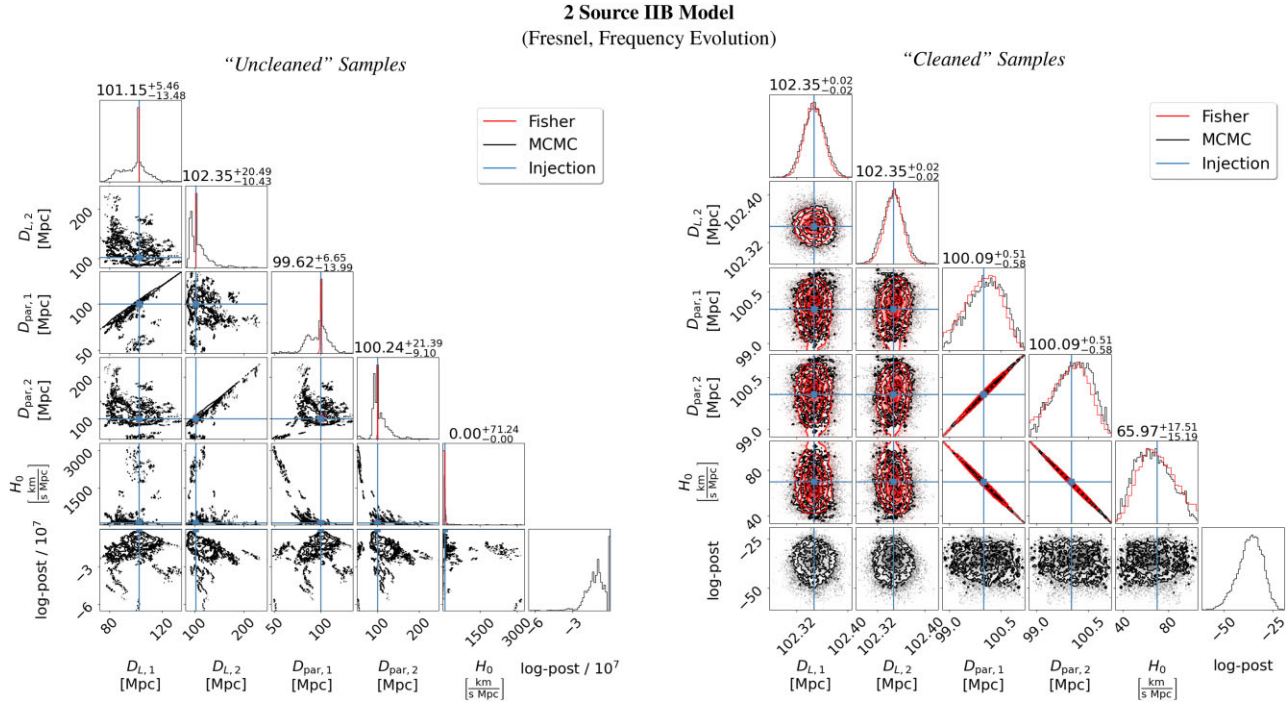


Figure 4. A targeted MCMC simulation of the recovery of H_0 for two sources (using the IIB model). For simplicity, both sources are identical except for being located on different parts of the sky. Source 1 is at the default location (and matches the single source in Fig. 2), while source 2 has sky angles $\theta_2 = \frac{\pi}{3}$ rad and $\phi_2 = \frac{\pi}{5}$ rad. Here 40 pulsars from our fiducial PTA are timed with $\sigma = 0.2$ ns uncertainty (same as in Fig. 2). Both panels are of the same simulation, showing just the recovery of the two sources’ distances (D_L and D_{par}), and the joint recovery of H_0 . The bottom row shows the (un-normalized) log-posterior values. For each pulsar, if we look at the first common mode where the common mode overlap strength criterion equation (29) is ≤ 2 , then for this given source and PTA our common modal overlap criterion predicts that there will be 0 overlapping, 0 intermediate, and 40 resolved pulsar error envelopes. For this reason, our prediction before even running the MCMC is that we should not find any strong support at secondary modes, making this is an ideal candidate for the H_0 measurement, which for this set of source and PTA parameters has a predicted $\text{CV}_{H_0} = 0.257$. Indeed, this is what we see here – unlike in Fig. 2, our walkers did not seem to find any high probability secondary modes. **Left-hand panel:** For the full simulation we initialized multiple ‘balls’ of walkers as described in Section 4.3. The bottom row shows many orders of magnitude difference in the values of the log-posterior for the walkers centred on the true mode versus the walkers away from the true mode. If we ran the MCMC simulation long enough, we would expect that these low-probability walkers would eventually find and settle on the true mode. **Right-hand panel:** This is the same corner plot, but now with all values of the log-posterior $\leq -2 \times 10^6$ removed from the data set. This ‘cleaned’ data reveals the sharply peaked true values in parameter space, which closely agree with the Fisher predictions.

from lower SNR MSPs ($\Delta T_a \propto \text{SNR}_a^{-1}$),

$$\sigma_a = \sigma_{\min} \left(\frac{P_a}{1 \text{ ms}} \right) \begin{cases} \left(\frac{\text{SNR}_a}{\text{SNR}_{\max}} \right)^{-1}, & \Delta T_a = \text{cst.} \\ \left(\frac{\text{SNR}_a}{\text{SNR}_{\max}} \right)^{-1/2}, & \Delta T_a \propto \text{SNR}_a^{-1}, \end{cases} \quad (24)$$

where we choose a characteristic, best-case timing uncertainty $\sigma_{\min} = \{1, 10\}$ ns and enforce $\sigma_a \geq \sigma_{\min}$. While somewhat arbitrary, we assume an $\text{SNR}_{\max} = 10^4$, corresponding to an SNR threshold above which pulsars with 1 ms pulse periods can be timed to σ_{\min} . Mock array pulsar distributions and timing uncertainties are shown in Appendix D.

3.4 Fisher matrix and MCMC

To recover best-fitting source and pulsar parameter values we carry out both Fisher matrix and Markov Chain Monte Carlo (MCMC) analyses which compare our model and mock timing residuals through the likelihood (equation 21).

A Fisher matrix is a useful tool for quick parameter estimation. Calculating the inverse of the Fisher matrix gives the estimated parameter covariance matrix $\mathbf{F} = \mathbf{C}^{-1}$ for the experiment. Therefore finding a model’s inverse Fisher matrix can help to tell us which parameters are covariant with each other, and roughly how well we might expect to recover each model parameter given our experimental set-up. Computationally, this is efficient and can quickly allow us to test many different experiments. Fisher-based surveys are useful for quickly searching large parts of parameter space for trying to understand what types of sources would result in good parameter estimation. But it does require that one assume a more robust search (such as with an MCMC analysis) could successfully identify the true mode from any potential secondary modes. This limitation is due to the fact that the Fisher matrix is only meant to approximate the shape of the posterior near the maximum likelihood. If for example the posterior is multimodal (see Section 4), then the Fisher matrix will not capture this behaviour. For a more extensive (and computationally expensive) targeted study of the posterior, we use MCMC.

From our likelihood equation (21) and the definition of the Fisher matrix as $F_{ij} \equiv \left\langle \left(\frac{\partial}{\partial X_i} \ln \mathcal{L}(\vec{\text{Res}} | \vec{X}_{\text{true}}) \right) \left(\frac{\partial}{\partial X_j} \ln \mathcal{L}(\vec{\text{Res}} | \vec{X}_{\text{true}}) \right) \right\rangle$ we have:

$$F_{ij} = \sum_a \frac{1}{\sigma_a^2} \left(\frac{\partial \text{Res}_a^{\text{GW}}(\vec{X}_{\text{true}})}{\partial X_i} \right) \left(\frac{\partial \text{Res}_a^{\text{GW}}(\vec{X}_{\text{true}})}{\partial X_j} \right). \quad (25)$$

In some instances, namely when using the Fresnel models, inverting the Fisher matrix cannot be done accurately because inclusion of the pulsar distance parameters introduces too much uncertainty and covariance amongst the other parameters (this typically happens with a single gravitational wave source). In these instances we add uncorrelated Gaussian pulsar distance priors to the Fisher matrix (Wittman, unpublished notes¹; Iacovelli et al. 2022). See Section 5 of MC21 for more details on adding these priors, and for a discussion of why this problem is more prominent in the Fresnel regimes.

Since all parameters in a Fisher matrix approximation are normally distributed, we often use the coefficient of variation (CV), the predicted fractional error on a given parameter, as a useful way of quantifying the measurability of a given parameter when performing a Fisher matrix analysis:

$$\text{CV}_x \equiv \frac{\text{Standard Deviation of } x}{\text{Expectation Value of } x} = \begin{cases} \frac{\sigma}{\mu}, & \text{Normal Distribution in } x \\ \sqrt{e^{\sigma^2 \ln(10)^2} - 1}. & \text{Log}_{10}\text{-Normal Distribution in } x \end{cases} \quad (26)$$

Here μ and σ are the distributions' parameters. If the parameter x is normally distributed then μ and σ are also the distribution's mean and standard deviation, and if the parameter is log-normally distributed then CV only depends on the lognormal σ parameter. A resulting CV of the order of unity or larger suggests the parameter would not be measureable given the source and experiment, while smaller CVs suggest better parameter recovery.

We use the PYTHON EMCEE package (Foreman-Mackey et al. 2013) to carry out MCMC sampling of the posterior. We use differential evolution jump proposals with the number of walkers set between 5 and 7 times the number of parameter dimensions, and the results shown in this paper ran between 350 000 and 500 000 iterations. We target our searches by initializing the MCMC walkers in very small Gaussian 'balls' about the true injected parameters, as well as around nearby theoretical secondary modes of interest. For the covariance matrix of these Gaussian balls we use the inverse of the Fisher matrix prediction, typically multiplied by some overall scale factor. This choice helps the MCMC simulations more quickly find and explore the local distributions around the true parameters.

While our MCMC simulations help provide proof-of-principle for the main ideas in this paper, future work should explore using different MCMC samplers to perform consistency checks on the work presented here, such as with parallel tempering or nested samplers (Samajdar et al. 2022). Further investigation of custom jump proposals should also be implemented to try and improve the exploration of walkers from one posterior mode to another.

4 A TWO-SOURCE SOLUTION TO THE PULSAR DISTANCE WRAPPING PROBLEM

The pulsar distance wrapping problem creates a challenge for parameter estimation in this work. As explained in detail in MC21, due to the way the pulsar distance L affects the phase of the pulsar term (equation 18), increasing or decreasing the pulsar distance by specific ΔL values will cause the phase to wrap around the interval $[0, 2\pi)$, resulting in the same timing residual for multiple pulsar distances.

However, we show here that the wrapping problem is prominent primarily when there is only one gravitational wave source in the signal. When there are two or more sources simultaneously creating residuals in every pulsar, the degeneracy of the wrapping cycle is broken. This has a significant effect on our ability to recover the parallax distance D_{par} and subsequently the Hubble constant, largely because breaking the wrapping degeneracy can strongly constrain the pulsar distances.

4.1 Classifying the error envelope (single source)

More formally, in the monochromatic regimes IA and IIA, the wrapping cycle degeneracy arises because the timing residual is identical for $L \pm \Delta L_n$:

$$\Delta L_n \equiv \begin{cases} n \frac{\lambda_{\text{gw}}}{(1-\hat{r} \cdot \hat{p})}, & (\text{Plane-Wave, monochromatic}) \\ -\left(\frac{D_{\text{par}}}{(1+\hat{r} \cdot \hat{p})} + L \right) + \sqrt{\left(\frac{D_{\text{par}}}{(1+\hat{r} \cdot \hat{p})} + L \right)^2 + n \frac{2\lambda_{\text{gw}} D_{\text{par}}}{(1-\hat{r} \cdot \hat{p})^2}}, & (\text{Heuristic Fresnel, monochromatic}) \end{cases} \quad (27)$$

for $n \in \mathbb{Z}$.² Hence, the wrapping cycle distance ΔL is a scaled version of the gravitational wave wavelength λ_{gw} , which includes geometric factors dependent on the relative positions of the pulsar and the source. Most notably, the more aligned a source and pulsar are on the sky (within 90° , $0 < \hat{r} \cdot \hat{p} < 1$), the larger the wrapping cycle becomes, and vice versa as the source becomes more anti-aligned (more than 90° , $-1 < \hat{r} \cdot \hat{p} < 0$).

¹Wittman D., no date, Fisher Matrix for Beginners, UC Davis, <http://wittman.physics.ucdavis.edu/Fisher-matrix-guide.pdf>.

²See MC21 and note that in equation (27) we replace $R \rightarrow D_{\text{par}}$ to fit with our generalization described in Section 2.1.

Consider the problem now in terms of parameter estimation using the likelihood function equation (21). If we work entirely in one of the monochromatic regimes IA or IIA, meaning we use one of these models to both calculate the timing residuals of our injected source $\vec{\text{Res}}$ and recover the parameters $\vec{\text{Res}}(\vec{X})$, then the likelihood function is perfectly multimodal at every $L \pm \Delta L_n$ for every pulsar in the PTA (see the left-hand panel of Fig. 2). Therefore in either of these regimes, we cannot identify the true pulsar distance with this likelihood since all modes have equal probability.

In the frequency evolution regimes IB and IIB this wrapping cycle does not formally exist as the frequencies are now time-dependent. While this technically breaks the degeneracy, there still exists some support at $L \pm \Delta L_n$. That is, the IB and IIB likelihood functions are still multimodal at each pulsar’s true distance modulo its wrapping cycle distance (see the right-hand panel of Fig. 2). However, as moving out a wrapping cycle no longer returns the same exact timing residual as the timing residual at the true pulsar distance, these modes become less and less probable for every additional cycle away from the true distance.

Additionally, every mode will have some width. If these modes are close enough to each other (relative to the size of their widths), then the modes will blend together. Corbin & Cornish identified this blending of the pulsar distance wrapping modes in their model as an ‘error envelope’ which forms about the true pulsar distance. The effect is that the true distance will be further buried within the uncertainty surrounding the secondary modes, making it much harder to measure in practice (see for example the top-most 1D histogram in both panels of Fig. 2). However, Corbin & Cornish did not fully classify the properties of the error envelope. Namely the geometric factors in the wrapping cycle (equation 27) are of crucial importance in determining the separation between the true mode and the secondary modes.

Therefore we construct a useful way of classifying the amount of ‘modal overlap’ between the primary (thick blue line) and secondary modes (dashed orange lines). Overlap between the first and second mode will depend on two factors: the width of each mode, and the separation between modes. If each mode is a Gaussian of approximately the same width, then the ratio of mode separation to mode width provides a numerical quantification of the amount of overlap (see the left-hand panel of Fig. 3). We define such a modal overlap criterion by calculating the ratio of the first wrapping cycle distance ΔL_1 (i.e. the distance between our primary and secondary modes) to the approximated Fisher width $\sigma_{L, \text{Fisher}}$ of each mode:

$$\frac{\Delta L_1}{\sigma_{L, \text{Fisher}}} \sim \begin{cases} < 4, & \text{Overlapping} \\ 4 - 8, & \text{Intermediate} \\ > 8, & \text{Resolved} \end{cases} \quad (28)$$

Here $\sigma_{L, \text{Fisher}}$ is the uncertainty of a given pulsar distance as predicted by the Fisher matrix analysis (that is, the standard deviation of a pulsar distance L calculated from the inverse Fisher matrix, $\mathbf{C} = \mathbf{F}^{-1}$). Conceptually, the values we choose here are motivated such that a secondary mode ‘overlapping’ with the primary mode will have its peak within $4\sigma_{L, \text{Fisher}}$ of the true value, and ‘resolved’ when its peak is greater than $8\sigma_{L, \text{Fisher}}$ from the true value. Fig. 2 shows an example of three pulsars from a simulation with overlapping, intermediate, and resolved modes clearly visible. A complete version is given in Fig. D3.

Note that in general the Fisher uncertainty predictions are only valid around the true mode, and therefore cannot accurately forecast the multimodal behaviour of the pulsar distances. However, in practice we observe that the secondary and primary modes have approximately the same width. Therefore, we use $\sigma_{L, \text{Fisher}}$ as our proxy in defining equation (28). This effectively assumes that the behaviour of the likelihood function near each wrapping cycle will look the same as the behaviour of the likelihood function near the true mode.

In practice, this specific modal overlap criterion also works best if no prior knowledge is placed on the pulsar distances. This is partly because $\sigma_{L, \text{Fisher}}$ would be affected by a prior (see discussion in Section 3.4), while ΔL_1 would not. In our MCMC simulations, we find that when a very constricting prior is placed on the pulsar distances (such as a prior of the order of the wrapping cycle distance $\sim \mathcal{O}(\Delta L_1)$), then the location of the secondary modes in the posterior can shift away from ΔL_1 and closer to the true mode. This results in a change in the numerical ranges for which we find the overlapping, intermediate, and resolved modes in equation (28).

One other important caveat to note is that this criterion also works best if there are no secondary modes in the other source parameters. Consider equation (27) for the wrapping cycle modes. In the case of the plane-wave, monochromatic wrapping cycle distance, $\Delta L_n = \Delta L_n(\theta, \phi, \omega_{0, \text{obs}})$. If the likelihood has secondary modes in the frequency parameter $\omega_{0, \text{obs}}$, then those secondary mode solutions can produce their own wrapping cycle modes, which we observed in several MCMC tests. Specifically, for some high mass systems $\mathcal{M}_{\text{obs}} > 10^9 M_{\odot}$, secondary modes (of lower probability) form in the \mathcal{M}_{obs} and $\omega_{0, \text{obs}}$ parameters, which then change the locations of the observed secondary modes in pulsar distances (meaning our criterion becomes less accurate in this case). In such cases, these new modes may be the result of stronger frequency evolution effects within the model, and should be kept in mind when performing these types of analyses.

Hence, equation (28) provides a new and more general criterion for quantifying the wrapping problem. While this criterion relies on the Fisher predicted uncertainty $\sigma_{L, \text{Fisher}}$, which does not offer an immediately intuitive connection to the system parameter values, equation (28) can still be calculated for all pulsars in a PTA without requiring a full MCMC simulation. Therefore we can quickly predict for any pulsar in our PTA whether or not that pulsar will have an error envelope. The more pulsars in the PTA which are predicted to have resolved modes, the more likely it will be that we can obtain a measurement of the desired parallax distance D_{par} and hence the Hubble constant.

4.2 Breaking the wrapping cycle degeneracy with two sources

Now consider two gravitational wave sources leaving their combined signal in the timing residuals of every pulsar. With a single source, every pulsar has a specific wrapping cycle distance (equation 27) which primarily depends on the frequency of the source and the angular

sky separation between the source and the pulsar. But with two sources, every pulsar will have two wrapping cycle distances. The key idea is that these two sets of wrapping cycle distances will not be the same as long as the two sources differ in either frequency and/or angular sky position. Therefore, the joint likelihood for both sources should only contain secondary modes in an individual pulsar's distance at common multiples of both the wrapping cycle distances. In principle adding even more sources would further break the degeneracy, because secondary modes should then only form in the joint likelihood at common multiples of all sources. We restrict our attention to two SMBHB sources in this work, and leave it open to future work to consider more sources.

Consider the right-hand panel of Fig. 3. The joint likelihood should find some support at common distances where the two separate wrapping cycle distance uncertainty modes overlap. The amount of support at those common distances will depend on how strong the modal overlap is. And if there is strong modal overlap at a common distance, we can then check if that common mode will overlap with the primary mode to create an error envelope (as we did in equation 28). Therefore we can define two new criteria. The first criterion predicts the 'common mode overlap strength,' and the second criterion classifies the 'common modal overlap' with the primary mode.

For the common mode overlap strength criterion we choose to write:

$$\frac{|\Delta L_n^{S1} - \Delta L_m^{S2}|}{\sigma_{L, \text{Fisher}}} \sim \begin{cases} 0 - 2, & \text{'strong' support} \\ 2 - 4, & \text{'weak' support} \end{cases}, \quad (29)$$

where ΔL_n^{S1} and ΔL_m^{S2} are the n th and m th wrapping cycle distances of source 1 and source 2, respectively. As an example, if this quantity equaled zero then the n th and m th modes would perfectly overlap, hence there would be strong support for a mode here. At a value of two, the point in between the n th and m th modes (the average distance) would be $1\sigma_{L, \text{Fisher}}$ from each peak, and at a value of four, the point in between the n th and m th modes would be $2\sigma_{L, \text{Fisher}}$ from each peak. We then expect that the common wrapping cycle distance is at approximately the average distance between these individual modes, that is $\Delta L_q^{\text{common}} \equiv \frac{\Delta L_n^{S1} + \Delta L_m^{S2}}{2}$, for the q th common wrapping cycle (between the n th and m th wrapping cycles of sources 1 and 2). Note that when calculating positive or negative common wrapping cycles, both ΔL_n^{S1} and ΔL_m^{S2} should be the same sign.

Note that for the purpose of our studies, we will only be searching for common modes where we think there is strong support, as we define in the criterion in equation (29). The reason for this is that 'weak' common modes will not have nearly as much support in the likelihood function. So even if the modal overlap between a weak common mode and the primary mode is such that it would produce an error envelope (using equation 30), we expect that this envelope would not contribute significantly to widening the uncertainty about the true pulsar distance.

For the common modal overlap criterion we write the following expression in analogy to equation (28) (see also the left-hand panel of Fig. 3):

$$\frac{\Delta L_1^{\text{common}}}{\sigma_{L, \text{Fisher}}} \sim \begin{cases} < 4, & \text{Overlapping} \\ 4 - 8, & \text{Intermediate} \\ > 8, & \text{Resolved} \end{cases} \quad (30)$$

This condition checks if the uncertainty around the 1st common wrapping cycle will be close enough to the uncertainty around the true distance as to blend those uncertainties into a greater envelope. Once again, in order to make this prediction we are assuming as a proxy that the width of the true mode predicted by the Fisher matrix for the joint source likelihood also approximates the secondary modes at the individual source wrapping cycles.

The key point is that with more than one source, it is more likely that common modes in the pulsar distance will form further away from the true distance mode. This means that the uncertainties surrounding these common modes will be less likely to blend with the uncertainty surrounding the true distance to form an error envelope, hence resulting in better recovery of the pulsar distances. Furthermore, recall that in the regimes IB and IIB, frequency evolution alone nominally breaks the wrapping cycle degeneracy (Section 4.1). Therefore two sources with significant frequency evolution will further constrain the pulsar distances. All of this coupled with prior knowledge on our pulsar distances thanks to electromagnetic observations provide three separate means of localizing the pulsars in our PTA. In application to our Hubble constant measurement, better pulsar distance measurements mean we can better recover the source parallax distance D_{par} from the Fresnel effects, which is what we show in Section 5.

4.3 Targeted MCMC results for two sources

Next we run a targeted MCMC simulation to test our discussion from Section 4 of how two sources break the wrapping cycle degeneracy. For two gravitational wave sources, our likelihood function (equation 21) now has $\vec{\text{Res}}^{\text{GW}} = \vec{\text{Res}}_1^{\text{GW}} + \vec{\text{Res}}_2^{\text{GW}}$ for sources '1' and '2.' Decomposing this joint likelihood function, we can write it as the product of the likelihood function of just source 1, the likelihood of just source 2, and their cross terms. Fortunately adding a second source does not double our parameter space, since both sources will share the pulsar distance parameters \vec{L} . We simply double the number of source parameters, so $\vec{X} = \vec{X}_1 \cup \vec{X}_2 = [\vec{s}_1, \vec{s}_2, \vec{L}]$. Furthermore, using the small redshift approximation (equation 20) to replace the parallax distance $D_{\text{par}, 1}$ and $D_{\text{par}, 2}$ parameters with H_0 reduces the model parameter dimensionality by one and gives us the joint posterior recovery of H_0 .

In order to test the ideas in Section 4.2, our methodological approach was to inject balls of walkers in specific areas of parameter space in order to strategically initialize them (as mentioned in Section 3.4). All walkers were initialized near the true injected source parameters, but the pulsar distance parameters were initialized distinctly. One ball of walkers was placed around the true mode pulsar distance parameter

values (L_{true}); two balls of walkers were placed at the source one wrapping cycles $L_{\text{true}} + \Delta L_{-1,+1}^{S1}$; two balls of walkers were placed at the source two wrapping cycles $L_{\text{true}} + \Delta L_{-1,+1}^{S2}$; and finally, two balls of walkers were placed at the first predicted *strong* (such that equation 29 was ≤ 2) common wrapping cycles $L_{\text{true}} + \Delta L_{-1,+1}^{\text{common}}$.

This was done because we wanted to see if indeed the walkers placed near the original wrapping cycles of the individual sources would no longer find any meaningful support in posterior, as they did in Fig. 2. The walkers near the new predicted common wrapping cycles were placed there in order to see if any support in the posterior would be found at those locations. However, depending on how far away those common modes are from the true mode we did not necessarily expect to find meaningful support there either. This is because we know that when frequency evolution is included in the model, secondary modes further away from the true mode become less probable (see again, the right-hand panel of Fig. 2).

The result of a targeted MCMC search is shown in Fig. 4, which is identical to the set-up in Fig. 2, but now with the inclusion of a second SMBHB source (and using the full 2 Source IIB model). For this example, our common modal overlap criterion predicts that all 40 pulsars would be resolved (again given that for each pulsar we look at the first common mode where the common mode overlap strength criterion equation (29) is ≤ 2). Compare this to the single source IB model results in Fig. D3, where only 10 pulsars were resolved. Therefore the addition of a second source changed this prediction such that the remaining 30 pulsars should be resolvable. Additionally, the predicted $\text{CV}_{H_0} = 0.257$. Therefore our expectation was that we would find no meaningful support in the posterior at the secondary modes as compared to the primary mode, resulting in a measurement of the Hubble constant. The left-hand panel of Fig. 4 shows that the posterior has sharp peaks at the true parameter values with broad bases surrounding them.

But if we consider the bottom row of the left panel, which shows the (un-normalized) log-posterior values of these MCMC results, we also see that the samples span a large range of parameter space with very low probability. Therefore if we try ‘cleaning’ this data by simply removing all samples with log-posterior values below a certain threshold (in this case, we cleaned all values for which the log-posterior $\leq -2 \times 10^6$), then we obtain the results shown in the right panel. This reveals that walkers near the true mode very closely traced the Fisher matrix prediction, walkers away from the true mode had very low-probability support, and that there were no clearly defined modal peaks – which starkly contrasts the results found in Fig. 2. Therefore these results support our conclusion about how two sources can help break the wrapping cycle degeneracy.

It is important to point out that we have shown proof-of-principle in the ability to disentangle secondary posterior modes from the primary mode. The caveat, as we see here is that the true mode is sharply peaked, and therefore may be difficult to find in a blind search over parameter space. Other techniques may be needed to find these true modes, such as parallel tempering MCMC, nested sampling, or different jump proposals to improve sampling efficiency. We leave this open to further investigation in future studies.

5 MEASUREMENT OF THE HUBBLE CONSTANT

5.1 Comparing H_0 measurements with one versus two gravitational wave sources

MC21 and DL21 found that with a single gravitational wave source we must know a priori the pulsar distances to sub-gravitational wavelength uncertainties in order to avoid the pulsar distance wrapping problem. Only then can D_{par} be recovered with sufficient accuracy to measure the Hubble constant through equation (7) (exact) or equation (20) (approximate). An example showing the recovery of D_L compared to D_{par} for a single source and assuming pulsar distance uncertainties of $\sigma_L = 0.1\lambda_{\text{gw}}$ is shown in Fig. 5. However, this level of precision in the pulsar distance measurement is far beyond current capabilities for all but the most nearby pulsars, making this a very challenging measurement restricted to gravitational wave sources within ≤ 100 Mpc (DL21 and MC21).

The top panel of Fig. 6 shows a direct comparison of measuring H_0 using one versus two gravitational wave sources. With one gravitational wave source, when performing a Fisher matrix analysis the only way to accurately invert the Fisher matrix requires that we first add in pulsar distance priors which would constrain the uncertainty on the pulsar distances down to the order of the wrapping cycle (see Section 3.4). Otherwise, the Fisher matrix is ill-conditioned, due to the strong covariances introduced from the pulsar distances and parallax distance parameters. So for a single source, the blue line in the top panel of Fig. 6 shows the recovery of H_0 as a function of the pulsar distance prior, which is a constraint applied to *all* of the pulsars in the PTA uniformly. We see that H_0 can only be recovered in this example for distance priors smaller than a few per cent of the gravitational wave wavelength.

With two gravitational wave sources, we can achieve the same level of accuracy in H_0 recovery without pulsar distance prior knowledge. The orange lines in the top panel of Fig. 6 demonstrate this for both a ‘good’ and ‘poor’ sky location. Simply having two sources at favorable sky separations and frequencies results in recovery of the Hubble constant better than a single source with sub-gravitational wavelength prior knowledge. Even unfavorable source sky positions still result in better recovery than with a single source.

The bottom panel of Fig. 6 demonstrates H_0 recovery for all sky realizations of the relative position of the second source, assuming the first source is located at the position of the red star. In the limit that the two sources become perfectly aligned with each other on the sky, the recovery of H_0 becomes effectively impossible. This is a rather interesting result, because from our previous discussions in Sections 4.1 and 4.2 we originally predicted that as long as the source frequency of both sources were different (even for the same sky position), the wrapping cycle degeneracy should break. Mathematically this still happens, but when testing various scenarios we found that in the case of perfect source sky alignment the model parameters for our PTA and sources needed to be highly favourable, making it very unlikely that such a circumstance would occur naturally.

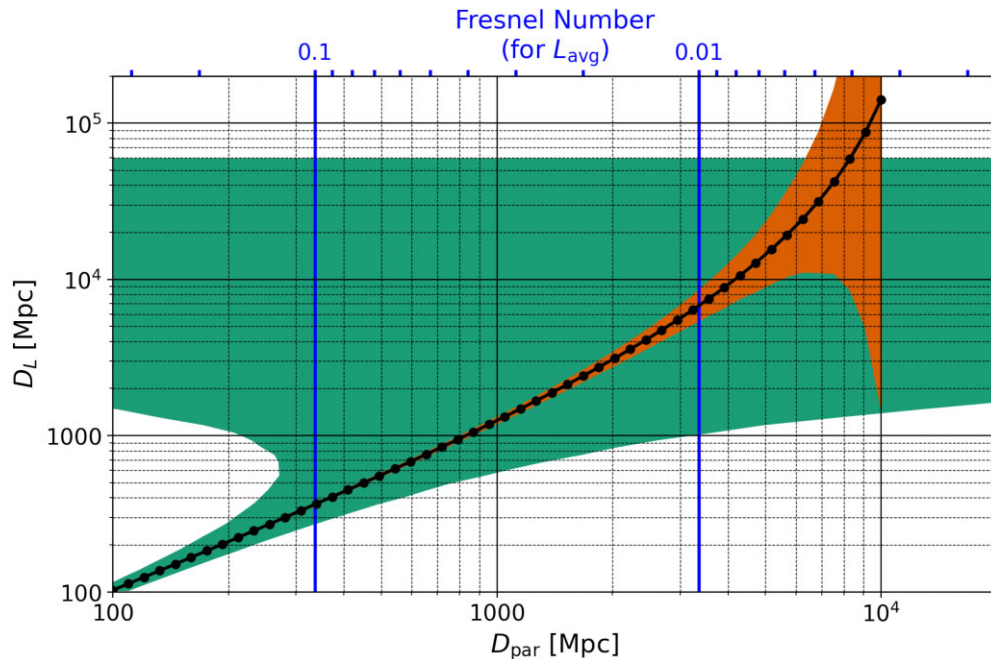


Figure 5. An example comparison of the uncertainties on the D_L and D_{par} parameters using a Fisher analysis for a single gravitational wave source (using the IIB model and default parameters). For this simulation, we set uniform uncertainties $\sigma = 10$ ns and $\sigma_L = 0.1\lambda_{\text{gw}}$. This PTA contains 100 pulsars from our fiducial PTA. The orange region shows the uncertainty range on the luminosity distance, while the green region shows the uncertainty range on the parallax distance. For reference, the Fresnel number (equation 5) for a pulsar at the average pulsar distance L_{avg} for this particular array is indicated on the top axis. Generally, the luminosity distance is easier to measure from the frequency evolution, as compared to the parallax distance from the Fresnel corrections, which for this example, is measurable for $D_{\text{par}} \lesssim 400$ Mpc.

Hence, we see that the Fisher matrix analysis demonstrates an important result, namely that two sources remove the need for prohibitively precise pulsar distance measurements a priori in order to measure the Hubble constant. Our one source simulation only obtained values of $\text{CV}_{H_0} \sim 0.1$ when all pulsars distances were known with precision $\sigma_L < 0.01\lambda_{\text{gw}}$, which is far beyond our current capabilities. However, two sources without any prior pulsar distance knowledge resulted in H_0 measurements ranging from a per cent to 10's of per cent, dependent on the relative sky positions of the two sources.

Fig. 7 shows Fisher analysis surveys of the H_0 measurability in terms of intrinsic source chirp mass and frequency, as well as PTA characteristics like timing accuracy and number of pulsars. Not surprisingly, we measure H_0 best from sources with high chirp mass and frequency, since these produce strong frequency evolution effects in the signal. For reference, we include a rough estimate of the NANOGrav 11 yr continuous wave strain upper limit, $h_{0,11\text{yr}} \approx 10^{-6} \frac{\omega_0}{\pi(1\text{ Hz})}$ (see fig. 3 of Aggarwal et al. 2019), and indicate the 1 kyr coalescence time $\Delta\tau_c$ contour. These two lines give a sense of what part of parameter space is most interesting. The lightly shaded region where $\Delta\tau_c < 1$ kyr begins to break the original model assumptions, where frequency evolution becomes very significant (for reference, see assumption xiv of MC21). The NANOGrav upper limit suggests that sources above the line would have already been seen (out to $D_L = 102.35$ Mpc, which is where the distance is fixed in this plot) in the data if they existed. This leaves us with a portion of parameter space for two sources with $\mathcal{M}_{\text{obs}} \lesssim 2 \times 10^9 M_\odot$ and frequencies $\omega_{0,\text{obs}} \gtrsim 20$ nHz where we may be able to find sources that could recover a measurement of H_0 in the future.

The right-hand panel of Fig. 7 shows that timing precision makes a significant difference in improving the measurement of H_0 . For the fiducial PTAs, we gain an order of magnitude improvement in CV_{H_0} for the same improvement in σ . Additionally, we see simply adding more fiducial PTA pulsars (randomly drawn from our distribution equation 22) improves the entire network's ability to recover H_0 . Overall, recovery of H_0 for the fiducial PTAs with constant timing uncertainties scales as $\text{CV}_{H_0} \sim \sigma/\sqrt{N_p}$.

Also plotted in Fig. 7 are the CV_{H_0} recovery uncertainties found from various SKA-era PTAs, with more realistic timing uncertainties (see Section 3.3.2). The upright triangles (optimistic) represent PTAs for which $\sigma \propto \text{SNR}^{-1/2}$ while the upside-down triangles (pessimistic) represent PTAs for which $\sigma \propto \text{SNR}^{-1}$ (equation 24). Grey (black) markers represent PTAs for which the best-timed pulsars have $\sigma = 1$ ns (10 ns). For small number arrays, the scaling with N_p approaches the $1/\sqrt{N_p}$ scaling of the constant σ case. For larger N_p this relation limits towards a constant value. This is due to the way in which we add pulsars to the array, in ranked order of SNR per radial bin, and to the limited number of high SNR pulsars at large distance. As N_p increases, we exhaust the supply of high-SNR pulsars at large distance from Earth so while N_p increases, N_p at large distance and high SNR effectively does not, and these are the pulsars which mostly strongly constrain D_{par} and hence H_0 . Given these specific SKA-era PTAs, we find that arrays composed of ~ 200 pulsars timed to a best case precision of $\sigma_{\text{min}} = 1$ ns (10 ns) could measure the Hubble constant to within 2 per cent (20 per cent) precision using the two-source method described here.

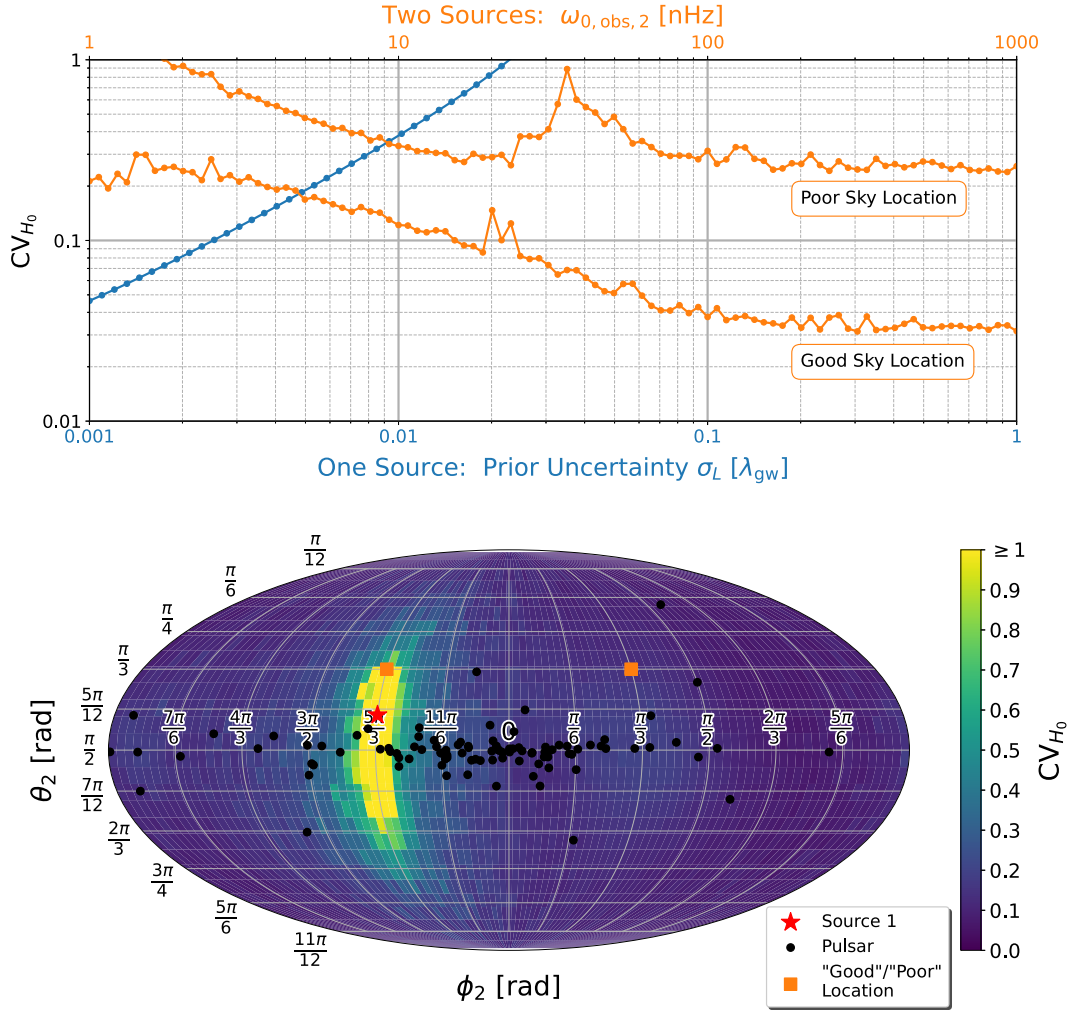


Figure 6. Fisher analysis predictions of the recovery of H_0 (using the IIB model). Both sources have $\mathcal{M}_{\text{obs}} = 2 \times 10^9 M_\odot$, default parameters otherwise, and the PTA contains 100 pulsars from the fiducial PTA with $\sigma = 1$ ns timing uncertainty. **Top panel:** A comparison of the recovery of the Hubble constant using one versus two gravitational wave sources. For one source, the recovery of H_0 is plotted as a function of the required pulsar distance prior knowledge, set uniformly on all pulsars in the PTA. For two sources, no prior knowledge on the pulsar distances is assumed. Instead, recovery of H_0 is shown as a function of the second source’s frequency. Both changing the second source’s frequency as well as its sky position relative to the first source breaks the pulsar distance wrapping cycle degeneracy, so here we show two cases: one where the second source is in a ‘good’ sky position ($\theta_2 = \pi/3$ rad, $\phi_2 = 2\pi/3$ rad), and one where it is in a ‘poor’ sky position ($\theta_2 = \pi/3$ rad, $\phi_2 = 5\pi/3$ rad). Both of these locations are marked on the sky plot in the bottom panel – these two positions lie on the same latitude, but on different parts of the sky. **Bottom panel:** H_0 recovery with two sources, as a function of the second source’s sky position. Here we see a degenerate region centred on the first source (red star) and reflected about the galactic plane. In the limit that the two sources become perfectly aligned the recovery of the Hubble constant becomes very difficult, even when the two sources have different frequencies.

5.2 Prospects for future SKA-like PTAs

To further explore the prospects of using future SKA-like PTAs to measure the Hubble constant, we examine the possibility that a detectable SMBHB source exists. Because closer sources within a few hundred Mpc will have larger Fresnel effects, and hence are more likely to produce a measurement of H_0 , we take an observational approach and consider the catalogue of PTA constraints on gravitational waves from massive SMBHBs in galaxies within 600 Mpc generated from the NANOGrav 11 yr data set in Arzoumanian et al. (2021).³ This includes 216 nearby galaxies for which PTA upper limits can constrain the mass ratios of putative SMBHBs at a given gravitational wave frequency (we consider the $f_{gw} = 10$ nHz, circular orbit constraints). In the left-hand panel of Fig. 8 we plot these limiting putative SMBHBs in chirp mass and distance space. Using a best-case, 438 pulsar, SKA-era PTA of Section 3.3.2 (5 per cent of all detected MSPs), we overlay contours of the corresponding H_0 recoverability CV_{H_0} . For this optimistic SKA-era PTA, there are 10’s of possible nearby sources.

The right-hand panel of Fig. 8 explores this further by taking each putative SMBHB (red dots) from the left-hand panel, assuming it has a twin at a different position on the sky, and computing the H_0 recovery uncertainty CV_{H_0} . The cumulative distribution reveals how increasing

³We used the data from Table 3 – the ‘Mass,’ ‘Dist,’ and ‘ q_{95} (10.0)’ columns.

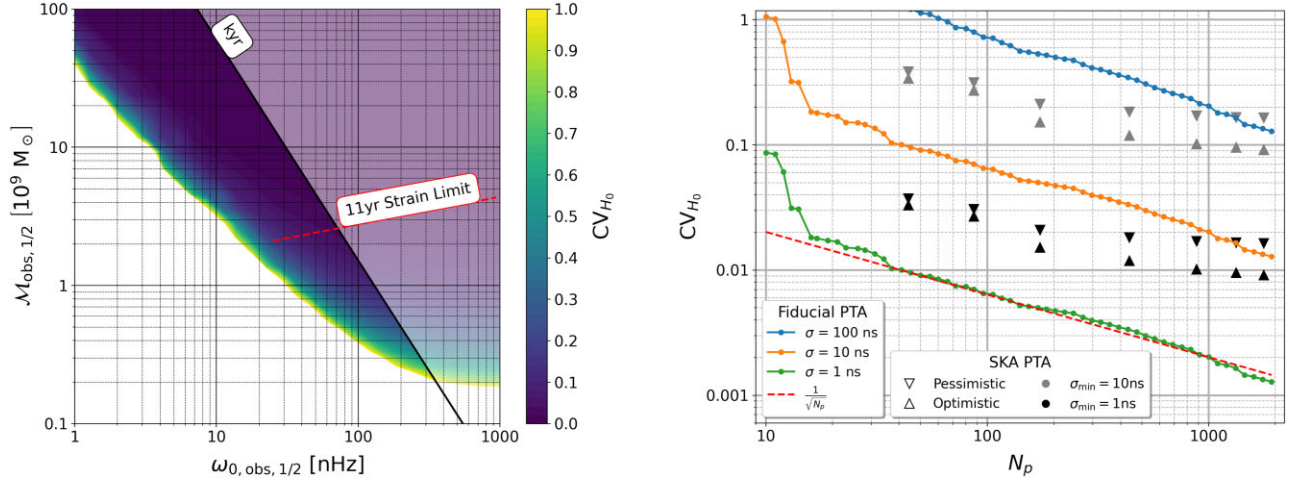


Figure 7. Fisher analysis predictions for the recovery of H_0 (using the IIB model) with two sources, as a function of the observed chirp mass and orbital frequency of the source, and as a function PTA timing uncertainty and number of pulsars. For simplicity, both sources are assumed to be identical (signified by the ‘1/2’ notation) except for being located on different parts of the sky. Source 1 is at the default location, while source 2 has sky angles $\theta_2 = \frac{\pi}{3}$ rad and $\phi_2 = \frac{\pi}{5}$ rad. **Left-hand panel:** Here 100 pulsars from our fiducial PTA are timed with $\sigma = 1$ ns uncertainty. Hypothetical sources with $\Delta\tau_c < 1$ kyr have been shaded out as these represent highly chirping systems (see the general discussion of assumptions in MC21). Additionally the results of Aggarwal et al. (2019) are used to include a rough estimate of NANOGrav’s 11 yr continuous wave strain upper limit, $h_{0,11\text{yr}} \approx 10^{-6} \frac{\omega_0}{\pi(1\text{ Hz})}$, as reference. **Right-hand panel:** Here both sources have $M_{\text{obs}} = 10^{10} M_\odot$. From our fiducial PTA which has constant timing uncertainty, we see that the recovery of H_0 scales as $CV_{H_0} \propto \sigma/\sqrt{N_p}$. For variable timing uncertainty, the SKA PTA no longer achieves the same improvement by adding more pulsars, but rather now scales as $CV_{H_0} \propto \sigma_{\min}$ at large N_p .

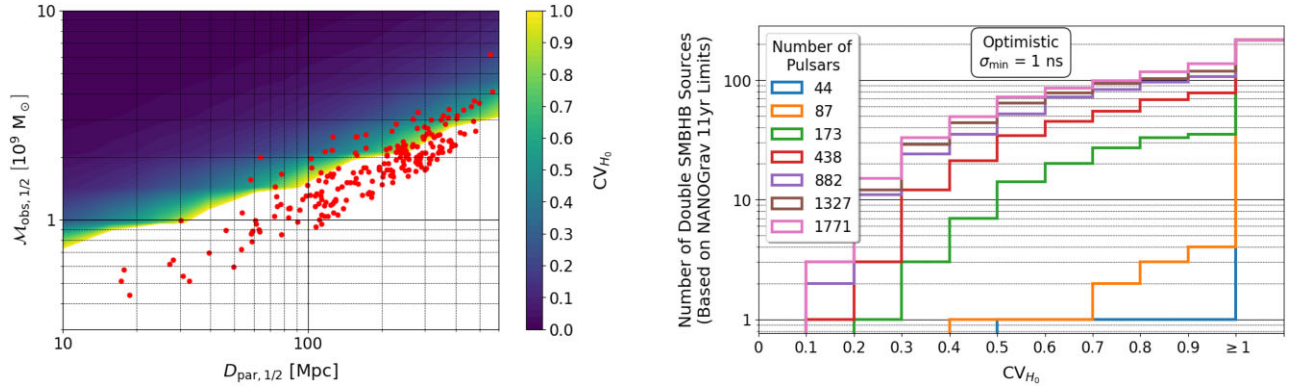


Figure 8. **Left-hand panel:** Fisher analysis predictions of the recovery of H_0 (using the IIB model) with two sources using 438 (5 per cent) of our ‘optimistic’ SKA-era pulsars with $\sigma_{\min} = 1$ ns, as a function of the observed chirp mass and parallax distance of the source. In the contour plot, for simplicity, both SMBHB sources are assumed to be identical (hence the ‘1/2’ notation) except for their sky location. Source 1 is at the default location, while source 2 has sky angles $\theta_2 = \frac{\pi}{3}$ rad and $\phi_2 = \frac{\pi}{5}$ rad. Additionally all sources are simulated with $f_{\text{gw}} = 10$ nHz (i.e. $\omega_{0,\text{obs}} \approx 31.4$ nHz). Plotted as red dots on top of the contours, we compare 216 hypothetical SMBHB sources within 600 Mpc published in Arzoumanian et al. (2021). This gives us a sense of the quality required of an SKA-era PTA in order to make the H_0 measurement. For the same number of pulsars, the pessimistic version of this PTA resulted in a marginal number of sources with $CV_{H_0} < 1$. **Right-hand panel:** For each of the 216 SMBHB sources indicated on the left-hand panel, we simulate two identical SMBHBs with those mass and distance values (and with the indicated SKA-era PTAs), and bin them based on their level of recovery of H_0 in a cumulative distribution. Both SMBHB sources are assumed to be identical with the same caveats as the left-hand panel. As expected from the right-hand panel of Fig. 7, adding more pulsars into the PTA improves the recovery of H_0 in more two-system scenarios. For the PTAs with > 173 pulsars, we begin to find double sources where $CV_{H_0} \leq 0.3$ (comparable to what is shown in Fig. 4). These plots do not rule out the possibility of a measurement of H_0 with a future SKA PTA, and instead suggests that there do exist hypothetical SMBHB candidates wherein this measurement may be viable.

the size of our PTA increases the number of twin SMBHBs for which H_0 recovery at a given precision is possible. With 438 pulsars or more, we find 10’s of systems with $CV_{H_0} \leq 0.5$, and over 30 per cent of all of these systems procure $CV_{H_0} \leq 0.9$. Over 5 per cent (i.e. more than ~ 12) of these putative systems result in $CV_{H_0} \leq 0.4$, and there are a few cases where we recover $CV_{H_0} \sim 0.1$.

Hence, an $\mathcal{O}(10\text{ per cent})$, purely gravitational wave H_0 measurement is possible if two of these optimal putative SMBHBs exist. Note that already for the ‘pessimistic,’ $\sigma_{\min} = 1$ ns SKA-era arrays (see the right-hand panel of Fig. 7), the overlap between putative sources and $CV_{H_0} \leq 1$ in Fig. 8 becomes marginal. Hence these best-case PTAs are likely required.

6 CONCLUSIONS

In this work, we have shown that future PTA experiments could make purely gravitational wave-based measurements of the Hubble constant. This is made possible by accounting for the Fresnel curvature effects in the wavefront across the Earth-pulsar baseline. By using the fully general Fresnel frequency evolution timing residual model, we can obtain two separate distance measurements to the source: the luminosity distance D_L (from the frequency evolution effects) and the parallax distance D_{par} (from the Fresnel effects).

The measurement of the parallax distance is particularly challenging due to the pulsar distance wrapping problem. Unless the distances to the pulsars in our array can be measured to sub-gravitational wavelength (\sim sub-parsec) precision, this measurement cannot be made with a single SMBHB source. However, we demonstrate that two SMBHB sources detected simultaneously break the wrapping cycle degeneracy, allowing a viable scenario for measuring the Hubble constant. Two or more sources will actively calibrate the PTA by helping to identify the true pulsar distances from the degenerate secondary modes. The ability to better resolve the pulsar distances using this method will be explored in a follow-up study (McGrath, D’Orazio, & Creighton, in preparation). Additionally, this method may better improve the recovery of the luminosity distance D_L (to both sources), even at higher redshifts where the Fresnel effects may be too small to measure D_{par} . It would be interesting for future studies to investigate how significant this improvement is, compared to measuring D_L of just a single source, since the luminosity distance by itself would still be useful for other astrophysical and cosmological studies.

For the two source problem, we explored the measurement of H_0 for a wide range of systems and PTAs. We developed new Fisher matrix-based criteria which can quickly predict the ability of a particular source-PTA set-up to measure H_0 , and facilitate follow-up with MCMC verification. While our MCMC techniques are less complicated than other samplers for PTA problems, we find they support our proof-of-principle calculations, and we leave it to future work to improve upon methods for efficiently extracting the model parameters (see for example, Samajdar et al. 2022).

With fiducial PTAs, with constant timing uncertainty σ and N_p pulsars, the precision in our simulated H_0 measurement scales as $\text{CV}_{H_0} \sim \sigma/\sqrt{N_p}$. With more realistic SKA-era PTAs, we observe a similar scaling for arrays with less than a few hundred pulsars and a saturation in measurement precision for larger arrays, where $\text{CV}_{H_0} \sim \sigma_{\text{min}}$. We also find that hypothetical two-source SMBHB systems, based on the NANOGrav 11 yr data in Arzoumanian et al. (2021), suggest that SKA-era PTAs may be capable of measuring H_0 to as low as $\text{CV}_{H_0} \sim \mathcal{O}(0.1)$. This will likely require fortuitous SMBHB sources and ‘best-case’ PTAs (such that $\sigma_{\text{min}} \sim 1$ ns to 10 ns) containing several hundred pulsars. While we have not considered the possibility of more than two sources here, this will likely improve measurement prospects.

In conclusion, the existence of two individually resolvable gravitational wave signals from inspiraling SMBHBs allows a purely gravitational wave-based measurement of the Hubble constant with PTAs in the SKA-era and beyond. Current and future H_0 measurements can achieve higher precision than what we envision here on that time-scale; for example, the standard sirens approach using binary neutron star mergers with electromagnetic counterparts could reach a 1 per cent level H_0 measurement by the 2030s (Chen, Fishbach & Holz 2018). However, as is apparent from the current Hubble tension, precision does not guarantee agreement between measurement methods. The value of the method discussed here, using only the gravitational messenger and geometry, is that it can provide a novel and independent measurement of H_0 .

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Aeronautics and Space Administration (NASA) under award number 80GSFC21M0002. This work was also supported by the National Science Foundation (NSF) PHY-1430284 [through the North American Nanohertz Observatory for Gravitational Waves (NANOGrav’s) Physics Frontier Center], PHY-1912649, and PHY-2207728, and by UW-Milwaukee’s computational resources PHY-1626190. DJD received funding from the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No. 101029157, and from the Danish Independent Research Fund through Sapere Aude Starting Grant No. 121587. We thank the referee, Neil Cornish, for a constructive report.

DATA AVAILABILITY

Calculations in this paper were performed using code developed by the authors, and the data are available on reasonable request to the authors.

REFERENCES

- Aggarwal K. et al., 2019, *ApJ*, 880, 116
- Arzoumanian Z. et al., 2018, *ApJS*, 235, 37
- Arzoumanian Z. et al., 2021, *ApJ*, 914, 121
- Babak S., Sesana A., 2012, *Phys. Rev. D*, 85, 044034
- Bates S. D., Lorimer D. R., Rane A., Swiggum J., 2014, *MNRAS*, 439, 2893
- Caldwell R. R., 1993, *Phys. Rev. D*, 48, 4688
- Chen H.-Y., Fishbach M., Holz D. E., 2018, *Nature*, 562, 545
- Corbin V., Cornish N. J., 2010, preprint ([arXiv:1008.1782](https://arxiv.org/abs/1008.1782))
- Cordes J. M., Chernoff D. F., 1997, *ApJ*, 482, 971
- Creighton J. D. E., Anderson W. G., 2011, *Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis*. Wiley-VCH, Weinheim

- D’Orazio D. J., Loeb A., 2021, *Phys. Rev. D*, 104, 063015
- Deller A. T. et al., 2019, *ApJ*, 875, 100
- Deng X., Finn L. S., 2011, *MNRAS*, 414, 50
- Feeney S. M., Peiris H. V., Williamson A. R., Nissanke S. M., Mortlock D. J., Alsing J., Scolnic D., 2019, *Phys. Rev. Lett.*, 122, 061105
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Ghosh T., Biswas B., Bose S., 2022, preprint (arXiv:2203.11756)
- Hogg D. W., 1999, preprint (arXiv:astro-ph/9905116)
- Holz D. E., Hughes S. A., 2005, *ApJ*, 629, 15
- Iacovelli F., Mancarella M., Foffa S., Maggiore M., 2022, preprint (arXiv:2207.02771)
- Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2018, *MNRAS*, 477, 964
- Lee K. J., Wex N., Kramer M., Stappers B. W., Bassa C. G., Janssen G. H., Karuppusamy R., Smits R., 2011, *MNRAS*, 414, 3251
- Lorimer D. R. et al., 2006, *MNRAS*, 372, 777
- Maggiore M., 2008, *Gravitational Waves: Theory and Experiments*. Vol. 1, Oxford Univ. Press, Oxford
- Maggiore M., 2018, *Gravitational Waves: Astrophysics and Cosmology*. Vol. 2, Oxford Univ. Press, Oxford
- McGrath C., Creighton J., 2021, *MNRAS*, 505, 3113
- Messenger C., Read J., 2012, *Phys. Rev. Lett.*, 108, 091101
- Petiteau A., Babak S., Sesana A., de Araújo M., 2013, *Phys. Rev. D*, 87, 064036
- Qian Y.-Q., Mohanty S. D., Wang Y., 2022, *Phys. Rev. D*, 106, 023016
- Samajdar A. et al., 2022, preprint (arXiv:2205.04332)
- Schneider P., 2015, *Extragalactic Astronomy and Cosmology: An Introduction*, 2nd edn. Springer, Berlin, p. 177
- Schutz B. F., 1986, *Nature*, 323, 310
- Shiralilou B., Raaijmakers G., Dubouef B., Nissanke S., Foucart F., Hinderer T., Williamson A., 2022, preprint (arXiv:2207.11792)
- Smits R., Kramer M., Stappers B., Lorimer D. R., Cordes J., Faulkner A., 2009, *A&A*, 493, 1161
- The LIGO Scientific Collaboration, the Virgo Collaboration, the 1M2H Collaboration, the Dark Energy Camera GW-EM Collaboration, the DES Collaboration, 2017, *Nature*, 551, 85
- The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, 2021, preprint (arXiv:2111.03604)
- Verbiest J. P. W., Shaifullah G. M., 2018, *Class. Quantum Gravity*, 35, 133001
- Wang L.-F., Zhang G.-P., Shao Y., Zhang X., 2022, preprint (arXiv:2201.00607)
- Weinberg S., 1972, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley, New York

APPENDIX A: COSMOLOGICAL DISTANCES

For a gravitational wave propagating from our source (at comoving coordinate distance R and redshift z) to the Earth (‘infalling’), we define the following ‘boundary’ conditions:

Source : $r = R, \quad z' = z, \quad t' = t_{\text{ret}},$

Earth : $r = 0, \quad z' = 0, \quad t' = t,$

where our reference frame is centred on the Earth. From the Friedmann equations and the definition of the redshift in terms of the present day t_0 scale factor, $1 + z = \frac{a(t_0)}{a(t)}$, we explicitly list the following important quantities (Weinberg 1972; Hogg 1999; Schneider 2015):

$$\text{Expansion Rate} \quad H \quad \equiv \left(\frac{\dot{a}(t)}{a(t)} \right) = H_0 E, \quad (\text{A1})$$

$$\begin{aligned} \text{Dimensionless Hubble Function} \quad E &\equiv \left[\frac{\Omega_r}{a^4(t)} + \frac{\Omega_m}{a^3(t)} + \Omega_\Lambda + \frac{\Omega_k}{a^2(t)} \right]^{1/2}, \\ &= \left[\frac{(1+z)^4}{a^4(t_0)} \Omega_r + \frac{(1+z)^3}{a^3(t_0)} \Omega_m + \Omega_\Lambda + \frac{(1+z)^2}{a^2(t_0)} \Omega_k \right]^{1/2}, \end{aligned} \quad (\text{A2})$$

$$\text{Curvature Density} \quad \Omega_k \quad \equiv a^2(t_0) \left[1 - \frac{\Omega_r}{a^4(t_0)} - \frac{\Omega_m}{a^3(t_0)} - \Omega_\Lambda \right] = -k D_H^2, \quad (\text{A3})$$

$$\text{Hubble Distance} \quad D_H \quad \equiv \frac{c}{H_0}, \quad (\text{A4})$$

$$\begin{aligned} \text{Line-of-Sight Comoving Distance} \quad D_c &\equiv D_H \int_0^z \frac{dz'}{E}, \\ &= c a(t_0) \int_{t_{\text{ret}}}^t \frac{dt'}{a}, \end{aligned} \quad (\text{A5})$$

$$\text{Parallax Distance} \quad D_{\text{par}} \quad \equiv a(t_0) \frac{R}{\sqrt{1 - k R^2}}, \quad (\text{A6})$$

where H_0 is the Hubble constant and Ω_r , Ω_m , and Ω_Λ are the radiation, matter, and vacuum density parameters, respectively. For completeness here we will continue to write the present day scale factor explicitly, but typical convention is to normalize this to $a(t_0) \equiv 1$. The two forms of the line-of-sight comoving distance in equation (A5) are made through the change of variables $t \rightarrow z$ using the definition of z and equation (A1).

The luminosity distance of our source is defined by considering the energy flux F_{obs} (in this case, of our gravitational waves) as measured by the observer (Maggiore 2008). In a cosmologically expanding universe the observed energy E_{obs} is redshifted compared to the energy that

was emitted in the source frame E_s , such that $E_s = (1 + z)E_{\text{obs}}$. Additionally, time dilation means the time measured by the observer and source clocks are related by $t_s = \frac{t_{\text{obs}}}{(1+z)}$. Finally, from the FLRW metric equation (1), the flux at time t will be spread over a total area of $A(t) = 4\pi a^2(t)R^2$. This means that we can write:

$$F_{\text{obs}}(t) \equiv \frac{L_{\text{obs}}}{A(t)} = \frac{L_s}{(1+z)^2 4\pi a^2(t)R^2} = \frac{L_s}{4\pi [(1+z)^2 a^2(t)R^2]} \equiv \frac{L_s}{4\pi D_L^2}, \quad (\text{A7})$$

where in the final equality we define our luminosity distance D_L such that the observed flux matches our standard notion of flux, but as a function of the source frame luminosity L_s . Therefore at the present time, we have:

$$D_L \equiv (1+z)a(t_0)R. \quad (\text{A8})$$

Next we draw the connection between the comoving coordinate distance R , the luminosity distance D_L , and the line-of-sight comoving distance D_c . Integrating the metric equation (1) along the radial ($d\theta = d\phi = 0$), null ($ds = 0$), infalling path of the gravitational wave, and using equation (A5) gives us:

$$-\int_R^0 \frac{dr}{\sqrt{1-kr^2}} = \int_0^t \frac{c}{a(t')} dt',$$

$$= \frac{D_c}{a(t_0)},$$

The solution of the left-hand integral depends on the sign of the curvature constant k (or alternatively, the density parameter Ω_k). Solving this integral we now have:

$$R \equiv D_M = \begin{cases} \frac{1}{\sqrt{k}} \sin\left(\frac{\sqrt{k}}{a(t_0)} D_c\right), & k > 0 \\ \frac{1}{a(t_0)} D_c, & k = 0 \\ \frac{1}{\sqrt{|k|}} \sinh\left(\frac{\sqrt{|k|}}{a(t_0)} D_c\right), & k < 0 \end{cases} = \begin{cases} \frac{D_H}{\sqrt{|\Omega_k|}} \sin\left(\frac{\sqrt{|\Omega_k|}}{a(t_0)D_H} D_c\right), & \Omega_k < 0 \\ \frac{1}{a(t_0)} D_c, & \Omega_k = 0 \\ \frac{D_H}{\sqrt{\Omega_k}} \sinh\left(\frac{\sqrt{\Omega_k}}{a(t_0)D_H} D_c\right), & \Omega_k > 0 \end{cases} = \frac{D_L}{a(t_0)(1+z)}, \quad (\text{A9})$$

where D_M is introduced as the ‘transverse comoving distance’ following Hogg, which is just a re-naming of the comoving coordinate distance. Thus depending on the global universe geometry, the relationship between the various distances R , D_L , D_c , and D_M are given by equation (A9).

One final and very important distance for our consideration is the ‘parallax distance’ D_{par} in equation (A6) (Weinberg 1972). Substituting equation (A9) into equation (A6) lets us write:

$$D_{\text{par}} \equiv \frac{a(t_0)R}{\sqrt{1-kR^2}} = \frac{a(t_0)D_M}{\sqrt{1-kD_M^2}} = \begin{cases} \frac{a(t_0)}{\sqrt{k}} \tan\left(\frac{\sqrt{k}}{a(t_0)} D_c\right), & k > 0 \\ D_c, & k = 0 \\ \frac{a(t_0)}{\sqrt{|k|}} \tanh\left(\frac{\sqrt{|k|}}{a(t_0)} D_c\right), & k < 0 \end{cases} = \frac{D_L}{(1+z)\sqrt{1-\frac{kD_L^2}{a^2(t_0)(1+z)^2}}}. \quad (\text{A10})$$

Notably in a flat universe (with $a(t_0) = 1$ normalization) we have $D_{\text{par}} = R = D_M = D_c = D_L/(1+z)$.

As a final note, the closed universe case needs further special consideration, as the parallax distance not only diverges, but can take on negative values. For closed geometry, the expression for D_{par} in equations (A6) and (A10) has a \pm sign in front of the expression. Conceptually, consider a point at the north pole of a globe generating some wave. As the wave propagates away from the source, in the Northern hemisphere two points on the wavefront (constant latitude) would produce a positive parallax measurement $D_{\text{par}} > 0$ back to the source (the wavefront curves towards the source and away from the direction of propagation). At the equator, however, all points along the wavefront will experience a plane-wave. Therefore there is no well-defined parallax distance for points along the equator, and mathematically we see that the value of $D_{\text{par}} \rightarrow \infty$ (i.e. at $R = R_{\text{max}} = \frac{1}{\sqrt{k}}$, or equivalently $\chi = \frac{1}{\sqrt{k}} \frac{\pi}{2}$ through equation 2).

Then interestingly, in the Southern hemisphere two points on the wavefront (constant latitude) would now produce a *negative* parallax measurement D_{par} back to the source (the wavefront curves away the source and towards the direction of propagation – towards the source antipode). Using equation (2) to replace R with χ in equation (A10), and including the \pm sign out front, we write:

$$D_{\text{par}} = \pm \frac{a(t_0) \sin(\sqrt{k}\chi)}{\sqrt{k}\sqrt{1-\sin^2(\sqrt{k}\chi)}} = \pm \frac{a(t_0) \sin(\sqrt{k}\chi)}{\sqrt{k}\sqrt{\cos^2(\sqrt{k}\chi)}} = \frac{a(t_0)}{\sqrt{k}} \tan(\sqrt{k}\chi). \quad (\text{for } k > 0)$$

Writing D_{par} in terms of χ takes care of the sign ambiguity in equations (A6) and (A10). For values $0 \leq \chi < \frac{1}{\sqrt{k}} \frac{\pi}{2}$ (i.e. the Northern hemisphere) the parallax distance is positive, for $\chi = \frac{1}{\sqrt{k}} \frac{\pi}{2}$ (i.e. the equator) the parallax distance diverges, and for values $\frac{1}{\sqrt{k}} \frac{\pi}{2} < \chi \leq \frac{1}{\sqrt{k}} \pi$ (i.e. the Southern hemisphere) the parallax distance is negative.

APPENDIX B: THE RETARDED TIME CALCULATION

The goal is to calculate $|\vec{x}' - \vec{x}|$ in equation (3), where \vec{x}' is the source location and \vec{x} is the field point. For our problem, we will primarily be interested in setting our field point at the pulsar. With the centre of our coordinate system at the Earth, let the source be at the coordinate $r = R$ (i.e. $\chi = \chi_s$) with unit vector \hat{r} , and let the pulsar be at the coordinate $r = L$ (i.e. $\chi = \chi_p$) with unit vector \hat{p} .

Next we write the law of cosines for Euclidean, spherical, and hyperbolic geometries:

$$\begin{cases} \cos(\sqrt{k}|\vec{x}' - \vec{x}|) = \cos(\sqrt{k}\chi_S) \cos(\sqrt{k}\chi_P) + \sin(\sqrt{k}\chi_S) \sin(\sqrt{k}\chi_P) (\hat{r} \cdot \hat{p}), & k > 0 \\ |\vec{x}' - \vec{x}|^2 = \chi_S^2 + \chi_P^2 - 2\chi_S\chi_P (\hat{r} \cdot \hat{p}), & k = 0 \\ \cosh(\sqrt{|k|}|\vec{x}' - \vec{x}|) = \cosh(\sqrt{|k|}\chi_S) \cosh(\sqrt{|k|}\chi_P) - \sinh(\sqrt{|k|}\chi_S) \sinh(\sqrt{|k|}\chi_P) (\hat{r} \cdot \hat{p}). & k < 0 \end{cases} \quad (\text{B1})$$

Recall from Section 2.2 that at the present time the Gaussian curvature is k . One approach to solving these equations is to Taylor expand in the small parameter $\sqrt{|k|}\chi_P$ in the closed and open universe cases, and in the small parameter $\frac{\chi_P}{\chi_S}$ the flat universe case. The result is:

$$|\vec{x}' - \vec{x}| \approx \chi_S - (\hat{r} \cdot \hat{p}) \chi_P + \frac{1}{2} (1 - (\hat{r} \cdot \hat{p})^2) \begin{cases} \frac{\chi_P^2}{\frac{1}{\sqrt{k}} \tan(\sqrt{k}\chi_S)}, & k > 0 \\ \frac{\chi_P^2}{\chi_S}, & k = 0 \\ \frac{\chi_P^2}{\frac{1}{\sqrt{|k|}} \tanh(\sqrt{|k|}\chi_S)}. & k < 0 \end{cases} \quad (\text{B2})$$

Additionally, using equation (2) we see that $\chi_P \approx L$ for each universe geometry case (Taylor expanding once again for the closed and open cases in $\sqrt{|k|}\chi_P$).

Consider for a moment a gravitational wave propagating directly from the source to the Earth. The wave leaves its position at redshift z and time t_{ret} , and arrives at Earth at redshift $z = 0$ at time t (recall the boundary conditions from Appendix A). In this case our field position $\vec{x} = 0$, so $\eta_{\text{ret}} = \eta - \frac{\chi_S}{c}$ from equation (3). Therefore using equation (A5) we can write:

$$\begin{aligned} -\frac{\chi_S}{c} &= \eta_{\text{ret}} - \eta = \int_t^{t_{\text{ret}}} \frac{dt'}{a} \\ &= -\frac{D_c}{a(t_0)c}. \end{aligned}$$

This gives us a connection between χ_S and the comoving coordinate distance D_c , namely that $\chi_S = \frac{D_c}{a(t_0)}$, which we can now substitute into equation (B2). The result is that we can now see that the relevant distance which appears in the Fresnel term in our Taylor expansion (for all three universe cases) is the parallax distance D_{par} (substitute in equation A10). Therefore the final form of equation (B2) which is then used in equation (4) is:

$$|\vec{x}' - \vec{x}| \approx \frac{D_c}{a(t_0)} - (\hat{r} \cdot \hat{p}) L + \frac{1}{2} (1 - (\hat{r} \cdot \hat{p})^2) a(t_0) \frac{L^2}{D_{\text{par}}}. \quad (\text{B3})$$

Note that in the flat universe case $D_{\text{par}} = D_c$, which is consistent with the result of DL21 given that this was their working assumption.

APPENDIX C: ASSUMPTIONS

Assumptions (ii)–(xvi) of MC21 carry over to this work. Additionally, we work under the following assumptions:

- (i) The universe is described by the FLRW metric with comoving coordinates (t, r, θ, ϕ) and curvature constant k .
- (ii) The FLRW scale factor a does not change appreciably over the time-scale of the traveltime of the photon from the pulsar to the Earth: $\dot{a}(t_0)\Delta t_{\text{p} \rightarrow \text{E}}$.
- (iii) All ‘local’ distances and times between the pulsar and the Earth are on a Minkowski background, $g_{\mu\nu} = \eta_{\mu\nu} = \text{diag}(-c^2, 1, 1, 1)$. The transition from the global cosmological FLRW background to the local Minkowski background requires that all points of interest have comoving coordinate distance separations much smaller than the background curvature ($r^2 \ll \frac{1}{|k|}$), and that the observation time of our experiment is much smaller than the present day age of the universe ($\frac{t_{\text{obs}}}{t_0} \ll 1$).
- (iv) For the results presented in this work we assume a flat universe ($k = 0$).

APPENDIX D: PULSAR POPULATIONS AND EXAMPLE PULSAR DISTANCE POSTERIOIRS

Two pulsar populations are used for different studies within this paper, a ‘fiducial PTA’ and an ‘SKA-era PTA’. The general characteristics of these PTAs are shown in Fig. D1.

To create the SKA-era PTAs, we use the `PSRPOP` package (Bates et al. 2014) to generate a realistic population of MSPs detectable in the SKA era. To generate the underlying population of pulsars we assume the default model in `PSRPOP` (except that we place the galactic center at 8 instead of 8.5 kpc), with a spatial pulsar distribution following Lorimer et al. (2006), but with the pulsar period distribution from Cordes & Chernoff (1997), chosen specifically for MSPs. We then follow (Smits et al. 2009) in assuming that there are 3×10^4 MSPs in the galaxy to generate the discoverable population.

We simulate an SKA-like survey by choosing survey parameters following Smits et al. (2009) (model A) and summarized in Table D1. We find that our SKA survey detects 8893 of these MSPs with SNR greater than a threshold value of 9, at distances ranging from 0.2 kpc out to 18 kpc. We assume that only some percentage of these MSPs will be suitable for high precision timing, and enforce this as described by the selection process in Section 3.3.2. We then compute a timing uncertainty for each MSP as per equation (24). Timing uncertainty distributions for the different choices of uncertainty scaling with SNR are plotted in Fig. D2.

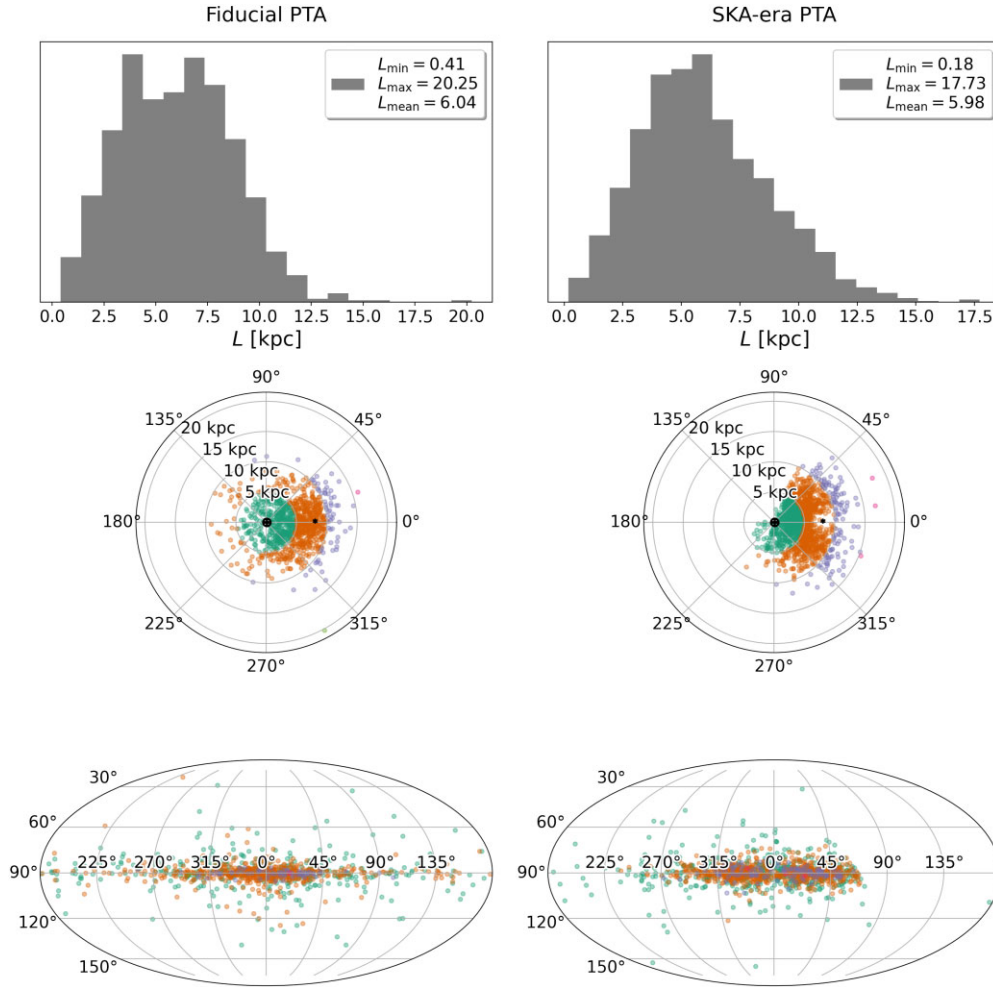


Figure D1. A side-by-side comparison of the fiducial PTA (left-hand column) and SKA-era PTA (right-hand column). These figures show the general distribution properties, in terms of the pulsar distances and locations throughout the Milky Way Galaxy. In the middle row, Earth is located at the origin and galactic centre is located at 8 kpc (at $\phi = 0^\circ$). **Left-hand panel:** This fiducial PTA contains 1000 pulsars generated from the distribution in equation (22). **Right-hand panel:** This SKA-era PTA contains 1194 pulsars.

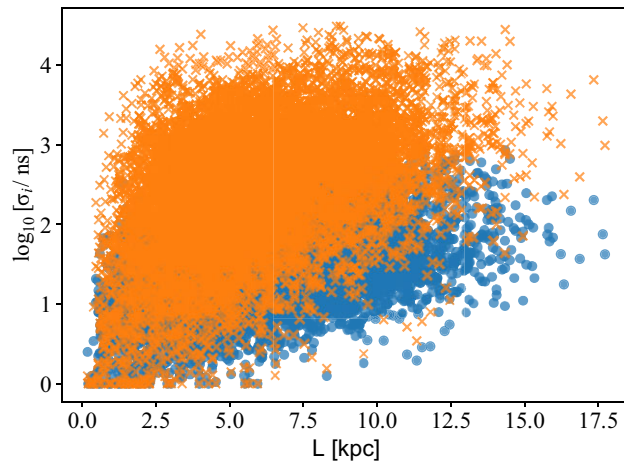


Figure D2. Timing uncertainty distributions of all of the MSPs detected in a mock SKA-era population, with $\sigma_{\min} = 1$ ns. The blue circles and orange crosses denote optimistic and pessimistic choices for σ_a , as written in equation (24). Mock SKA-era PTAs are generated by choosing MSPs from this set, as outlined in the text.

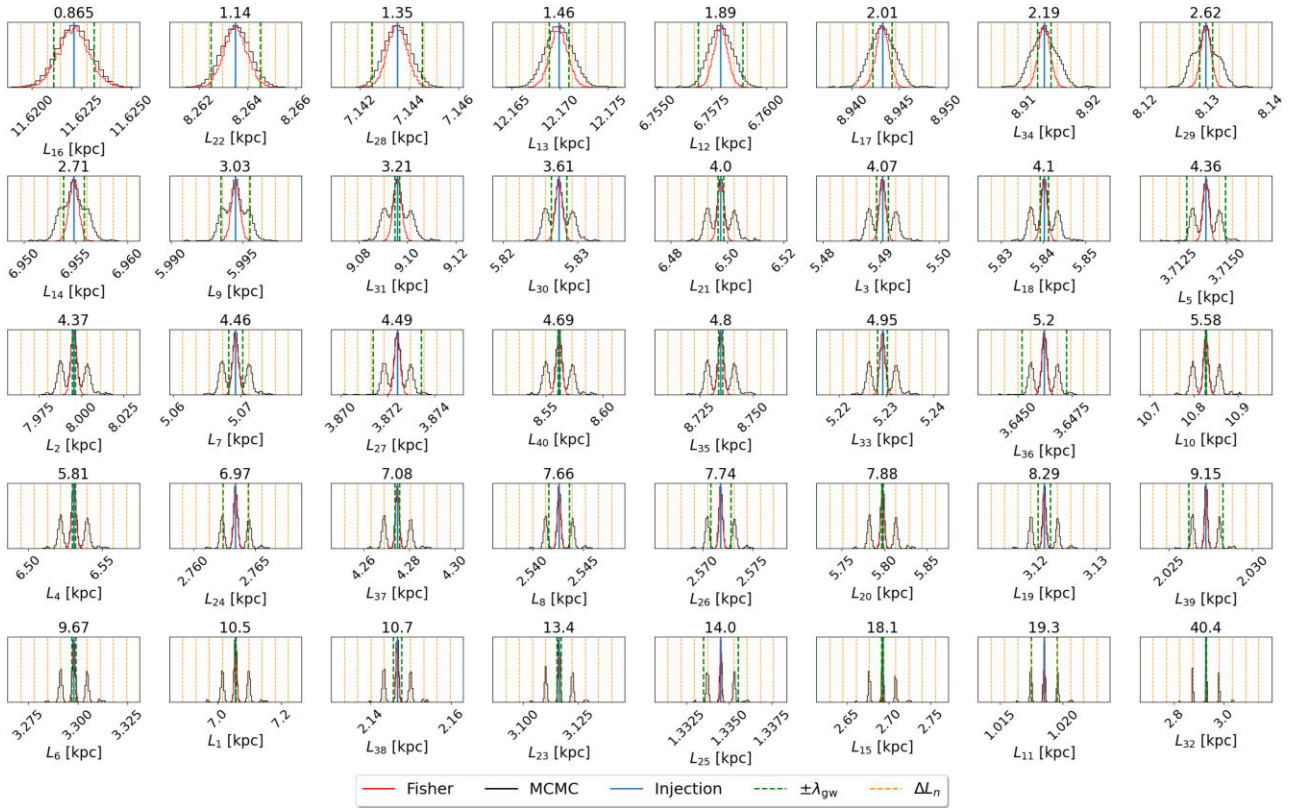


Figure D3. The complete set of 1D posteriors for the same results shown in the right-hand panel of Fig. 2 using the IB model. Here we have sorted the pulsars in ascending value by the modal overlap criterion ratio defined in equation (28) (given above each figure). As we can see, this particular criterion successfully shows the progression of the error envelope from ‘overlapping’ to ‘resolved,’ with a total of 12 overlapping, 18 intermediate, and 10 resolved pulsars. For ratios less than about 4, we see that uncertainties about each mode are both sufficiently large and sufficiently close to the primary mode that they blend together around the true value. However, for ratios between about 4 and 8 these secondary modes begin to separate away from the primary mode far enough that they start to become distinguishable from the true mode. Finally, for ratios greater than about 8, we see that the secondary modes are distinguishable from the true mode. In practice this therefore motivates us to search our results to see what the value of this ratio is for each of our pulsars in our desired source-PTA simulation. If it is greater than 8, then we would expect that the true distance to that pulsar would be identifiable. Note the samples in this figure have been cleaned, such that all samples with log-posterior ≤ -100 have been removed from the data set (see the discussion of cleaning in Section 4.3).

Table D1. SKA survey parameters motivated by Smits et al. (2009) (their Model A), used to generate mock PTAs via the PSRPOPpy code (Bates et al. 2014).

Name	Value
Survey degradation factor	1.0
Antenna gain (K Jy^{-1})	130
Integration time (s)	1800
Sampling time (ms)	0.1
System temperature (K)	30
Centre frequency (MHz)	1400
Bandwidth (MHz)	500
Channel bandwidth (MHz)	0.009
No. polarizations	2
FWHM (arcmin)	65.5
Min RA (deg)	0
Max RA (deg)	360
Min DEC (deg)	-90
Max DEC (deg)	30
Frac. survey coverage.	1
SNR threshold	9

Finally, Fig. D3 shows the complete set of 1D posteriors for the 40 pulsar PTA shown in the right-hand panel of Fig. 2, which motivates the modal overlap criterion defined in equation (28).