



## UvA-DARE (Digital Academic Repository)

### European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive slack hole binaries

Babak, S.; Petiteau, A.; Sesana, A.; Brem, P.; Rosado, P.A.; Taylor, S.R.; Lassus, A.; Hessels, J.W.T.; Bassa, C.G.; Burgay, M.; Caballero, R.N.; Champion, D.J.; Cognard, I.; Desvignes, G.; Gair, J.R.; Guillemot, L.; Janssen, G.H.; Karuppusamy, R.; Kramer, M.; Lazarus, P.; Lee, K.J.; Lentati, L.; Liu, K.; Mingarelli, C.M.F.; Osłowski, S.; Perrodin, D.; Possenti, A.; Purver, M.B.; Sanidas, S.; Smits, R.; Stappers, B.; Theureau, G.; Tiburzi, C.; van Haasteren, R.; Vecchio, A.; Verbiest, J.P.W.

**DOI**

[10.1093/mnras/stv2092](https://doi.org/10.1093/mnras/stv2092)

**Publication date**

2015

**Document Version**

Final published version

**Published in**

Monthly Notices of the Royal Astronomical Society

[Link to publication](#)

**Citation for published version (APA):**

Babak, S., Petiteau, A., Sesana, A., Brem, P., Rosado, P. A., Taylor, S. R., Lassus, A., Hessels, J. W. T., Bassa, C. G., Burgay, M., Caballero, R. N., Champion, D. J., Cognard, I., Desvignes, G., Gair, J. R., Guillemot, L., Janssen, G. H., Karuppusamy, R., Kramer, M., ... Verbiest, J. P. W. (2015). European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive slack hole binaries. *Monthly Notices of the Royal Astronomical Society*, 455(2), 1665-1679. <https://doi.org/10.1093/mnras/stv2092>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

# European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive black hole binaries

S. Babak,<sup>1★</sup> A. Petiteau,<sup>2★</sup> A. Sesana,<sup>1,3★</sup> P. Brem,<sup>1</sup> P. A. Rosado,<sup>4,5</sup> S. R. Taylor,<sup>6,7</sup> A. Lassus,<sup>8,9</sup> J. W. T. Hessels,<sup>10,11</sup> C. G. Bassa,<sup>10,12</sup> M. Burgay,<sup>13</sup> R. N. Caballero,<sup>8</sup> D. J. Champion,<sup>8</sup> I. Cognard,<sup>9,14</sup> G. Desvignes,<sup>8</sup> J. R. Gair,<sup>7</sup> L. Guillemot,<sup>9,14</sup> G. H. Janssen,<sup>10,12</sup> R. Karuppusamy,<sup>8</sup> M. Kramer,<sup>8,12</sup> P. Lazarus,<sup>8</sup> K. J. Lee,<sup>15</sup> L. Lentati,<sup>16</sup> K. Liu,<sup>8</sup> C. M. F. Mingarelli,<sup>3,8,17</sup> S. Osłowski,<sup>8,18</sup> D. Perrodin,<sup>13</sup> A. Possenti,<sup>13</sup> M. B. Purver,<sup>12</sup> S. Sanidas,<sup>11,12</sup> R. Smits,<sup>10</sup> B. Stappers,<sup>12</sup> G. Theureau,<sup>9,14,19</sup> C. Tiburzi,<sup>13,20</sup> R. van Haasteren,<sup>17</sup> A. Vecchio<sup>3</sup> and J. P. W. Verbiest<sup>8,18</sup>

*Affiliations are listed at the end of the paper*

Accepted 2015 September 6. Received 2015 August 31; in original form 2015 June 24

## ABSTRACT

We have searched for continuous gravitational wave (CGW) signals produced by individually resolvable, circular supermassive black hole binaries (SMBHBs) in the latest European Pulsar Timing Array (EPTA) data set, which consists of ultraprecise timing data on 41-ms pulsars. We develop frequentist and Bayesian detection algorithms to search both for monochromatic and frequency-evolving systems. None of the adopted algorithms show evidence for the presence of such a CGW signal, indicating that the data are best described by pulsar and radiometer noise only. Depending on the adopted detection algorithm, the 95 per cent upper limit on the sky-averaged strain amplitude lies in the range  $6 \times 10^{-15} < A < 1.5 \times 10^{-14}$  at  $5 \text{ nHz} < f < 7 \text{ nHz}$ . This limit varies by a factor of five, depending on the assumed source position and the most constraining limit is achieved towards the positions of the most sensitive pulsars in the timing array. The most robust upper limit – obtained via a full Bayesian analysis searching simultaneously over the signal and pulsar noise on the subset of our six best pulsars – is  $A \approx 10^{-14}$ . These limits, the most stringent to date at  $f < 10 \text{ nHz}$ , exclude the presence of sub-centiparsec binaries with chirp mass  $\mathcal{M}_c > 10^9 M_\odot$  out to a distance of about 25 Mpc, and with  $\mathcal{M}_c > 10^{10} M_\odot$  out to a distance of about 1 Gpc ( $z \approx 0.2$ ). We show that state-of-the-art SMBHB population models predict <1 per cent probability of detecting a CGW with the current EPTA data set, consistent with the reported non-detection. We stress, however, that PTA limits on individual CGW have improved by almost an order of magnitude in the last five years. The continuing advances in pulsar timing data acquisition and analysis techniques will allow for strong astrophysical constraints on the population of nearby SMBHBs in the coming years.

**Key words:** black hole physics – gravitational waves – pulsars: general.

## 1 INTRODUCTION

The direct detection of gravitational waves (GWs) is one of the primary goals of contemporary observational astrophysics. The access

to GW information alongside well-established electromagnetic observations will be a milestone in our investigation of the Universe, opening the era of multimessenger astronomy.

Precision timing of an array of millisecond pulsars (MSP; i.e. a pulsar timing array, PTA) provides a unique opportunity to get the very first low-frequency (nHz) GW detection. PTAs exploit the effect of GWs on the propagation of radio signals from ultrastable MSPs to the Earth (e.g. Sazhin 1978; Detweiler 1979),

\* E-mail: [stba@aei.mpg.de](mailto:stba@aei.mpg.de) (SB); [petiteau@apc.in2p3.fr](mailto:petiteau@apc.in2p3.fr) (AP); [asesana@star.sr.bham.ac.uk](mailto:asesana@star.sr.bham.ac.uk) (AS)

producing a characteristic fingerprint in the times of arrival (TOAs) of radio pulses. In the timing analysis, TOAs are fitted to a physical model accounting for all the known processes affecting the generation, propagation and detection of the radio pulses. The timing residuals are the difference between the observed TOAs and the TOAs predicted by the best-fitting model, and they carry information about unaccounted noise and potentially unmodelled physical effects, such as GWs, in the datastream (e.g. Hellings & Downs 1983; Jenet et al. 2005). The European Pulsar Timing Array (EPTA; Kramer & Champion 2013), the Parkes Pulsar Timing Array (PPTA; Hobbs 2013) and the North American Nanohertz Observatory for Gravitational Waves (NANOGrav; McLaughlin 2013), joining together in the International Pulsar Timing Array (IPTA; Hobbs et al. 2010; Manchester & IPTA 2013), are constantly improving their sensitivity in the frequency range of  $\sim 10^{-9}$ – $10^{-6}$  Hz.

The primary GW source in the nHz window is a large population of adiabatically inspiralling supermassive black hole binaries (SMBHBs), formed following the frequent galaxy mergers occurring in the Universe (Begelman, Blandford & Rees 1980). Signals from a cosmic string network (see e.g. Vilenkin 1981; Vilenkin & Shellard 1994) or from other physical processes occurring in the early Universe (see e.g. Grishchuk 2005) are also possible, but we will concentrate on SMBHBs in this paper. Consisting of a superposition of several thousands of sources randomly distributed over the sky (Sesana, Vecchio & Colacino 2008), the signal has classically been described as a stochastic GW background (GWB; Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana et al. 2004). Consequently, in the last decade several detection techniques have been developed in this direction (e.g. Anholm et al. 2009; van Haasteren et al. 2009; Lentati et al. 2013; Chamberlin et al. 2015) and applied to the EPTA, PPTA and NANOGrav data sets to get limits on the amplitude of a putative isotropic GWB (Jenet et al. 2006; Yardley et al. 2011; van Haasteren et al. 2011; Demorest et al. 2013; Shannon et al. 2013; Lentati et al. 2015).

However, Sesana, Vecchio & Volonteri (2009) (see also Ravi et al. 2012) first showed that the signal is dominated by a handful of sources, some of which might be individually resolvable. The typical evolution time-scale of those SMBHBs is thousands to millions of years, far exceeding the observational baseline of PTA experiments (about two decades); therefore, their signals can be modelled as non-evolving continuous GWs (CGWs, Sesana & Vecchio 2010). Resolvable sources are particularly appealing because, if detected and localized on the sky, they can also be followed up electromagnetically, thus providing a multimessenger view (Sesana et al. 2012; Tanaka, Menou & Haiman 2012; Burke-Spolaor 2013; Rosado & Sesana 2014).

This prospect triggered a burst of activity in the development of search and parameter estimation algorithms for CGWs from circular SMBHBs (Sesana & Vecchio 2010; Lee et al. 2011; Babak & Sesana 2012; Ellis, Jenet & McLaughlin 2012; Petiteau et al. 2013; Taylor, Ellis & Gair 2014), and more recently led to the development of the first pipelines for eccentric binaries (Taylor et al. 2015b). The pioneering work of Yardley et al. (2010) was the first to produce sensitivity curves and set upper limits using a power spectral summation method. More recently, Arzoumanian et al. (2014) applied the frequentist and Bayesian methods for evolving and non-evolving signals described in Ellis et al. (2012) to the NANOGrav 5-year data set (Demorest et al. 2013), whereas Zhu et al. (2014) applied a frequentist method to the PPTA data release (DR1) presented in Manchester et al. (2013). Those limits are usually cast in terms of the intrinsic strain amplitude of the wave,  $h_0$  or its inclination-averaged version (which is a factor of 1.26 larger) as a function of

frequency, both averaged over the entire sky or as a function of sky location. The best sky-averaged 95 per cent confidence upper limit on  $h_0$  quoted to date is  $1.7 \times 10^{-14}$  at 10 nHz (Zhu et al. 2014).

Here, we investigate the presence of non-evolving continuous waves from circular binaries in the latest EPTA data release (Desvignes et al., in preparation). We perform a comprehensive study applying both frequentist (Babak & Sesana 2012; Ellis et al. 2012; Petiteau et al. 2013) and Bayesian (Ellis 2013; Taylor et al. 2014; Lassus et al., in preparation) methods, and searching for both evolving and non-evolving GW signals. The paper is organized as follows. In Section 2, we introduce the EPTA data set and the adopted gravitational waveform model. Section 3 is devoted to the description of the techniques developed to analyse the data, divided into frequentist and Bayesian methods. Our main results (upper limit, sensitivity curves, sky maps) are presented in Section 4, and their astrophysical interpretation is discussed in Section 5. Finally, we summarize our study in Section 6. Throughout the paper, we use geometrical units  $G = c = 1$ .

This research is the result of the common effort to directly detect GWs using pulsar timing, known as the EPTA (Kramer & Champion 2013).<sup>1</sup>

## 2 EPTA DATA SET AND GW MODEL

### 2.1 The EPTA data set

In this paper, we make use of the full EPTA data release described in Desvignes et al. (in preparation), which consists of 42 MSPs monitored for timespans ranging from 7 to 24 yr. However, we exclude PSR J1939+2134 from our analysis because it shows a large, unmodelled red-noise component in its timing residuals. The remaining 41 MSPs show well-behaved rms residuals between 130 ns and 35  $\mu$ s. For each of these pulsars, a full timing analysis has been performed using a time-domain Bayesian method based on MULTINEST (Feroz, Hobson & Bridges 2009), which simultaneously includes the white noise modifiers EFAC and EQUAD<sup>2</sup> for each observing system, as well as intrinsic red noise and (observational) frequency dependent dispersion measure (DM) variations. Variations in the DM are due to a changing line of sight through the interstellar medium towards the pulsar. Hereinafter, we refer to this timing analysis as single pulsar analysis (SPA). As a sanity check, parallel analyses have also been done in the frequency domain using the TempoNest plugin (Lentati et al. 2014) for the TEMPO2 pulsar timing package (Hobbs, Edwards & Manchester 2006), and in time domain using search method combining a genetic algorithm Petiteau et al. (2010) with MCMCHammer Foreman-Mackey et al. (2013). The results of the three methodologies are consistent.

For the searches performed in this paper, we use the results of the SPA produced by MULTINEST. Those consist of posterior probability distributions for the relevant noise parameters (EFAC, EQUAD, DM and intrinsic red noise), together with their maximum likelihood (ML) values. Extensive noise analyses on the same data set are fully detailed in Janssen et al. (in preparation) and Caballero et al. (in preparation), and the posterior distributions of the noise parameters of the three most sensitive pulsars in our array (J1909–3744,

<sup>1</sup> [www.epta.eu.org/](http://www.epta.eu.org/)

<sup>2</sup> EFAC is used to account for possible miscalibration of the radiometer noise and it acts as a multiplier for all the TOA error bars. EQUAD represents some additional (unaccounted) source of time independent noise and it is added in quadratures to the TOA error bars.

**Table 1.** Summary of the searches performed in this study. Column 1: name of the search; column 2: treatment of the noise in the search; column 3: number of pulsars considered in the data set ( $N$ ); column 4: dimensionality of the parameter space to search over; column 5: adopted signal model; column 6: notes about the treatment of the likelihood function. The different signal models are: Earth term only (E), Earth plus pulsar term at the same frequency (E+P NoEv), Earth plus pulsar term at different frequencies (E+P Ev).

Search ID	Noise treatment	Pulsars ( $N$ )	Parameters	Signal model	Likelihood
<i>Fp-ML</i>	Fixed ML	41	1	E+P NoEv	Maximization over $2N$ constant amplitudes
<i>Fp</i>	Sampling posterior	41	1	E+P NoEv	Maximization over $2N$ constant amplitudes
<i>Fe</i>	Fixed ML	41	3	E	Maximization over four constant amplitudes
<i>Bayes_E</i>	Fixed ML	41	7	E	Full
<i>Bayes_EP</i>	Fixed ML	6	$7 + 2 \times 6$	E+P Ev	Full
<i>Bayes_EP_NoEv</i>	Fixed ML	41	7	E+P NoEv	Marginalization over pulsar phases
<i>Bayes_EP_NoEv_noise</i>	Searched over	6	$7 + 5 \times 6$	E+P NoEv	Marginalization over pulsar phases

J1713+0747, J1744–1134) are also given in fig. 3 of the companion paper Lentati et al. (2015). In most cases, we fix the noise parameters to their ML value, but we also perform separate searches sampling from the posterior distribution or searching simultaneously over the signal and the noise parameters, as described in Section 3 and summarized in Table 1.

## 2.2 GW model

In this subsection, we introduce the mathematical description of the GW signal from a binary in circular orbit and the associated PTA response. The main purpose here is to introduce the notation; we refer the reader to Sesana & Vecchio (2010) for a full derivation of the relevant equations. The timing residuals of radio pulses due to the propagation of the electromagnetic waves in the field of an intervening GW can be written as

$$r_a(t) = \int_0^t \frac{\delta v_a}{v_a}(t') dt', \quad (1)$$

where

$$\frac{\delta v_a}{v_a} = \frac{1}{2} \frac{\hat{p}_a^i \hat{p}_a^j}{1 + \hat{p}_a \cdot \hat{\Omega}} \Delta h_{ij}. \quad (2)$$

Here  $v_a$  is the frequency of the TOAs (i.e. the spin frequency of the pulsar) and  $\delta v_a$  its deviation. The index  $a$  labels a particular pulsar ( $a = 1, \dots, N$  where  $N = 41$  is the number of the pulsars in our array) and indicates that a given quantity explicitly depends on the parameters of the individual pulsar,  $\hat{p}$  denotes the position of the pulsar on the sky and  $\hat{\Omega}$  is the direction of the GW propagation. The last factor in equation (2) depends on the strain of the GW at the location of the pulsar  $h_{ij}(t_a^p)$  and on Earth  $h_{ij}(t)$ :

$$\Delta h_{ij} = h_{ij}(t_a^p) - h_{ij}(t). \quad (3)$$

The pulsar time  $t_a^p$  is related to the Earth time  $t$  as

$$t_a^p = t - L_a(1 + \hat{\Omega} \cdot \hat{p}_a) \equiv t - \tau_a, \quad (4)$$

where  $L_a$  is the distance to the pulsar. We consider a non-spinning binary system in quasi-circular orbit. To leading order, the response of a particular pulsar to a passing GW (that is, the induced timing residual) is given by

$$r_a(t) = r_a^p(t) - r_a^e(t), \quad (5)$$

where

$$r_a^e(t) = \frac{\mathcal{A}}{\omega} \left\{ (1 + \cos^2 \iota) F_a^+ [\sin(\omega t + \Phi_0) - \sin \Phi_0] \right. \\ \left. + 2 \cos \iota F_a^\times [\cos(\omega t + \Phi_0) - \cos \Phi_0] \right\},$$

$$r_a^p(t) = \frac{\mathcal{A}_a}{\omega_a} \left\{ (1 + \cos^2 \iota) F_a^+ [\sin(\omega_a t + \Phi_a + \Phi_0) \right. \\ \left. - \sin(\Phi_a + \Phi_0)] + 2 \cos \iota F_a^\times [\cos(\omega_a t + \Phi_a + \Phi_0) \right. \\ \left. - \cos(\Phi_a + \Phi_0)] \right\}. \quad (6)$$

Equation (6) contains all the relevant features of the signal, and deserves some detailed explanation. Here,  $\iota$  is the inclination of the SMBHB orbital plane with respect to the line of sight, and  $\mathcal{A}$  (sometimes referred to as  $h_0$  in the PTA literature) is the amplitude of the GW strain given by

$$\mathcal{A} = 2 \frac{\mathcal{M}_c^{5/3}}{D_L} (\pi f)^{2/3}. \quad (7)$$

Throughout the paper we consider a SMBHB with *redshifted* chirp mass  $\mathcal{M}_c = M\eta^{3/5}$ , where  $M = (m_1 + m_2)$  and  $\eta = m_1 m_2 / M^2$  are the total mass and symmetric mass ratio of the binary system;  $D_L$  is the luminosity distance to the source and  $f = \omega / (2\pi)$  is the *observed* GW frequency, which is twice the SMBHB observed orbital frequency.<sup>3</sup> We specify  $\mathcal{A}$  and  $\mathcal{A}_a$  in equation (6) because the GW frequency might be different in the pulsar and Earth terms, implying different amplitudes. In fact, in the quadrupole approximation, the evolution of the binary orbital frequency  $\omega_{\text{orb}} = \omega / 2$  and GW phase can be written as

$$\omega_{\text{orb}}(t) = \omega_{\text{orb}} \left( 1 - \frac{256}{5} \mathcal{M}_c^{5/3} \omega_{\text{orb}}^{8/3} t \right)^{-3/8}, \quad (8)$$

$$\Phi(t) = \Phi_0 + \frac{1}{16\mathcal{M}_c^{5/3}} \left( \omega_{\text{orb}}^{-5/3} - \omega_{\text{orb}}(t)^{-5/3} \right). \quad (9)$$

In equation (6),  $\omega$  and  $\omega_a = \omega(t - \tau_a)$  are the GW frequencies of the Earth term and pulsar term, respectively. Over the typical duration of a PTA experiment (decades), these two frequencies can be approximated as constants (see e.g. Sesana & Vecchio 2010), and we drop the time dependence accordingly. However, the delay  $\tau_a$  between the pulsar and the Earth term is of the order of the pulsar-Earth light travel time, and can be thousands of years. This

<sup>3</sup> The relation between redshifted chirp mass and rest-frame chirp mass is  $\mathcal{M}_c = (1+z)\mathcal{M}_{c,\text{rf}}$ ; likewise, the relation between observed and rest-frame frequency is  $f_{\text{rf}} = (1+z)f$ . The GW community works with redshifted quantities because this is what is directly measured in the observations, and because this way, the  $1+z$  factors cancel out in the equations, making the math cleaner. We note, moreover, that current PTAs might be able to resolve SMBHBs out to  $z \sim 0.1$ , implying only a small difference between rest-frame and redshifted quantities.

is comparable with the evolutionary time-scale of the SMBHB’s orbital frequency (Sesana & Vecchio 2010). In particular  $\omega_a = 2\omega_{\text{orb},a}$  is obtained by setting  $t = -\tau_a$  in the right-hand side of equation (8), and can be generally different from one pulsar to the other and from  $\omega$ . Combining equations (8) and (9), one can show that  $\Phi_a \approx -\omega\tau_a$ .

If  $T$  is the timing baseline of a given PTA’s set of observations, then its nominal Fourier frequency resolution bin is given by  $\Delta f \approx 1/T$ . We therefore have two possibilities.

(i) If  $(\omega_a - \omega)/(2\pi) > \Delta f$  for the majority of the MSPs in the array, then the pulsar and the Earth terms fall in different frequency bins; all the Earth terms can be added-up coherently and the pulsar terms can be considered either as separate components of signal or as an extra incoherent source of noise.

(ii) Conversely, if  $(\omega_a - \omega)/(2\pi) < \Delta f$  for the majority of the MSPs, then the pulsar terms add up to the respective Earth terms, destroying their phase coherency.

This distinction has an impact on the detection methodology that should be adopted, as we will see in the next section.

The antenna response functions  $F_a^{+, \times}$  of each pulsar to the GW signal depend on the mutual pulsar-source position in the sky and are given by

$$F_a^+ = \frac{1}{2} \frac{(\hat{p}^a \cdot \hat{u})^2 - (\hat{p}^a \cdot \hat{v})^2}{1 + \hat{p}^a \cdot \hat{\Omega}}, \quad (10)$$

$$F_a^\times = \frac{(\hat{p}^a \cdot \hat{u})(\hat{p}^a \cdot \hat{v})}{1 + \hat{p}^a \cdot \hat{\Omega}}, \quad (11)$$

where

$$\hat{\Omega} = -\{\sin \theta_S \cos \phi_S, \sin \theta_S \sin \phi_S, \cos \theta_S\}, \quad (12)$$

$$\hat{u} = \hat{n} \cos \psi - \hat{m} \sin \psi, \quad \hat{v} = \hat{m} \cos \psi + \hat{n} \sin \psi, \quad (13)$$

$$\hat{n} = \{\cos \theta_S \cos \phi_S, \cos \theta_S \sin \phi_S, -\sin \theta_S\}, \quad (14)$$

$$\hat{m} = \{\sin \phi_S, -\cos \phi_S, 0\}. \quad (15)$$

Here  $\theta_S, \phi_S$  define the source sky location (respectively latitude and longitude), and  $\psi$  is the GW polarization angle.

### 2.3 Likelihood function and noise model

The core aspect of all searches performed in this study (both frequentist and Bayesian) is the evaluation of the likelihood that some signal described by equation (6) is present in the time series of the pulsar TOAs. We use the expression for the likelihood marginalized over the timing parameters as described in detail in van Haasteren & Levin (2013):

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda} | \delta \mathbf{t}) &= \frac{1}{\sqrt{(2\pi)^{n-m} \det(\mathbf{G}^T \mathbf{C} \mathbf{G})}} \\ &\times \exp\left(-\frac{1}{2}(\delta \mathbf{t} - \mathbf{r})^T \mathbf{G}(\mathbf{G}^T \mathbf{C} \mathbf{G})^{-1} \mathbf{G}^T (\delta \mathbf{t} - \mathbf{r})\right). \end{aligned} \quad (16)$$

Here,  $n$  is the length of the vector  $\delta \mathbf{t} = \cup x_a$  obtained by concatenating the individual pulsar TOA series  $x_a$ ,  $m$  is the total number

of parameters describing the timing model, and the matrix  $\mathbf{G}$  is related to the design matrix (see van Haasteren & Levin 2013, for details). The variance-covariance matrix  $\mathbf{C}$ , in its more general version, contains contributions from the GWB and from the white and (in general) red noise:  $\mathbf{C} = \mathbf{C}_{\text{gw}} + \mathbf{C}_{\text{wn}} + \mathbf{C}_{\text{rn}}$ . We refer the reader to van Haasteren et al. (2009) and to our companion paper Lentati et al. (2015) for exact expressions of the noise variance-covariance matrix. In our analysis, we use both the time (van Haasteren & Levin 2013) and frequency domain (Lentati et al. 2013) representation of this matrix. Both approaches give qualitatively and quantitatively consistent results (as we checked during our analysis). Therefore, we will not specify which representation was used for each of the employed methods. Moreover, we have excluded  $\mathbf{C}_{\text{gw}}$  from our analysis assuming the hypothesis that the data consists of noise and a single CGW source.

The model parameters in equation (16) are split in two groups: (i) parameters describing the CGW signal ( $\boldsymbol{\lambda} : \mathbf{r} = \mathbf{r}(\boldsymbol{\lambda})$ ), and (ii) parameters describing the noise model  $\boldsymbol{\theta}$ . The waveform of a non-spinning circular binary given by equation (6) is generally described by  $7 + 2N$  parameters. The Earth term (a single sinusoidal GW) is fully described by seven parameters:  $(\mathcal{A}, \theta_S, \phi_S, \Psi, \iota, \omega, \Phi_0)$ , and each pulsar term adds two additional parameters: frequency and phase  $(\omega_a, \Phi_a)$ . As mentioned above, the pulsar term might fall at the same frequency as the Earth term, in which case we have only one extra parameter per pulsar (since  $\omega_a = \omega$ ). In principle, even for  $\omega_a \neq \omega$ ,  $\omega_a$  and  $\Phi_a$  are uniquely connected through the pulsar distance  $L_a$  as shown by equations (8) and (9). However this implicitly assumes that we have an exact model for the evolution of the binary, which in this case is perfectly circular, non-spinning and GW driven. Even tiny deviations from these assumptions (i.e. small residual eccentricity, partial coupling with the environment, spins), very likely to occur in nature, will invalidate the  $\omega_a - \Phi_a$  connection, and in the most general case, the two parameters must be considered separately. We will specify the details of the adopted waveform model for each individual method in the next section.

Some of the noise parameters (like pulsar-intrinsic red noise or clock and ephemeris errors) are correlated with the GW parameters and one should in principle fit for noise and GW parameters simultaneously. However, such a multidimensional search is computationally very expensive, and in most of the searches detailed below, we use noise characteristics derived from the SPA introduced in Section 2.1 and extensively described in Janssen et al. (in preparation) and Caballero et al. (in preparation). We characterize the noise by considering the data from each pulsar separately (as given by the SPA), and to exclude any potential bias we also considered a model “noise + monochromatic signal”. The results of the latter are usually consistent with the ‘noise-only’ model, except for one pulsar, J1713+0747, where we have found some correlation between the parameters describing the red noise and the extra monochromatic signal. Each SPA returns posterior distributions for the array of noise parameters  $\boldsymbol{\theta}_a$ : slope and amplitude of the red noise, slope and amplitude of DM variations (both red noise and DM variation were modelled as single power laws) and EFAC and EQUAD for each backend. We have used this information in two ways: (i) use the ML estimator for all the parameters ( $\boldsymbol{\theta}_{\text{ML}} = \cup \boldsymbol{\theta}_{\text{ML},a}$ ) and assume that the noise is represented by that model, (ii) draw the parameters describing the noise from the posterior distributions obtained in the SPA. The first choice (fixed noise at  $\boldsymbol{\theta}_{\text{ML}}$ ) is computationally cheaper as we need to compute the noise variance matrix,  $\mathbf{C}(\boldsymbol{\theta})$ , only once, while in the second choice we need to recompute it for each draw of the noise parameters. Arzoumanian et al. (2014) found that fixing the noise to the ML values will bias the results of the

search and will yield a better upper limit compared to the full search including noise parameters. The effect is, however, not dramatic, as they found their upper limit is less than a factor  $\approx 1.5$  worse in the latter case, over the full frequency range. To test the robustness of our results, we also ran a full analysis searching simultaneously on the GW and noise parameters on a restricted data set of six pulsars (see Section 3.3 for the definition of this restricted data set).

### 3 METHODS

As described in the previous section, the two parts of the GW signal (Earth term and pulsar term) might or might not fall at the same frequency, which has implications for the form of the likelihood function given in equation (16). On top of that, both frequentist and Bayesian methods can be used to analyse the data. We therefore identify four separate classes of analysis: frequentist non-evolving, frequentist evolving, Bayesian non-evolving and Bayesian evolving. In each individual case, the signal (and sometimes also the noise) can be treated differently, and the likelihood might undergo peculiar manipulations (maximization, marginalization, etc.); moreover, we sometimes explore two variations of the same method. The result is a variety of complementary searches, which we now describe in detail and which are comparatively summarized in Table 1.

#### 3.1 Frequentist methods

The basis for the frequentist analysis is to postulate that we have a deterministic signal in the data which is either corrupted (in case of detection) or dominated by noise (no detection). We then define appropriate statistical distributions (or simply ‘statistics’) based on the likelihood function, those describe the data in absence and presence of a signal. Those statistics must be chosen in such a way that the detection rate is maximized for a fixed value of the false alarm probability (FAP), which is also known as the Neyman–Pearson criterion. The aim is to check the null hypothesis (i.e. whether the data are described by noise only), and in case of ‘no detection’ set an upper limit on the GW amplitude. In building our statistics we always assume that the noise is Gaussian and, unless otherwise stated, characterized by ML parameters estimated during the SPA. We then fix the FAP at 1 per cent to set the detection threshold. In case of no detection, frequency dependent upper limits are obtained by splitting the frequency range in small bins, performing a large number of signal injections with varying amplitude in each bin, and computing the associated detection statistics. In the next two subsections, we describe two particular implementations of this procedure, known as  $\mathcal{F}_p$  and  $\mathcal{F}_e$  statistics.

##### 3.1.1 $\mathcal{F}_p$ -statistic

In the case of non-evolving sources (i.e.  $\omega_a = \omega$ ), Ellis et al. (2012) showed that, for each pulsar  $a$ , one can write equation (6) factorizing out the  $\omega$  dependence

$$r_a(t) = \sum_{j=1}^2 b_{(j,a)}(\mathcal{A}, \theta_S, \phi_S, \Psi, \iota, \Phi_0, \Phi_a) \kappa_{(j)}(\omega, t). \quad (17)$$

Explicit expressions for  $b_{(j,a)}$  and  $\kappa_{(j)}$  can be found in Ellis et al. (2012). We merely stress here that  $\kappa_{(j)}$  are independent on the considered pulsar  $a$ . The log of expression (16) can then be maximized over the constant  $2N$  coefficients  $b_{(1,a)}$ ,  $b_{(2,a)}$  assuming that they are independent, resulting in what is commonly known as the  $\mathcal{F}_p$ -statistic. This latter assumption is not actually true if the number of

pulsars is larger than six. We make use of the full 41 pulsar data set in the  $\mathcal{F}_p$ -statistic evaluation. However, six pulsars dominate (i.e. give 90 per cent of) the S/N, as discussed in Section 3.3; moreover, we can simply postulate this form of statistic. Assuming independence in the maximization process will somewhat degrade the detection power of the  $\mathcal{F}_p$ -statistic; nonetheless, this is a very simple detection statistic which depends only on one parameter: the frequency of the GW. The  $\mathcal{F}_p$ -statistic is in essence an ‘excess-power’ statistic, in which we basically search for extra power – compared to the expectations of the statistic describing the null hypothesis – at a given frequency in each pulsar’s data (summing the weighted square of the Fourier amplitudes in each pulsar). By mixing the GW phases  $\Phi_0$  and  $\Phi_a$  at the Earth and at the pulsar in the coefficients  $b_{(j,a)}$ , by construction  $\mathcal{F}_p$  assumes that there is no coherence between GW signals across each individual pulsar’s data.

With the assumption of Gaussian noise,  $2\mathcal{F}_p$  follows a central  $\chi^2$  distribution,  $p_0(\mathcal{F}_p)$  – non-central,  $p_1(\mathcal{F}_p, \rho)$ , if a signal with optimal S/N  $= \rho$  is present – with  $n = 2N$  degrees of freedom (which follows from  $2N$  maximizations of the log likelihood), given by

$$p_0(\mathcal{F}_p) = \frac{\mathcal{F}_p^{n/2-1}}{(n/2-1)!} \exp(-\mathcal{F}_p), \quad (18)$$

$$p_1(\mathcal{F}_p, \rho) = \frac{(2\mathcal{F}_p)^{(n/2-1)/2}}{\rho^{n/2-1}} I_{n/2-1}(\rho \sqrt{2\mathcal{F}_p}) e^{-\mathcal{F}_p - \frac{1}{2}\rho^2}, \quad (19)$$

where  $I_{n/2-1}(x)$  is the modified Bessel function of the first kind of order  $n/2 - 1$ . We divide our data set in bins  $\Delta f = 1/T$ , where  $T$  is the longest pulsar observation time, and we evaluate  $\mathcal{F}_p$  independently in each bin. At any given  $f$ , we consider only pulsars with observation time  $T > 1/f$  when evaluating  $\mathcal{F}_p$ . Since the implementation of the  $\mathcal{F}_p$ -statistic is computationally cheap, we run two flavors of it: (i) one in which we fix the noise to the ML values ( $\mathcal{F}_p$ -ML hereinafter), (ii) one in which we take into account the uncertainty in the noise parameters by sampling from their posterior distribution derived from the SPA (simply  $\mathcal{F}_p$  hereinafter).

##### 3.1.2 $\mathcal{F}_e$ -statistic

If the source is evolving, then  $\omega_a \neq \omega$  and one can choose to consider as ‘signal’ only the Earth portion of equation (6), and treat the pulsar term as a source of noise. In this case, we take only the Earth term of  $\mathbf{r}$  in the likelihood expression (16). Babak & Sesana (2012) showed that it is convenient to rewrite the antenna pattern expressions (11) separating the terms containing the polarization angle  $\psi$ :

$$F_a^+ = F_c^a \cos(2\psi) + F_s^a \sin(2\psi) \quad (20)$$

$$F_a^\times = -F_s^a \cos(2\psi) + F_c^a \sin(2\psi), \quad (21)$$

thus rearranging equation (6) in the form

$$r_a(t) = \sum_{j=1}^4 a_{(j)}(\mathcal{A}, \iota, \psi, \Phi_0) h_{(j)}^a(\omega, \theta_S, \phi_S, t). \quad (22)$$

Explicit expressions for  $a_{(j)}$ ,  $h_{(j)}^a$ ,  $F_c^a$ ,  $F_s^a$  can be found in Babak & Sesana (2012), but note that, contrary to the  $\mathcal{F}_p$  case, the coefficients  $a_{(j)}$  do not depend on the considered pulsar  $a$ . As done for the  $\mathcal{F}_p$  case, we can now maximize over the four constants  $a_{(j)}$ , constructing a statistic – the  $\mathcal{F}_e$ -statistic – which is a function of three parameters; namely the GW frequency,  $f$ , and source sky location,  $\theta_S, \phi_S$ . If the noise is Gaussian,  $\mathcal{F}_e$  follows a  $\chi^2$  distribution,

$p_0(\mathcal{F}_e)$  – non-central,  $p_1(\mathcal{F}_e, \rho)$ , if a signal with optimal S/N= $\rho$  is present – with four degrees of freedom (which follows from the four maximizations of the log likelihood), given by:

$$p_0(\mathcal{F}_e) = \mathcal{F}_e e^{-\mathcal{F}_e}, \quad (23)$$

$$p_1(\mathcal{F}_e, \rho) = \frac{(2\mathcal{F}_e)^{1/2}}{\rho} I_1(\rho\sqrt{2\mathcal{F}_e}) e^{-\mathcal{F}_e - \frac{1}{2}\rho^2}, \quad (24)$$

Where  $I_1(x)$  is the modified Bessel function of the first kind of order 1.

Note that the  $\mathcal{F}_e$ -statistic adds the data from the pulsars coherently (taking the phase information into account), and being a function of the source position, it allows direct sky localization. However, if the signal is non-evolving, its efficiency drops significantly and the estimation of the source sky location can be severely biased. The evaluation of  $\mathcal{F}_e$  is significantly more computationally expensive than  $\mathcal{F}_p$ , as we need to take into account cross-pulsar correlations. We therefore continue to use the full 41-pulsar data set, but we fix the noise parameters at the ML values found in the SPA, without attempting any sampling (contrary to the  $\mathcal{F}_p$  case).

For searching over the 3-D parameter space ( $f, \theta, \phi$ ), we use the multimodal genetic algorithm described in Petiteau et al. (2013). A detailed description of the method can be found there; here we give only a brief overview. In the first step, we run a genetic algorithm (with 64 organisms) over 1000 generations tuned for an efficient exploration of the whole parameter space. Then we identify the best spot in the 3-D space and seed there a variation of the Markov chain Monte Carlo ‘MCMC Hammer’ (Foreman-Mackey et al. 2013), which serves as a sampler and returns the effective size of the ‘mode’ and the correlations among parameters. Here ‘mode’ stands for a local maximum of the likelihood. We then again run the genetic algorithm, exploring the remaining parameter space (i.e. excluding the mode we already found) and searching for other potential local maxima in the likelihood. At each distinct maximum that is found, we run ‘MCMC hammer’, and we iterate this procedure five times. The end result is a set of independent local maxima in likelihood, among which we choose the largest (highest in value). We repeat the entire procedure three times to verify convergence. This algorithm has proven to give very robust results, even in the case of pathological likelihood surfaces with multiple maxima.

### 3.2 Bayesian analysis

In the Bayesian approach, the parameters describing the model are treated as random variables. The principles of Bayesian statistics provide a robust framework to obtain probability distributions of those parameters for a given set of observations, while also folding in our prior knowledge of them.

Bayes’ theorem states that the *posterior* probability density function (PDF),  $p(\boldsymbol{\mu}|\mathbf{d}, \mathcal{H})$ , of the parameters  $\boldsymbol{\mu}$  within a hypothesised model  $\mathcal{H}$  and given data  $\mathbf{d}$  is

$$p(\boldsymbol{\mu}|\mathbf{d}, \mathcal{H}) = \frac{\mathcal{L}(\boldsymbol{\mu}, \mathcal{H}|\mathbf{d})\pi(\boldsymbol{\mu}|\mathcal{H})}{p(\mathbf{d}|\mathcal{H})}, \quad (25)$$

where  $\mathcal{L}(\boldsymbol{\mu}, \mathcal{H}|\mathbf{d})$  is the likelihood of the parameters given the data assuming the model  $\mathcal{H}$  with parameters  $\boldsymbol{\mu}$ . The prior PDF of the model parameters,  $\pi(\boldsymbol{\mu}|\mathcal{H})$ , incorporates our preconceptions and prior experience of the parameter space. The Bayesian evidence,

$p(\mathbf{d}|\mathcal{H}) \equiv \mathcal{Z}$ , is the probability of the observed data given the model  $\mathcal{H}$ ,

$$\mathcal{Z} = \int \mathcal{L}(\boldsymbol{\mu})\pi(\boldsymbol{\mu})d^N\boldsymbol{\mu}. \quad (26)$$

For posterior inference within a model,  $\mathcal{Z}$  plays the role of a normalization constant and can be ignored. However, if we want to perform model selection then the evidence value becomes a key component in the computation of the posterior odds ratio:

$$\frac{p(\mathcal{H}_2|\mathbf{d})}{p(\mathcal{H}_1|\mathbf{d})} = \frac{p(\mathbf{d}|\mathcal{H}_2)p(\mathcal{H}_2)}{p(\mathbf{d}|\mathcal{H}_1)p(\mathcal{H}_1)} = \frac{\mathcal{Z}_2 \times p(\mathcal{H}_2)}{\mathcal{Z}_1 \times p(\mathcal{H}_1)}. \quad (27)$$

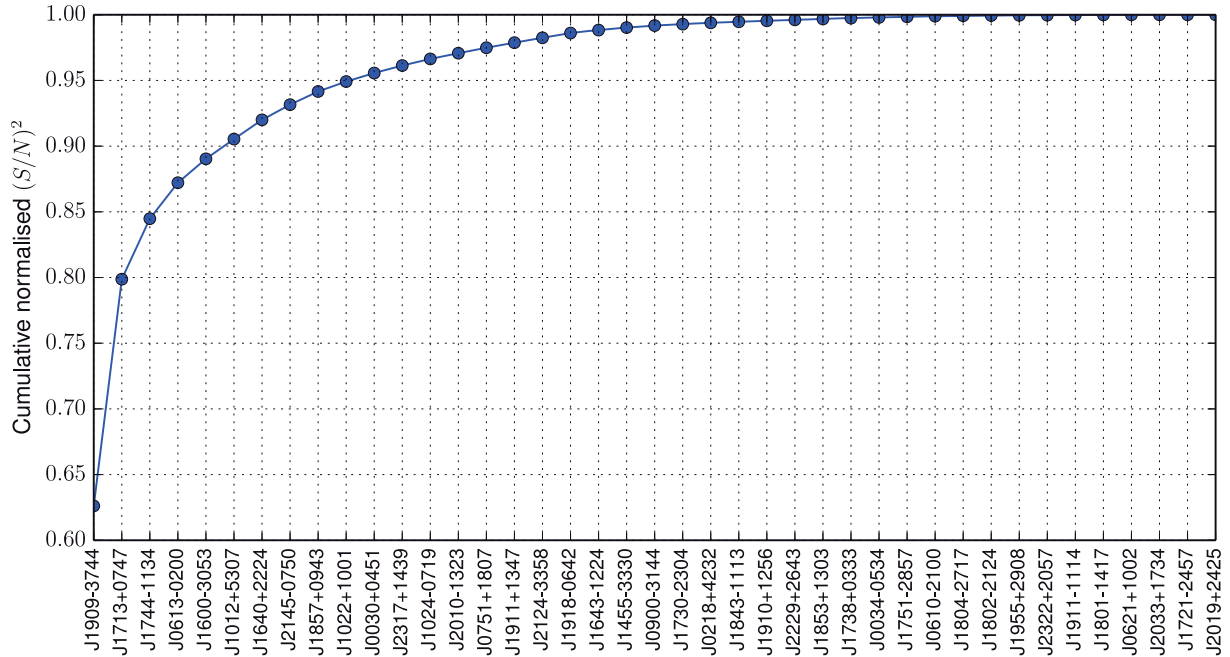
Here  $\mathcal{H}_1, \mathcal{H}_2$  are the two models under comparison,  $\mathcal{Z}_2/\mathcal{Z}_1$  is the Bayes factor and  $p(\mathcal{H}_2)/p(\mathcal{H}_1)$  is the prior probability ratio for the two competing models, which is often set to be one; the posterior odds ratio is then just the Bayes factor.

When we specialize the formalism above to PTA data, the likelihood  $\mathcal{L}$  is given by equation (16), the data  $\mathbf{d}$  is the concatenation of the TOA series  $\delta t$  and the model’s parameters  $\boldsymbol{\mu}$  are identified with  $\boldsymbol{\theta}, \boldsymbol{\lambda}$ . The presence of a signal in the data is assessed through the odds ratio, equation (27), where model  $\mathcal{H}_1$  corresponds to data described by noise only and model  $\mathcal{H}_2$  corresponds to data described by noise plus a CGW. We use non-informative and conservative priors in our analysis: they are always uniform in all parameters. It is especially important to emphasize that we have used a broad uniform prior in the signal’s amplitude  $\mathcal{A}$ , which will provide robust and conservative upper limits on the strain. Unless otherwise stated, we fix the stochastic noise parameters to the ML values found in the SPA, and we employ either MULTINEST (Feroz et al. 2009) or a parallel tempering adaptive MCMC (Lassus et al., in preparation) to return samples from the posterior distributions, and thus reconstruct the posterior PDFs. Both techniques also permit a recovery of the aforementioned model-comparison statistic known as the *Bayesian evidence*. We describe the different searches in detail below.

#### 3.2.1 Phase-marginalized Bayesian analysis

For the non-evolving CGW signal searches, one should sample a  $7 + N$  multidimensional posterior, corresponding to the parameters  $(\mathcal{A}, \theta_S, \phi_S, \Psi, \iota, \omega, \Phi_0, \Phi_a)$ . For a 41-pulsar data set, that amounts to sampling a 48-dimensional parameter space. We mitigate the computational cost implied by such high dimensionality by numerically marginalizing (integrating) our posterior distribution over all of these nuisance parameters  $\Phi_a$ , thereby collapsing the search to more manageable dimensions (i.e. to seven parameters only, Taylor et al. 2014). By doing so, we not only rapidly recover the posterior PDFs, but also achieve quick and accurate Bayesian evidence values. This method is close in spirit to the  $\mathcal{F}_p$ -statistic, we call it *phase-marginalized Bayesian analysis* (labelled as *Bayes-EP-NoEv*, for Bayes, Earth+pulsar, non-evolving source). The sampling of the posterior is performed by MULTINEST.

The high performance of the MULTINEST sampler allowed us to also run a full search, including noise parameters  $\boldsymbol{\theta}$ , on a restricted data set composed of the six best pulsars –contributing to 90 per cent of the total S/N, see Section 3.3 and Fig. 1 – in our array (labelled *Bayes-EP-NoEv-noise*). We use non-informative priors also for the pulsar noise parameters: uniform in the range [1,7] for the slopes of both red noise and DM; uniform in the range [−20, −10] for the logarithm (base 10) of their amplitudes; uniform in the range [0,10] for the global EFAC (see Lentati et al. 2015, for a definition of global EFAC). The posterior spans now a  $7 + 5N = 37$  dimension parameter space. We designed this search as a benchmark for



**Figure 1.** Cumulative normalized  $(S/N)^2$ . We rank pulsars according to their relative contribution to the total  $S/N$ , and we plot the quantity  $\sum_{i < a} \rho_a / \rho$ .

our different fixed-noise analysis; to speed up the sampling, we restricted the frequency prior range to  $[5, 15]$  nHz, which turns out to be the sweet spot of our array sensitivity.

### 3.2.2 Full Bayesian analysis

We also employed the Bayesian formalism to construct a search for evolving signals. In this case, we can either (i) use the full signal of equation (6), or (ii) restrict ourselves to the Earth term only. This latter case (ii) is similar in spirit to the frequentist  $\mathcal{F}_e$ -statistic, with the difference that we now search over the full 7-D source parameter space with the whole 41-pulsar array; we label this analysis *Bayes-E*. The full-signal analysis (i) is very computationally expensive because we do not assume any model for the source evolution, therefore adding two extra parameters ( $\omega_a$ ,  $\Phi_a$ ) for each pulsar. Note that our parametrization of the full signal does not rely on any knowledge about the distance to the pulsar. We name this search *Bayes-EP-Ev*. We use non-informative priors also on those parameters, imposing the only constraint that the pulsar-term frequencies cannot be higher than the Earth-term one ( $\omega_a \leq \omega$ ). Since we need to cover a  $7 + 2N$  parameter space, we limit this search to the best 6 pulsars, for a total of 19 parameters. In both searches, the multidimensional posteriors are stochastically sampled with a parallel tempering adaptive MCMC (Lassus et al., in preparation).

### 3.3 Restricted data set: ranking pulsar residuals

As mentioned above, some of the searches are extremely computationally expensive, involving sampling of highly dimensional parameter spaces. A way to mitigate the computational cost is to run the algorithms on a “restricted data set”, which includes only the best pulsars for these purposes. We therefore need a metric to rank the quality of each pulsar. We choose our metric to be the relative  $S/N$  contribution of each pulsar to a putative detectable source. We conduct extensive Monte Carlo simulations, in which we in-

ject CGW signals with an astrophysically motivated distribution of parameters  $\lambda$  in the EPTA data set. For each signal, we compute the total  $S/N$  and the relative contribution of each individual pulsar according to the standard inner product definition:

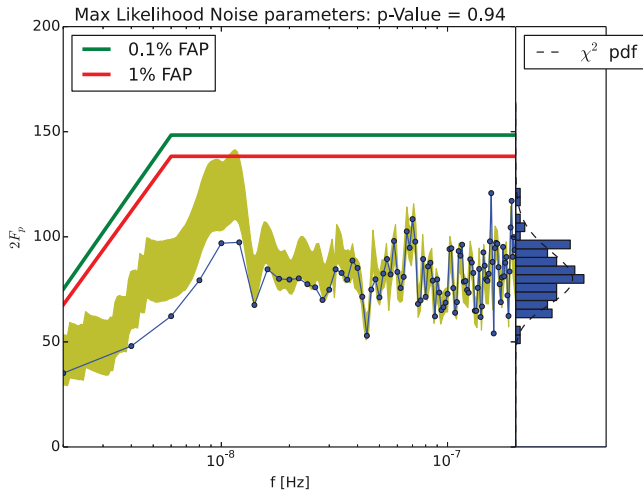
$$(S/N)^2 = (h(\lambda)|h(\lambda)) \equiv (h(\lambda)^T \mathbf{G})(\mathbf{G}^T \mathbf{C}(\theta) \mathbf{G})^{-1} (\mathbf{G}^T h(\lambda)). \quad (28)$$

For each injected CGW, we draw the noise parameters of each pulsar ( $\theta_a$ ) from the corresponding posterior distribution of the SPA. We injected 1000 detectable CGWs, and each signal contained both Earth and pulsar terms. In Fig. 1 we plot the average (over the 1000 simulations performed) build-up of  $\rho = (S/N)^2$  as we add pulsars to the array, from the best to the worst. The plot shows the cumulative  $S/N$  square over the total  $S/N$  square, i.e.  $\sum_{i < a} \rho_a / \rho$ . The ranking reveals that the array is heavily dominated by PSR J1909–3744, contributing more than 60 per cent of  $\rho$ , followed by PSR J1713+0747 at about 20 per cent. We decided to form a data set considering only the best pulsars building up 90 per cent of  $\rho$ . We ended up with six pulsars (PSRs J1909–3744, J1713+0747, J1744–1134, J0613–0200, J1600–3053 and J1012+5307) which constitute the restricted data set mentioned above, and is the same used in our companion isotropic and anisotropic GW background searches (Lentati et al. 2015; Taylor et al. 2015a).

## 4 RESULTS

We now turn to the description of the main results of our analysis. As in the previous section, we present the outcome for the frequentist and Bayesian analyses separately, discussing first the detection results and then the upper limit computation for the various techniques. No evidence for CGW signals was found in the data; a summary of all the 95 per cent upper limits on the GW strain amplitude  $\mathcal{A}$  as a function of frequency is collected in Fig. 6.





**Figure 2.**  $\mathcal{F}_p$ -statistics. The blue line represents  $\mathcal{F}_p$  evaluated at 99 independent frequencies, for 41 pulsars, using the ML noise parameters. The right-hand panel shows the histogram of the  $\mathcal{F}_p$  values at all frequencies, and the dashed line is a central  $\chi^2$ -distribution, which is the expected distribution of the  $\mathcal{F}_p$  statistic in absence of a GW signal. The two are consistent with a p-value of 0.93, which is indicative of excellent agreement. The yellow area denotes the central 90 per cent of  $\mathcal{F}_p$  evaluated across the whole sample of noise parameters. The thresholds turn over below 6 nHz because of the reduced number of pulsars that have enough observing timespan.

## 4.1 Frequentist analysis

### 4.1.1 $\mathcal{F}_p$ -statistic

To tackle the issue of detection, we have evaluated  $\mathcal{F}_p$  at  $N_f = 99$  independent frequencies in the range  $[1/T, 2 \times 10^{-7} \text{ Hz}]$  in steps of  $\Delta f = 1/T$ , where  $T$  is the maximum observation time. In this data set  $1/T = 2.0 \times 10^{-9} \text{ Hz}$ . In the absence of a signal,  $2\mathcal{F}_p$  follows the  $\chi^2$  PDF with  $n = 2N$  degrees of freedom given by equation (18). The FAP associated with a given threshold  $F_0$  is simply given by the integral of the PDF and takes the form:

$$P(\mathcal{F}_p > F_0) = \int_{F_0}^{\infty} p_0(\mathcal{F}_p) d\mathcal{F}_p = \exp(-F_0) \sum_{k=0}^{n/2-1} \frac{F_0^k}{k!}. \quad (29)$$

To assess the global probability of finding a given value of  $\mathcal{F}_p$  in our data set, we need to take into account that we are evaluating it at 99 different frequency bins, i.e. we are performing 99 independent trials. The global FAP is therefore given by

$$\text{FAP} = 1 - [1 - P(\mathcal{F}_p > F_0)]^{N_f}. \quad (30)$$

We choose thresholds in  $F_0$  that correspond to FAPs of 0.01 and 0.001. The results are presented in Fig. 2. For the ML noise parameters (solid line with circles and corresponding histogram on the right-hand), the data is consistent with the noise description with a p-value of 0.93 and that there is no excess at any frequency above the 0.01 FAP threshold. However, the choice of the noise parameters, and hence the covariance matrix  $C$ , is crucial in evaluating  $\mathcal{F}_p$ , and the SPA reveals that many parameters are poorly constrained. We therefore chose to create a whole distribution of  $\mathcal{F}_p$  at each frequency, sampling over noise parameters from the posterior distributions produced by the SPA. This is overplotted as the yellow area in Fig. 2. The results obtained in this way are independent of the particular ML value and give a better view of the uncertainties involved. At each frequency,  $f$ , only pulsars with baselines  $T > 1/f$  have been taken into account. This explains the rising FAP thresh-

olds. The uncertainty in  $\mathcal{F}_p$  induced by the uncertainty in the noise parameters is much larger at lower frequencies, where red noise and DM dominate the noise model. For the lowest frequencies, the ML evaluation of  $\mathcal{F}_p$  does not even lie within the 90 per cent region shown in the plot, which is a consequence of the fact that we sum the contributions of 41 broad, and often skewed, distributions. In several pulsars the ML red noise or DM amplitudes tend to lie at the upper end of the distribution, which leads to small values of  $\mathcal{F}_p$  and results in this offset.

After confirming that the SPA describes the data appropriately as only noise, we want to know how large a CGW contribution must be in order to make the  $\mathcal{F}_p$ -distribution non-central and clearly distinguishable from the noise. This yields an upper limit on the GW strain  $\mathcal{A}$  that might be present in our data and still consistent with the observed  $\mathcal{F}_p$  value. A standard way to do this in frequentist analysis is through signal injections. We first fix the noise to the SPA-ML value; in this case the procedure to obtain the upper limit on  $\mathcal{A}$  at each frequency  $f$  is as follows (see e.g. Ellis et al. 2012):

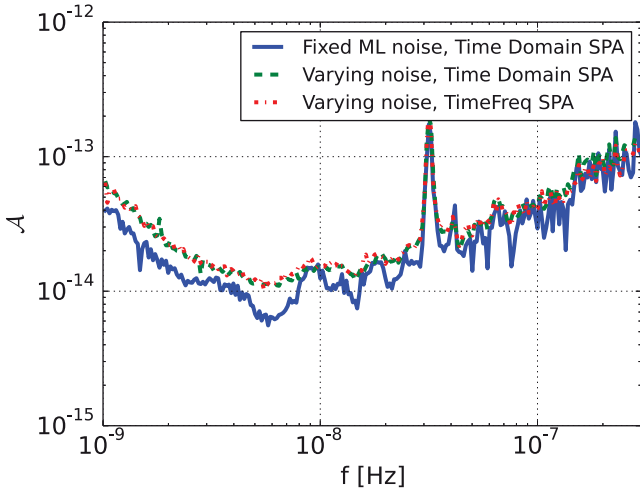
- (i) compute  $\mathcal{F}_{p,0}$  using the data set;
- (ii) create 1000 different mock data sets  $i$ , by injecting in each of them one source with fixed strain  $\mathcal{A}$  but otherwise random parameters, and compute  $\mathcal{F}_{p,i}$ ;
- (iii) compute the fraction  $y$  of mock data sets where  $\mathcal{F}_{p,i} > \mathcal{F}_{p,0}$ ;
- (iv) repeat steps (ii) and (iii) with different  $\mathcal{A}$  until  $y = 0.95$ .

In practice, we iterate over a grid in  $\log_{10} \mathcal{A}$  with step size 0.1 and interpolate to find the point for which  $y = 0.95$ . We also want to obtain an upper limit that takes into account the uncertainty of the single-pulsar noise parameters. In doing this, the procedure outlined above is modified as follows:

- (i) compute a distribution of 1000  $\mathcal{F}_{p,0}$  at each frequency  $f$  using the data set and 1000 noise parameters drawn from the posterior noise PDF obtained from the SPA;
- (ii) create 1000 different mock data sets  $i$ , by injecting in each of them one source with fixed strain  $\mathcal{A}$  but otherwise random parameters in each, and compute  $\mathcal{F}_{p,i}$ , each time drawing different noise parameters from the posterior noise PDF;
- (iii) compute the fraction  $y$  of mock data sets where  $\mathcal{F}_{p,i} > \bar{\mathcal{F}}_{p,0}$ , where the single reference  $\bar{\mathcal{F}}_{p,0}$  is chosen to be the mean of the distribution obtained in step (i);
- (iv) repeat steps (ii) and (iii) with different  $\mathcal{A}$  until  $y = 0.95$ .

In other words, we want 95 per cent of the  $\mathcal{F}_{p,i}$  distribution to lie above the mean value  $\bar{\mathcal{F}}_{p,0}$  of the observed distribution. The motivation behind criterion (iii) is that it resembles the criterion for the ML upper limit and it is much more conservative than a Kolmogorov–Smirnov test. It is also possible to choose even more conservative criteria, susceptible to possible non-Gaussian tails in either distribution, but we do not explore this possibility here.

Upper limits as a function of  $f$  are shown in Fig. 3. In all cases, the most stringent limit is around 6–7 nHz, and reaches down to  $\mathcal{A} = 6 \times 10^{-15}$  for the  $\mathcal{F}_p$ -ML case. We note, however, that when we allow the noise to vary, we get a limit which is a factor  $\approx 2$  worse at low frequency, yielding a value of  $\mathcal{A} = 1.1 \times 10^{-14}$ . This is consistent with Fig. 2; at low frequency the ML estimator of the noise parameters is not representative of the noise posterior distribution, resulting in  $\mathcal{F}_p$  values which are biased towards low values. Injections with lower  $\mathcal{A}$  will therefore result in a detectable excess  $\mathcal{F}_p$ , pushing down the upper limit. Note that our upper limit gets significantly worse at  $f < 3 \times 10^{-9} \text{ Hz}$ , because red noise becomes significant for some pulsars, and not all 41 pulsars in our array contribute down to those frequencies, having an observation



**Figure 3.** 95 per cent upper limit on the GW strain  $\mathcal{A}$  obtained with the  $\mathcal{F}_p$ -statistic. The blue line corresponds to noise parameters fixed to the ML values obtained in the time-domain SPA, while the dashed lines take into account the full uncertainty in the noise estimation by sampling from the PDF distributions of either the pure time domain (green) or the time-frequency (red) SPA. At 1 per year (the peak), the limit is poor because the GW signal at this frequency is absorbed in the fitting of the pulsar positions.

time  $T < 10$  years. At high frequency, the 95 per cent upper limit degrades approximately linearly with  $f$ , consistent with a white-noise-dominated data set.

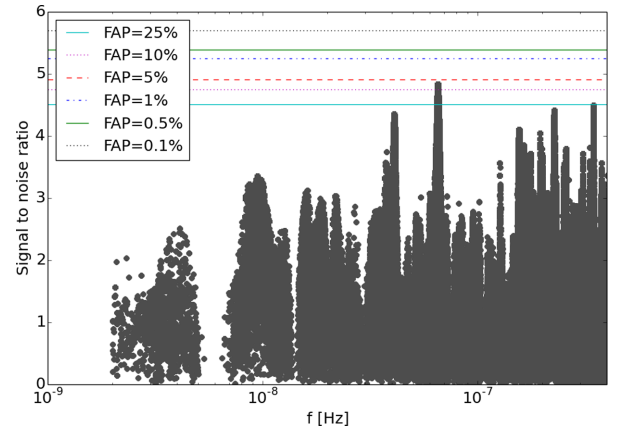
#### 4.1.2 $\mathcal{F}_e$ -statistic

Since  $\mathcal{F}_e$  is also a frequentist technique, the procedures to assess detection and to place upper limits are analogous to the  $\mathcal{F}_p$  case. Here, in absence of signal,  $2\mathcal{F}_e$  follows the  $\chi^2$  PDF with  $n = 4$  degrees of freedom given by equation (23). The FAP associated with a given threshold  $F_0$  is simply given by the integral of the PDF and takes the form:

$$P(\mathcal{F}_e > F_0) = \int_{F_0}^{\infty} p_0(\mathcal{F}_e) d\mathcal{F}_e = (1 + F_0)e^{-F_0}. \quad (31)$$

Again, the global false alarm rate depends on the number of trials, according to equation (30), which is now given by the number of independent templates in the sky location-frequency 3-D parameter space. The vast majority of templates we have used in the search are strongly correlated. We estimate the number of independent trials by constructing a stochastic 3-D template bank (see Babak 2008; Harry, Allen & Sathyaprakash 2009). We use a minimal match equal to 0.5 as the criterion of independence among different templates, and obtain 4276 independent points in the searched parameter space (full sky and frequency band restricted to 2–400 nHz). Fig. 4 presents the result of the detection analysis. The maximum of  $\mathcal{F}_e$ ,  $\mathcal{F}_{e,\max}$ , is found at  $f = 66$  nHz and  $\{\theta_S, \phi_S\} = \{51.9, 136.4\}$ . It corresponds to a FAP of 7 per cent, which is compatible with a non-detection. Note that a GW signal with  $S/N=5$ , if present in the data, would correspond to a FAP of  $\approx 5$  per cent, which is clearly too high to make any confident claim of detection, as shown in Fig. 4.

Since the data are compatible with describing only noise, we can again compute the 95 per cent upper limit on the strain amplitude  $\mathcal{A}$  of a putative CGW as a function of  $f$  by means of signal injections. The procedure is similar to the one employed in the  $\mathcal{F}_p$ -ML analysis. Here, we construct a grid of frequencies, and at each grid point we



**Figure 4.** Result of the  $\mathcal{F}_e$ -statistic detection analysis. The points are the trials of the search. The horizontal lines are the detection thresholds for different FAPs.

consider a small frequency interval  $\Delta f = 1$  nHz which is sufficient to capture the injected Earth term:

- (i) compute  $\mathcal{F}_{e,\max,0}$ , i.e. the maximum of  $\mathcal{F}_e$  on the whole sky and in the narrow frequency band  $\Delta f$  in the raw data set;
- (ii) create 1000 different mock data sets  $i$ , by injecting in each of them one source with fixed strain  $\mathcal{A}$  but otherwise random parameters, including both Earth and pulsar terms;
- (iii) for each data set  $i$ , run a search (stochastic bank + multisearch genetic algorithm<sup>4</sup>) to find  $\mathcal{F}_{e,\max,i}$ , i.e. the maximum of  $\mathcal{F}_e$  on the whole sky and in the narrow frequency band  $\Delta f$ ;
- (iv) compute the fraction  $y$  of the mock data sets in which  $\mathcal{F}_{e,\max,i} > \mathcal{F}_{e,\max,0}$ ;
- (v) repeat steps (ii) and (iii) increasing  $\mathcal{A}$  until  $y = 0.95$ .

The 95 per cent sky-averaged upper limit obtained in this way is shown by the red curve in Fig. 6 and is in agreement with the results obtained with other methods.

## 4.2 Bayesian analysis

The frequentist analysis presented above already provided strong evidence against the presence of a signal in the data. None the less, this can also be addressed in the Bayesian framework through the computation of the odds ratio defined by equation (27). Since we give a priori no preference to the presence or absence of a signal in the data, we set the prior probability ratio to unity, and the odds ratio coincides with the Bayes factor. The Bayes factor is then simply the ratio of the Bayesian evidence computed for the hypothesis  $\mathcal{H}_1$  and hypothesis  $\mathcal{H}_0$  as given by equation (27), which in our case reduces to:

$$\mathcal{B} = \frac{\int \mathcal{L}(\theta, \lambda | \delta t) \pi(\theta, \lambda) d\theta d\lambda}{\int \mathcal{L}(\theta | \delta t) \pi(\theta) d\theta}. \quad (32)$$

<sup>4</sup> A full use of the multimodal genetic algorithm for each injection is computationally expensive and not needed, in practice. We therefore use a lighter and faster search for the injected signals. We construct a stochastic bank with minimal match 0.95 and we filter the data through this bank. We then identify the maximum of  $\mathcal{F}_e$  across the bank and refine our search running the genetic algorithm with 64 organisms evolved over 1000 generations. The stochastic bank is generated only once for the full parameter space and contains 532 488 templates. In each search we use only the portion of template bank covering the parameter space region around the injected signal.

In the case of fixed noise, we assume that the noise parameters  $\theta$  are known exactly (fixed at their ML value), and the Bayes factor is directly computed from the likelihood ratio multiplied by the priors, integrated over the source parameters  $\lambda$ . We compute the evidence using both MULTINEST and parallel tempering MCMC searches. In all the Bayesian searches with fixed noise we obtain Bayes factors close to zero, consistent with a non-detection and with the outcome of the frequentist analysis. In particular, we get  $\log(\mathcal{B}) = -0.27$  for the search *Bayes\_E*, and  $\log(\mathcal{B}) = -0.31$  for the search *Bayes\_EP\_NoEv*.

The Bayesian analysis also returns samples from the joint posterior probability distribution of all model parameters. The marginalized distribution of any parameter of interest can then be evaluated by integrating (i.e. marginalizing) the joint posterior distribution over all other parameters. We are particularly interested in the strain amplitude  $\mathcal{A}$ . We can then split the vector parameter  $\lambda = (\mathcal{A}, \lambda')$  and integrate over  $\lambda'$  to obtain the marginalized posterior for the parameter  $\mathcal{A}$ . In practice, we divide the frequency range in small bins in which we carry out this marginalization procedure separately. The 95 per cent upper limit at each frequency corresponds to the value  $\tilde{\mathcal{A}}$  for which 95 per cent of the posterior distribution lies at  $\mathcal{A} < \tilde{\mathcal{A}}$ ; namely

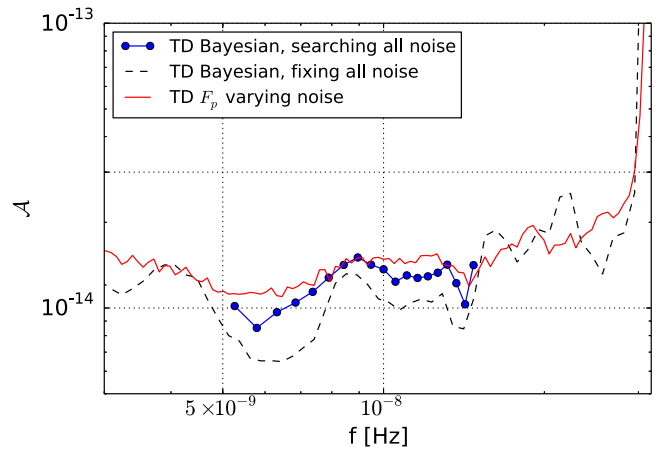
$$0.95 = \int_0^{\tilde{\mathcal{A}}} d\mathcal{A} \int d\lambda' \mathcal{L}(\mathcal{A}, \lambda' | \delta t) \pi(\mathcal{A}) \pi(\lambda'). \quad (33)$$

Results are shown in Fig. 6, which compares all the upper limits on the GW strain achieved by all methods presented in this paper. For the non-evolving source case, the *Bayes\_EP\_NoEv* upper limit agrees particularly well with the fixed-noise  $\mathcal{F}_p$ -ML statistic. This is encouraging, since the two methods are similar in spirit as they adopt the same signal model and assume fixed/known noise parameters. For the evolving source case, the  $\mathcal{F}_e$  upper limit is very similar to both  $\mathcal{F}_p$ -ML and *Bayes\_EP\_NoEv*, mimicking almost perfectly their behaviour at low frequency. The upper limits obtained by both the *Bayes\_E* and the *Bayes\_EP* searches are noisier and slightly higher, but overall consistent with the others within a factor of two.

As mentioned in Section 3, we also ran a full 37-dimensional search over noise and signal parameters on the restricted set of the six best pulsars in our PTA (c.f. Fig. 1), and in a restricted frequency range of 5–15 nHz where we have the best sensitivity. We used the phase-marginalised Bayesian analysis for non-evolving sources, and labelled the run *Bayes\_EP\_NoEv\_noise*. The 95 per cent upper limit obtained in this case is shown in Fig. 5, together with the fixed noise *Bayes\_EP\_NoEv* and the noise-sampling  $\mathcal{F}_p$  results. The *Bayes\_EP\_NoEv\_noise* limit lies a factor 1.1–1.5 above the *Bayes\_EP\_NoEv* one. This is in line with the findings of Arzoumanian et al. (2014), and confirms that our ML fixed noise upper limits are reliable within a factor  $\lesssim 1.5$ . It is also interesting to see that the *Bayes\_EP\_NoEv\_noise* limit agrees fairly well with the noise-sampling  $\mathcal{F}_p$  one. By analysing Fig. 6 we can conclude that all the upper limits yielded by the different techniques agree within a factor of two. We also observe that methods based on fixed noise (ML) parameters slightly underestimate the upper limit, which could be because the ML values are not always representative of the posterior distribution of the noise parameters.

### 4.3 Sky maps

Most of the searches outlined above are also sensitive to the source location on the sky. We can therefore extend our study and produce



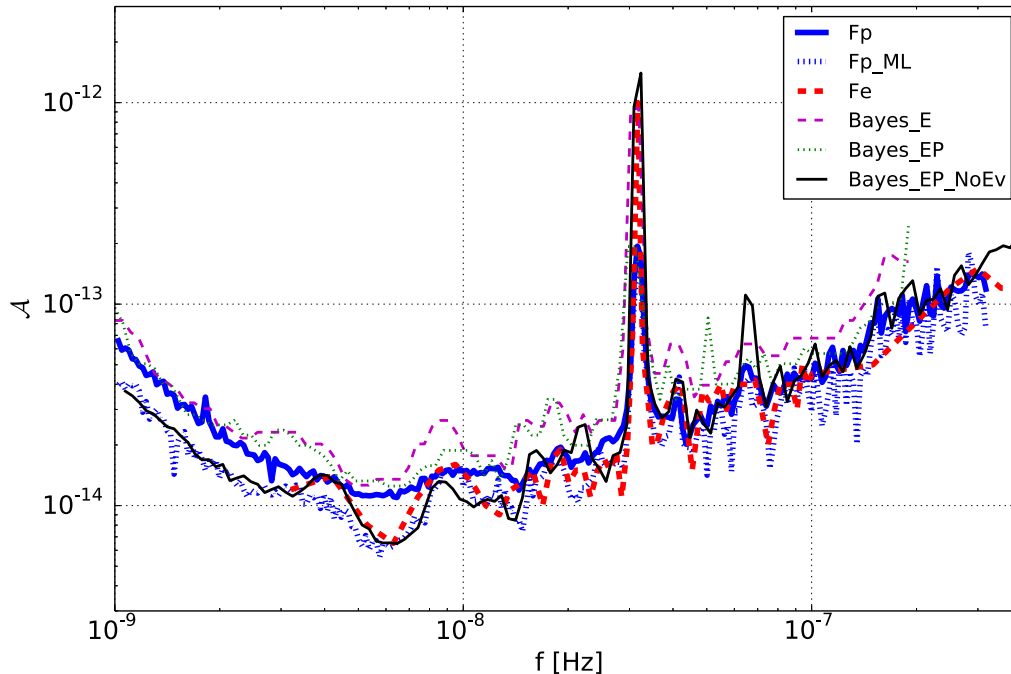
**Figure 5.** The 95 per cent upper limit on the GW strain obtained with the phase-marginalized Bayesian analysis, searching on noise and signal simultaneously (*Bayes\_EP\_NoEv\_noise*, blue-circled curve), is compared to the same analysis with ML-fixed noise (*Bayes\_EP\_NoEv*, dashed-black curve) and to the noise-sampling  $\mathcal{F}_p$  analysis (solid-red curve). Details of each specific analysis are given in Table 1. ‘TD’ stands for time-domain analysis as opposed to the time-frequency approach developed in Lentati et al. (2013).

sky maps of the 95 per cent upper limits provided by our analysis as a function of frequency.

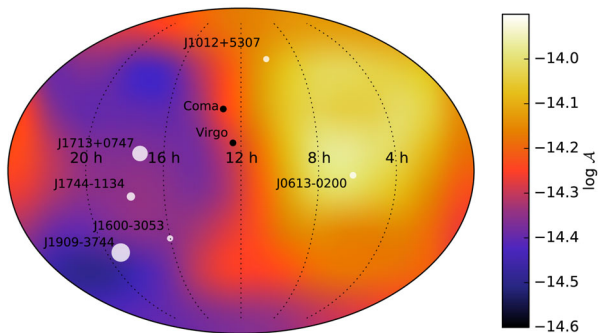
In the frequentist framework, this is straightforward to do in the context of the  $\mathcal{F}_e$ -statistic, since it is sensitive to sky position (unlike  $\mathcal{F}_p$ ). As already mentioned, the upper limit is evaluated through massive signal injections according to the procedure outlined in Section 4.1.2. The difference is that now we have to divide the sky into ‘cells’ and inject 500 sources at each cell location. This is much more computationally expensive than the evaluation of the sky-averaged limit; we therefore generate the sky map at 6.3 nHz only, corresponding to our best sky-averaged limit. This is shown in Fig. 7. As expected we are more sensitive in the region of the best pulsars.

We can also produce targeted upper limits as a function of sky location by means of Bayesian techniques. The problem here is that by splitting the posterior samples on a 3-D frequency-sky location grid, we end up with only a handful of points per cell, which are not enough to derive a reliable 95 per cent upper limit. To mitigate this issue, we divided the sky into  $4 \times 2$  patches on the  $\phi$  and  $\theta$  coordinates, respectively. A dedicated Bayesian analysis (fixed noise with marginalization of pulsar phases) on each patch yielded enough samples to sub-divide the region into a further  $4 \times 4$  sectors, for a total of  $16 \times 8 = 128$  resolution elements across the whole sky. Fig. 8 illustrates the sky map obtained in this way at a GW frequency of 7 nHz (a movie showing the evolution of the sky map across the relevant frequency range is available at <http://www.epta.eu.org/aom.html>).

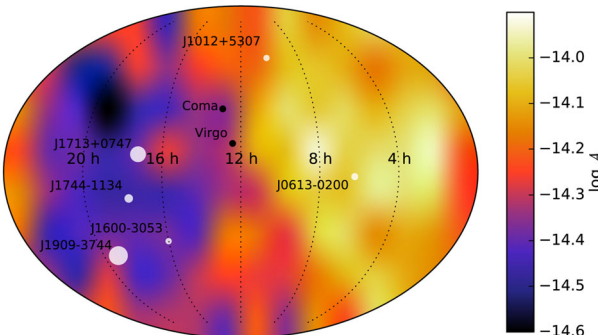
The qualitative agreement between the two maps is quite good. In both of them, the best pulsars are shown as white dots, with size proportional to their contribution to the square of the S/N. As expected, the most constraining (i.e. lowest) upper limits on the strain of a putative CGW lie around the location of the best pulsars in the array, and the sky maps shows a clear dipolar pattern. The closest galaxy clusters in the Universe, i.e. Virgo and Coma, are located at the transition between the two regions of the dipolar pattern, in an area of ‘average sensitivity’.



**Figure 6.** The 95 per cent upper limit on the GW strain for the three frequentist methods, i.e.  $\mathcal{F}_p$  varying noise ( $F_p$ ),  $\mathcal{F}_p$  fixed noise ( $F_{p\_ML}$ ) and  $\mathcal{F}_e$ , and the three bayesian methods, i.e. “evolving source” with Earth term only ( $Bayes\_E$ ) and with Earth and Pulsar terms ( $Bayes\_EP$ ) and ‘non-evolving source with Earth and Pulsar terms ( $Bayes\_EP\_NoEv$ ), see Table 1 for details.



**Figure 7.** Sensitivity sky map at  $f = 6.3$  nHz computed with the  $\mathcal{F}_e$  computed with 500 injections in 48 directions in the sky (‘cells’). The colour scale corresponds to  $\log_{10}$  of the 95 per cent upper limit on the strain amplitude  $\mathcal{A}$ . The white points indicate the positions of the six best pulsars with sizes corresponding to their contribution to the S/N. Black dots indicate the location of the Virgo and the Coma clusters.



**Figure 8.** Sensitivity sky map at  $f = 7$  nHz computed with the phase-marginalized Bayesian technique for a ‘non-evolving source’. The colour scale and points are the same as for Fig. 7.

## 5 ASTROPHYSICAL INTERPRETATION

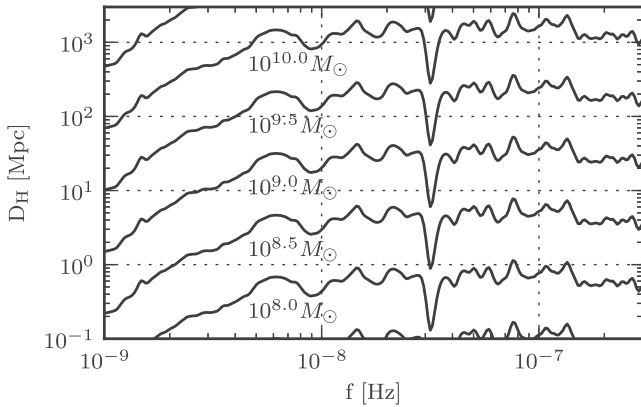
The upper limits on CGWs from SMBHBs presented in this paper are currently the most stringent in the literature. We turn now to investigate their impact on the astrophysics of SMBHBs.

### 5.1 Horizon distance

Each of the 95 per cent upper limits on  $\mathcal{A}$  derived in the previous section can be easily converted into a horizon distance for CGW detection as a function of mass and frequency using equation (7). If  $\mathcal{A}_{95 \text{ per cent}}(f)$  is the strain upper limit as a function of frequency obtained with a specific method, then

$$D_H(f, \mathcal{M}_c) = 2 \frac{\mathcal{M}_c^{5/3}}{\mathcal{A}_{95 \text{ per cent}}(f)} (\pi f)^{2/3}. \quad (34)$$

In a frequentist sense, this has to be interpreted as the distance at which, on average, a source of mass  $\mathcal{M}_c$  emitting at frequency  $f$  located anywhere on the sky would result in a value of the detection statistics *higher than what we measure in the data* with 95 per cent probability, if it was there. As an example, results for the  $\mathcal{F}_p$ -ML statistic are presented in Fig. 9. An interesting feature of the plot is that, for a given  $\mathcal{M}_c$ ,  $D_H$  is essentially constant (slowly declining) for  $f > 5 \times 10^{-9}$  Hz. This is because of the cancellation effect between the rising CGW amplitude with frequency,  $\mathcal{A} \propto f^{-2/3}$  and the PTA sensitivity, which degrades almost linearly with  $f$  (see Figs 3 and 6). In this frequency range, and with the current sensitivity, we can exclude the presence of an SMBHB with  $\mathcal{M}_c > 10^9 M_\odot$  out to a distance of about 25 Mpc, i.e. well beyond the distance to the Virgo cluster, and with  $\mathcal{M}_c > 3 \times 10^9 M_\odot$  out to a distance of about 200 Mpc, i.e. twice the distance to the Coma cluster. Note that Virgo and Coma themselves are located in a region of ‘average sensitivity’ in our sky sensitivity map (see Fig. 8), meaning that we can rule out the presence of SMBHBs (with the characteristics



**Figure 9.** Horizon distance as a function of GW frequency for selected values of  $\mathcal{M}_c$ , based on the  $\mathcal{F}_p$ -ML upper limit.

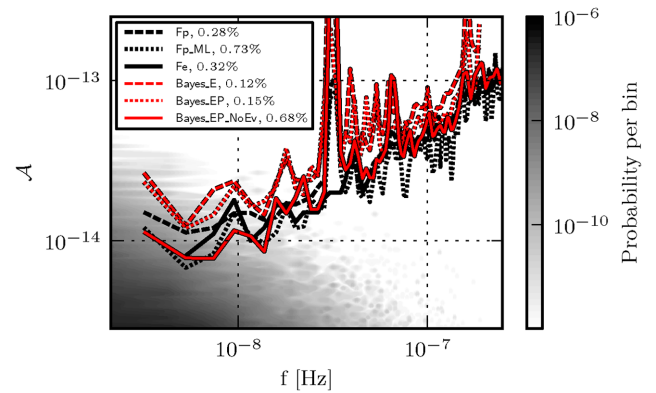
described above) in these specific clusters. We remind the reader that these numbers are for SMBHBs with a given *redshifted* chirp mass; because of the  $1+z$  factor, horizon distances evaluated at the same values of the *intrinsic* SMBHB mass will be slightly larger.

A number of potentially interesting sources have been proposed in the literature. Among these are the two blazars OJ287 (Valtonen et al. 2008) and PG 1302–102 (Graham et al. 2015). Both objects are located at  $z \approx 0.3$  corresponding to  $D_L \approx 1.5\text{Gpc}$ . At such distances, we cannot rule out any system below  $\mathcal{M}_c = 10^{10} M_\odot$ , which far exceeds the plausible range of chirp masses inferred for these two objects. Should their binary nature be confirmed, these two systems would likely be among the thousands of contributors to the stochastic GW background. At this stage, SMBHBs need to be more massive and/or nearby to be resolved by a PTA.

## 5.2 Probability of detection

A natural question that arises at this point is: could we expect a detection of a CGW signal with the current EPTA sensitivity? We now evaluate the probability of detecting CGWs from an individual SMBHB with an array like the current EPTA, by using a large set of observationally based simulations of the cosmic SMBHB population. Each simulation represents a particular realization of the ensemble of SMBHBs. In a nutshell, galaxy merger rates are obtained using a selection of galaxy mass functions and close galaxy pair fractions from the literature; merging galaxies are populated with SMBHBs following empirical black hole–galaxy host relations; finally, each binary is assumed to emit GWs while inspiralling in a quasi-circular orbit. Given the broad range of different models taken into account and their uncertainties, numerous simulations are created in order to cover all possible configurations consistent with the observations. The SMBHB populations obtained in this way are consistent with the results of semi-analytic halo merger trees and cosmological  $N$ -body and hydrodynamical simulations. More details on the simulations and the models employed to produce them can be found in Sesana (2013).

One can perform signal injections drawing the sources from these models and run all the different detection pipelines detailed above, to assess detection probabilities. However, this is an expensive task, and we do not need such a refined analysis at this stage. We instead simplify the problem following a similar approach as in Rosado, Sesana & Gair (2015). For a given realization of the SMBHB population, we group GW sources in frequency bins  $\Delta f = 1/T$ , and



**Figure 10.** GW strain amplitude versus GW observed frequency. The coloured lines represent the different upper limits presented in this work. The shading gives the probability of detecting an SMBHB in a particular interval of strain and frequency. That detection probability increases towards lower frequencies and smaller values of strain (on the lower-left corner). In the legend, the percentage of detection probability is given for each of the upper limits.

compute the characteristic strain

$$h_c^2 = \frac{\sum_k h_k^2 f_k}{\Delta f}, \quad (35)$$

the sum runs over all binaries falling in the frequency bin. We then identify the loudest source in each bin, and compute its S/N following Sesana & Vecchio (2010) assuming that the noise is given by the sum of the strains of the GWs produced by all other binaries. In practice, we are assuming that all other sources produce an ‘unresolved background’, and we check whether the loudest source ‘sticks out’ of it. We assume a detection statistic described by a  $\chi^2$  distribution with four degrees of freedom, and we consider ‘individually resolvable’ only those sources with S/N surpassing the FAP threshold of 0.1 per cent related to this distribution.<sup>5</sup>

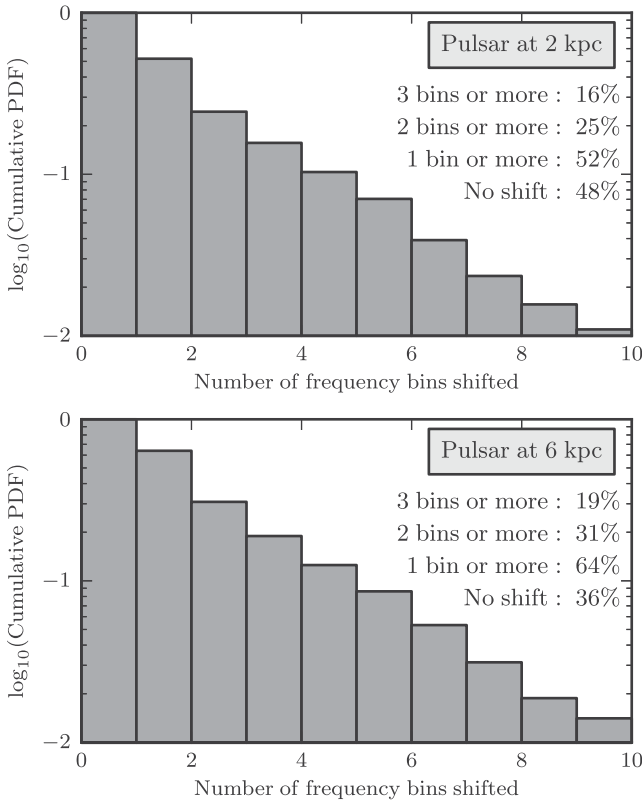
Let us assume a particular upper limit on the GW strain amplitude, among those presented in Fig. 6, and call it  $UL^i$ . At a particular frequency bin  $f_j$ , we simply estimate the probability of detecting an SMBHB with such sensitivity as the fraction of realisations in which a resolvable binary produces a strain amplitude  $\mathcal{A} > \mathcal{A}_{95} \text{ per cent}(f_j)$ . We call this detection probability  $p(\text{D}|UL^i, f_j)$ . The probability of *not* detecting a binary at that frequency is thus  $p(\text{N}|UL^i, f_j) = 1 - p(\text{D}|UL^i, f_j)$ . Assuming that the probabilities of different frequency bins are independent, the probability of not detecting a binary in any frequency bin is the product of the individual values  $p(\text{N}|UL^i, f_j)$ . Hence,

$$p(\text{D}|UL^i) = 1 - \prod_j (1 - p(\text{D}|UL^i, f_j)) \quad (36)$$

is the probability of detecting an SMBHB at any frequency bin, for the upper limit  $UL^i$ .

The detection probability at any frequency obtained for each of the upper limits is given in the legend of Fig. 10. The maximum detection probability achieved with the EPTA upper limits is below  $\sim 1$  per cent. Therefore, we can safely conclude that a non-detection is consistent with the theoretical expectations.

<sup>5</sup> By assuming a  $\chi^2$  distribution with four degrees of freedom, we are assuming  $\mathcal{F}_e$  as detection statistic. This is an arbitrary choice dictated by computational convenience only. Results are, however, qualitatively unchanged if a different statistic (e.g.  $F_p$ ) is assumed.



**Figure 11.** Cumulative PDF of the frequency shift between pulsar and Earth terms, in units of the adopted frequency bin  $\Delta f = 15 \text{ yr}^{-1}$ . The upper and the lower panels assume a pulsar distance of 2 kpc (average distance of the EPTA pulsars) and 6 kpc (maximum distance), respectively.

### 5.3 Frequencies of the Earth and the pulsar terms

In our searches, we distinguished between evolving and non-evolving GW signals, presenting distinct search methods for each of them. One may therefore ask, whether one type of signal is more likely than the other, in order to better focus development efforts on specific analysis pipelines. We can use the same simulated SMBHB populations discussed above to answer this question. As shown by Fig. 10, only a small percentage of them leads to a detection with the sensitivity of the current EPTA, being therefore inconsistent with observations. Nevertheless, it is still meaningful to study the outcome of those realizations, since they would resemble the true ensemble of SMBHBs in the fortunate case of a detection in the near future.

For each of the observable SMBHBs in those realizations, we calculate, according to equation (8), the frequency evolution of the emitted GW after a time lapse of  $4\text{kpc}/c$ , which would be the maximum time difference between pulsar and Earth terms for a pulsar located at 2kpc (which is approximately the mean distance to the EPTA pulsars). Frequency shifts are defined as the difference between the GW frequency before and after that time lapse in units of the frequency resolution bin of the array (assumed to be  $15 \text{ yr}^{-1}$  here). Their distribution is shown on the upper plot of Fig. 11. The lower plot is analogous, but assuming a pulsar located at 6 kpc (which corresponds to approximately the largest distance to a pulsar in the EPTA). Should EPTA detect an individual source in the near future, it could either be evolving or non-evolving with nearly equal probability. Even considering the shifts produced in the furthest pulsar only ( $\leq 6 \text{ kpc}$ ), there is still a 36 per cent probability that

the source would be non-evolving. Decreasing the PTA sensitivity floor would make it sensitive to lower mass binaries (which evolve faster), but would also improve the chance of detection at higher frequency, where evolution is more likely. Likewise, extending the observation time will allow us to see binaries at lower frequency, where evolution is less likely; it will, however, also shrink the size of the frequency bin ( $\Delta f = 1/T$ ), making it easier for a source to sweep through different resolution elements. Detection strategy development for both classes of sources is therefore warranted.<sup>6</sup>

## 6 CONCLUSIONS

In this paper, we searched for continuous GW signals in the latest release of the EPTA data set. We adopted both frequentist and Bayesian techniques, searching for both frequency-evolving and strictly monochromatic signals. In most of the cases, we fixed the value of the noise parameters in each pulsar to the ML estimated in a separate SPA. This choice was primarily dictated by computational feasibility, but is certainly non-optimal, since pulsar noise and GW signals might be degenerate and one should include both simultaneously in the search. To validate our results we therefore also performed a frequentist analysis, sampling from the posterior distributions of the noise parameters returned by the SPA (simply labeled  $\mathcal{F}_e$ ), and a full Bayesian search over both the noise and the signal (labeled *Bayes\_EP\_NoEv\_noise*). Because of the high dimensionality of the search parameter space, the latter has been conducted on a restricted data set including only the six best pulsars in the array.

None of the analysis yielded any evidence of the presence of a signal, and only upper limits on the amplitude  $\mathcal{A}$  of a putative CGW could be placed. The excellent quality and length of the data set allowed us to place limits comparable to those of Zhu et al. (2014) at  $f > 10 \text{ nHz}$  and a factor of two better at  $f < 10 \text{ nHz}$ , yielding the overall most stringent constraints to date. All the employed methods yield 95 per cent upper limits on  $\mathcal{A}$  ( $\mathcal{A}_{95}$  per cent) consistent within a factor of two across the whole 2–400nHz frequency range. Our best sensitivity is in the 5–7nHz interval, where we find  $6 \times 10^{-15} < \mathcal{A}_{95} \text{ per cent} < 1.5 \times 10^{-14}$ , depending on the adopted method. The most robust analysis (*Bayes\_EP\_NoEv\_noise*) results in  $\mathcal{A}_{95} \text{ per cent} = 9 \times 10^{-15}$  at 6 nHz. Limits on the strain amplitude can be converted to horizon distances as a function of source mass and frequency. We exclude the existence of SMBHBs with separation  $< 0.01 \text{ pc}$  and  $\mathcal{M}_c > 10^9 M_\odot$  out to a distance of about 25 Mpc (well beyond Virgo), and with  $\mathcal{M}_c > 10^{9.5} M_\odot$  out to a distance of about 200 Mpc (twice the distance to Coma). In recent years, several ‘overmassive’ black holes have been found in the local Universe, with measured masses in excess of  $10^{10} M_\odot$ . Our analysis excludes that any such system lives in a compact binary within a distance of about 1 Gpc ( $z \approx 0.2$ ). Finally, we compared our limits to the predictions of state of the art models of the cosmic population of SMBHBs. We found a detection probability of  $\lesssim 1$  per cent at current sensitivity, consistent with the null result of our searches.

The present analysis has also highlighted a few interesting technical issues related to the search methods and to the nature of the data set. Despite not being robust for detection purposes, as pointed out by Arzoumanian et al. (2014), fixed noise analysis upper limits are consistent within 50 per cent of those obtained by searches over

<sup>6</sup> A more detailed study of the expected properties of the first detectable SMBHBs can be found in Rosado et al. (2015).

the full parameter space (i.e. including signal and noise simultaneously). Therefore, so long as the data do not support the presence of a signal, a computationally cheap analysis of this type can be carried out over an extensive data set of numerous pulsars, possibly yielding more interesting astrophysical constraints on the low-redshift SMBHB population in the near future. Eventually, simultaneous searches over the signal and noise parameters will be required for a confident detection claim. However, those are extremely expensive, and novel techniques capable of efficiently handling parameter spaces of 100+ dimensions must be developed. The reason why the results of the full search on the restricted data set of six pulsars is consistent with those provided by fixed noise analysis on the full set of 41 pulsars, is that the current EPTA array is heavily dominated by a handful of ultrastable MSPs. In particular, PSRs J1909–3744 and J1713+0747 combined account for 80 per cent of the EPTA sensitivity to CGWs. As a result, the EPTA data set sensitivity has a strongly dipolar pattern across the sky, varying by almost a factor of four over the celestial sphere. The discovery of new ultrastable MSPs will therefore be crucial to provide a better sky coverage, ensuring that no ‘blind spots’ are left, and thus enhancing the probability of detecting CGWs in the coming decade.

## ACKNOWLEDGEMENTS

Part of this work is based on observations with the 100-m telescope of the Max-Planck-Institut für Radioastronomie (MPIfR) at Effelsberg. The Nançay radio Observatory is operated by the Paris Observatory, associated with the French Centre National de la Recherche Scientifique (CNRS). We acknowledge financial support from ‘Programme National de Cosmologie and Galaxies’ (PNCG) of CNRS/INSU, France. Pulsar research at the Jodrell Bank Centre for Astrophysics and the observations using the Lovell Telescope is supported by a consolidated grant from the STFC in the UK. The Westerbork Synthesis Radio Telescope is operated by the Netherlands Institute for Radio Astronomy (ASTRON) with support from The Netherlands Foundation for Scientific Research NWO.

This research was performed using several supercomputers: the CCIN2P3 computer cluster of the CNRS-IN2P3 (Lyon-France), the ARAGO computer cluster of the François Arago Centre (Paris-France), the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (<http://www.hpc.cam.ac.uk/>), provided by Dell Inc using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council, and the Vulcan cluster of MPIfG-AEI (Golm-Germany). The authors acknowledge the support of VirtualData from LABEX P2IO for providing computing resources through its StratusLab cloud. This work was supported in part by the National Science Foundation under grant no. PHYS-1066293 and the hospitality of the Aspen Center for Physics.

LL was supported by a Junior Research Fellowship at Trinity Hall College, Cambridge University. ST was supported by appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, administered by Oak Ridge Associated Universities through a contract with NASA. CMFM was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme. AS and JG are supported by the Royal Society. SAS acknowledges funding from an NWO Vidi fellowship (PI JWTH). RNC acknowledges the support of the International Max Planck Research School Bonn/Cologne and the Bonn-Cologne Graduate School. KJL is supported by the National Natural Science Foundation of China (grant no.11373011). RvH is supported by

NASA Einstein Fellowship grant PF3-140116. JWTH acknowledges funding from an NWO Vidi fellowship and ERC Starting Grant ‘DRAGNET’ (337062). PL acknowledges the support of the International Max Planck Research School Bonn/Cologne. SO is supported by the Alexander von Humboldt Foundation.

## REFERENCES

- Anholm M., Ballmer S., Creighton J. D. E., Price L. R., Siemens X., 2009, *Phys. Rev. D*, 79, 084030
- Arzoumanian Z. et al., 2014, *ApJ*, 794, 141
- Babak S., 2008, *Class. Quantum Gravity*, 25, 195011
- Babak S., Sesana A., 2012, *Phys. Rev. D*, 85, 044034
- Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307
- Burke-Spolaor S., 2013, *Class. Quantum Gravity*, 30, 224013
- Chamberlin S. J., Creighton J. D. E., Demorest P. B., Ellis J., Price L. R., Romano J. D., Siemens X., 2015, *Phys. Rev. D*, 91, 044048
- Demorest P. B. et al., 2013, *ApJ*, 762, 94
- Detweiler S., 1979, *ApJ*, 234, 1100
- Ellis J. A., 2013, *Class. Quantum Gravity*, 30, 224004
- Ellis J. A., Jenet F. A., McLaughlin M. A., 2012, *ApJ*, 753, 96
- Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Graham M. J. et al., 2015, preprint ([arXiv:1501.01375](https://arxiv.org/abs/1501.01375))
- Grishchuk L. P., 2005, *Phys.-Usp.*, 48, 1235
- Harry I. W., Allen B., Sathyaprakash B., 2009, *Phys. Rev. D*, 80, 104014
- Hellings R. W., Downs G. S., 1983, *ApJ*, 265, L39
- Hobbs G., 2013, *Class. Quantum Gravity*, 30, 224007
- Hobbs G. B., Edwards R. T., Manchester R. N., 2006, *MNRAS*, 369, 655
- Hobbs G. et al., 2010, *Class. Quantum Gravity*, 27, 084013
- Jaffe A. H., Backer D. C., 2003, *ApJ*, 583, 616
- Jenet F. A., Hobbs G. B., Lee K. J., Manchester R. N., 2005, *ApJ*, 625, L123
- Jenet F. A. et al., 2006, *ApJ*, 653, 1571
- Kramer M., Champion D. J., 2013, *Class. Quantum Gravity*, 30, 224009
- Lee K. J., Wex N., Kramer M., Stappers B. W., Bassa C. G., Janssen G. H., Karuppusamy R., Smits R., 2011, *MNRAS*, 414, 3251
- Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, *Phys. Rev. D*, 87, 104021
- Lentati L., Alexander P., Hobson M. P., Feroz F., van Haasteren R., Lee K. J., Shannon R. M., 2014, *MNRAS*, 437, 3004
- Lentati L. et al., 2015, *MNRAS*, 453, 2576
- McLaughlin M. A., 2013, *Class. Quantum Gravity*, 30, 224008
- Manchester R. N., IPTA, 2013, *Class. Quantum Gravity*, 30, 224010
- Manchester R. N. et al., 2013, *PASA*, 30, 17
- Petiteau A., Shang Y., Babak S., Feroz F., 2010, *Phys. Rev. D*, 81, 104016
- Petiteau A., Babak S., Sesana A., de Araújo M., 2013, *Phys. Rev. D*, 87, 064036
- Rajagopal M., Romani R. W., 1995, *ApJ*, 446, 543
- Ravi V., Wyithe J. S. B., Hobbs G., Shannon R. M., Manchester R. N., Yardley D. R. B., Keith M. J., 2012, *ApJ*, 761, 84
- Rosado P. A., Sesana A., 2014, *MNRAS*, 439, 3986
- Rosado P. A., Sesana A., Gair J., 2015, *MNRAS*, 451, 2417
- Sazhin M. V., 1978, *SvA*, 22, 36
- Sesana A., 2013, *MNRAS*, 433, L1
- Sesana A., Vecchio A., 2010, *Phys. Rev. D*, 81, 104008
- Sesana A., Haardt F., Madau P., Volonteri M., 2004, *ApJ*, 611, 623
- Sesana A., Vecchio A., Colacino C. N., 2008, *MNRAS*, 390, 192
- Sesana A., Vecchio A., Volonteri M., 2009, *MNRAS*, 394, 2255
- Sesana A., Roedig C., Reynolds M. T., Dotti M., 2012, *MNRAS*, 420, 860
- Shannon R. M. et al., 2013, *Science*, 342, 334
- Tanaka T., Menou K., Haiman Z., 2012, *MNRAS*, 420, 705
- Taylor S., Ellis J., Gair J., 2014, *Phys. Rev. D*, 90, 104028
- Taylor S. R. et al., 2015a, *Phys. Rev. Lett.*, 115, 041101
- Taylor S. R., Huerta E. A., Gair J. R., McWilliams S. T., 2015b, preprint ([arXiv:1505.06208](https://arxiv.org/abs/1505.06208))
- Valtonen M. J. et al., 2008, *Nature*, 452, 851

van Haasteren R., Levin Y., 2013, MNRAS, 428, 1147  
 van Haasteren R., Levin Y., McDonald P., Lu T., 2009, MNRAS, 395, 1005  
 van Haasteren R. et al., 2011, MNRAS, 414, 3117  
 Vilenkin A., 1981, Phys. Lett. B, 107, 47  
 Vilenkin A., Shellard E. P. S., 1994, Cosmic Strings and Other Topological Defects. Cambridge Univ. Press, Cambridge  
 Wyithe J. S. B., Loeb A., 2003, ApJ, 590, 691  
 Yardley D. R. B. et al., 2010, MNRAS, 407, 669  
 Yardley D. R. B. et al., 2011, MNRAS, 414, 1777  
 Zhu X.-J. et al., 2014, MNRAS, 444, 3709

<sup>1</sup>Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Golm, Germany

<sup>2</sup>Université Paris-Diderot-Paris7 APC - UFR de Physique, Bâtiment Concorcet, 10 rue Alice Domont et Léonie Duquet, F-75205 Paris, CEDEX 13, France

<sup>3</sup>School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

<sup>4</sup>Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

<sup>5</sup>Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Callinstraße 38, D-30167 Hannover, Germany

<sup>6</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

<sup>7</sup>Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

<sup>8</sup>Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany

<sup>9</sup>Laboratoire de Physique et Chimie de l'Environnement et de l'Espace LPC2E CNRS-Université d'Orléans, F-45071 Orléans, France

<sup>10</sup>ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, NL-7990 AA Dwingeloo, the Netherlands

<sup>11</sup>Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands

<sup>12</sup>Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester M13 9PL, UK

<sup>13</sup>INAF – Osservatorio Astronomico di Cagliari, via della Scienza 5, I-09047 Selargius, Italy

<sup>14</sup>Station de radioastronomie de Nançay, Observatoire de Paris, CNRS/INSU F-18330 Nançay, France

<sup>15</sup>Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China

<sup>16</sup>Astrophysics Group, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK

<sup>17</sup>TAPIR (Theoretical Astrophysics), California Institute of Technology, Pasadena, CA 91125, USA

<sup>18</sup>Fakultät für Physik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany

<sup>19</sup>Laboratoire Univers et Théories LUTH, Observatoire de Paris, CNRS/INSU, Université Paris Diderot, 5 place Jules Janssen, F-92190 Meudon, France

<sup>20</sup>Dipartimento di Fisica, Università di Cagliari, Cittadella Universitaria, I-09042 Monserrato, Italy

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.