The optimal schedule for pulsar timing array observations

K. J. Lee, 1,2*, C. G. Bassa, 2 G. H. Janssen, 2 R. Karuppusamy, 1,2 M. Kramer, 1,2 R. Smits, 2,3 and B. W. Stappers 2

1 Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany
2 Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester M13 9PL
3 Stichting ASTRON, Postbus 2, 7990 AA Dwingeloo, the Netherlands

ABSTRACT

In order to maximize the sensitivity of pulsar timing arrays to a stochastic gravitational wave background, we present computational techniques to optimize observing schedules. The techniques are applicable to both single- and multitelescope experiments. The observing schedule is optimized for each telescope by adjusting the observing time allocated to each pulsar while keeping the total amount of observing time constant. The optimized schedule depends on the timing noise characteristics of each individual pulsar as well as the performance of instrumentation. Several examples are given to illustrate the effects of different types of noise. A method to select the most suitable pulsars to be included in a pulsar timing array project is also presented.

Key words: gravitational waves – pulsars: general.

1 INTRODUCTION

Millisecond pulsars (MSPs) are stable celestial clocks, so that the timing residuals, the differences between the observed and the predicted time of arrival (TOA) of their pulses, are usually minute compared to the total length of the data span. A stochastic gravitational wave (GW) background leaves angular dependent correlations in the timing residuals of widely separated pulsars (for general relativity see Hellings & Downs 1983; for alternative gravity theories see Lee, Jenet & Price 2008; Lee et al. 2010), i.e. the correlation coefficient between timing residuals of a pulsar pair is a function of the angular distance between the two pulsars. Such a spatial correlation in pulsar timing signals makes it possible to directly detect GW using pulsar timing arrays (PTAs; Hellings & Downs 1983; Foster & Backer 1990). Previous analyses (Jenet et al. 2005) have calculated PTA sensitivity to a stochastic GW background generated by supermassive black hole (SMBH) binaries at cosmological distances (Jaffe & Backer 2003; Sesana et al. 2004). They have shown that a positive detection of the GW background is feasible, if one uses state-of-the-art pulsar timing technologies. Such encouraging results triggered consequent observational efforts.

At present, several groups are trying to detect GWs using PTAs: (i) the European Pulsar Timing Array (EPTA; Stappers et al. 2006; Ferdman et al. 2010; Lazaridis 2010; van Haasteren et al. 2011) with a subproject, the Large European Array for Pulsars (LEAP; Ferdman et al. 2010; Kramer & Stappers 2010), combining data from the Lovell telescope, the Westerbork Synthesis Radio Telescope, the Effelsberg 100-m Radio Telescope, the Nançay Decimetric Radio Telescope and the Sardinia Radio Telescope; (ii) the Parkes Pulsar Timing Array (PPTA; Manchester 2008; Hobbs et al. 2009; Verbiest et al. 2010) using observations with the Parkes Radio Telescope augmenteded by public domain observations from the Arecibo Observatory; (iii) the North American Nanohertz Observatory for Gravitational waves (NANOGrav; Jenet et al. 2009) using data from the Green Bank Telescope and the Arecibo Observatory and (iv) Kalyazin Radio Astronomical Observatory timing (Rodin 2011). Besides these ongoing projects, international cooperative efforts, e.g., the International Pulsar Timing Array (IPTA; Hobbs et al. 2010) or future telescopes with better sensitivity, e.g., the Five-hundred-meter Aperture Spherical Radio Telescope (Nan et al. 2006; Smits et al. 2009b) and the Square Kilometre Array (Smits et al. 2009a; Kramer & Stappers 2010), are planned to join the PTA projects to increase the chances of detecting GWs.

Operational questions arise naturally from such PTA campaigns, e.g., how should the observing schedule be arranged to maximize our opportunity to detect the GW signal? How much will we benefit from such optimization? In this paper, we try to answer these questions. The paper is organized as follows: in Section 2, we extend the formalism of Jenet et al. (2005) to calculate the GW-detection significance as a function of observing schedules, i.e. the telescope time allocation to each pulsar. Then we describe the technique to maximize the GW-detection significance in Section 3. Frameworks of the optimization problem are described in Section 3.1, and the

*E-mail: kjlee@mpifr-bonn.mpg.de

1 The Sardinia Radio Telescope was in the commissioning phase at the time of writing this paper.
algorithm to optimize a single- and multiple-telescope array are given in Sections 3.2 and 3.3, respectively. The results are presented in Section 4 and we discuss related issues in Section 5.

2 ANALYTICAL CALCULATION FOR GW-DETECTION SIGNIFICANCE

In this section, we calculate the statistical significance $S$ for detecting the stochastic GW background using PTAs. We consider TOAs from multiple pulsars, where each set may be collected from different telescopes or data acquisition systems. To detect the GW background, one correlates the TOAs between pulsar pairs and checks if the GW-induced correlation is significant. Jenet et al. (2005) have calculated the GW-detection significance for the case, where the noise in TOAs is of a white spectra with equal root-mean-square (RMS) level for all pulsars. To investigate the optimal observing schedule, we have to generalize the calculation, such that we can explicitly check the dependence of the GW-detection significance on the noise properties of each individual pulsar.

Under the influence of a stochastic GW background, the pulsar timing residual $R$ from a standard pulsar timing pipeline contains two components, the GW-induced signal $s$ and noise from other contributions $n$. In this section, we determine the statistical properties of $s$ and $n$ first, and then calculate the GW-detection significance.

2.1 Statistics for GW-induced pulsar timing signal

The spectrum of the stochastic GW background is usually assumed to be a power law, in which the characteristic strain ($h_c$) of the GW background is $h_c = A_0 f_0^{-5/3}$. Here, $A_0$ is the dimensionless amplitude for the background at $f_0 = 1$ yr$^{-1}$ and $\alpha$ is the spectral index. Under the influence of such a GW background, the power spectrum $S(f)$ of the GW-induced pulsar timing residual $s$ is (Jenet et al. 2005)

$$S(f) = \frac{A_0^2 f^{2\alpha-3}}{12\pi^3 f_0^6}.$$  

GWs perturb the space–time metric at the Earth. This introduces a correlation in the timing signal of the two pulsars. The correlation coefficient $H(\theta)$ between the GW-induced signals of the two pulsars with an angular separation of $\theta$ is called the Hellings and Downs function (Hellings & Downs 1983) given as

$$H(\theta) = \begin{cases} \frac{3\cos\theta}{8} - \frac{3(\cos\theta-1)}{2} \ln\left(\sin\left(\frac{\theta}{2}\right)\right), & \text{if } \theta \neq 0, \\ 1, & \text{if } \theta = 0. \end{cases}$$  

(2)

The spectral properties, equation (1), together with the spatial correlation, equation (2), fully characterize the statistical properties of the GW-induced signals.

For an isotropic GW background, the correlations between the GW-induced signals are

$$\langle s_i s'_j \rangle = \sigma_s^2 H(\theta_{ij}) \delta_{ik}.$$  

(3)

Here, we follow the notation that the subscript on the right is the index of the sampling and the superscript on the left is the index for the pulsar. For example, we denote the $k$th measurement of a timing residual of the $i$th pulsar at $R_k$, the GW-induced signal as $s_i$, and other noise contributions as $n_k$. $\delta_{ik}$ is the RMS level for the GW-induced signal, $\theta_{ij}$ is the angular distance between the $i$th and $j$th pulsar and $\delta_{ik}$ is the temporal correlation coefficient between the $k$th and $k'$th sampling. $\sigma_s$ and $\gamma_{ik}$ are numerically calculated from the GW spectrum as shown in Appendix A.

2.2 Statistics of noise components from other contributions

A purely theoretical modelling of the noise part $n$ is complex, because it (Foster & Backer 1990; Cordes & Shannon 2010; Shannon & Cordes 2010) depends on the properties of each individual pulsar, the instrumentation and radio frequency interference. We therefore model the noise phenomenologically using observational knowledge. In this paper, the noise part of a pulsar timing residual is modelled as a superposition of a white noise component and a red noise component, where the white noise is designated to the measurement uncertainty on the TOA due to radiometer noise and pulse jitter noise (Liu et al. 2012). Timing residuals of several MSPs show clear evidence of temporally correlated noise, although its origin is not yet clear (Verbiest et al. 2009). The red noise components are used to empirically model such effects. We further assume that the noise components are not correlated between any two different pulsars.

For each pulsar, three parameters are used to characterize the noise spectrum. These parameters are the RMS level for the white noise $\sigma_w$, the RMS level for the red noise $\sigma_r$ and the spectral index for the red noise $\beta$. The white noise spectrum is

$$S_w(f) = \frac{\sigma_w^2}{\sigma_w^2} ,$$  

and the red noise spectrum is

$$S_r(f) = \frac{\sigma_r^2 f^\beta}{\sigma_r^2} ,$$  

where $\sigma_w$ and $\sigma_r$ are for normalization. From the spectrum, one can derive the correlation between noise components

$$\langle n_i n_j \rangle = \left( \sigma_w^2 \delta_{ij} + \sigma_r^2 \gamma_{ij} \right) \delta_{ij} .$$  

(6)

Following our conventions, $n_i$ is the noise for the $i$th sampling of the $i$th pulsar. The $\delta_{ij}$ is the ‘Kronecker delta’ symbol, i.e. $\delta_{ij} = 1$, if $i = j$, otherwise $\delta_{ij} = 0$. The parameters $\sigma_w$, $\sigma_r$ and correlation coefficients $\gamma_{ij}$ are calculated using a numerical simulation shown in Appendix A.

2.3 GW background detection significance

Following Jenet et al. (2005), we calculate the cross power $\langle c \rangle$ between timing residuals and then compare it with the predicted correlation coefficient $H(\theta)$ to check whether the GW signal is significant. The cross power $\langle c \rangle$ is

$$\langle c \rangle = \frac{1}{m} \sum_{k=1}^{m} R_k^* R_i ,$$  

(7)

where $R_k^*$ and $R_i$ are the residual of the $i$th and $j$th pulsars, respectively, for the $k$th timing session, and $m$ is the number of data points for a given pulsar. It is assumed that the data from different pulsars overlap with each other and the number of data points is identical for all the pulsars, similar to the discussion in Yardley et al. (2011). These assumptions are good approximations in calculating the GW-detection sensitivity. In the case where the TOAs are misaligned or the number of data points is not identical, one can combine data points so that the number of data points is identical. This operation retains most of the GW-detection sensitivity due to the following reasons: (i) the RMS error reduces when data are

---

2 The correlated noise such as clock noise is discussed in Section 5 and Appendix E.
combined, and such a reduction in RMS error compensates the reduction in the number of data points. (ii) The spectrum of the GW-induced timing signal is steep; thus, most of the GW-induced signal is in the low-frequency components, which are preserved during the combining operation.

The comparison between the \(1/c\) and the Hellings–Downs function is carried out by doing another correlation, which gives the GW-detection significance \(S\) as

\[
S = \sqrt{M} \sum_{i \neq j} \langle (1/c - \bar{c})(H(1/\eta) - \bar{H}) \rangle \left( \sum_{i \neq j} (H(1/\eta) - \bar{H})^2 \right)^{-1/2},
\]

where the summation \(\sum_{i \neq j}\) sums over all independent pulsar pairs except the case where \(i = j\), i.e.

\[
\sum_{i \neq j} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i \neq j} = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i \neq j}
\]

and

\[
\bar{c} = \frac{1}{M} \sum_{i \neq j} 1/c
\]

and

\[
\bar{H} = \frac{1}{M} \sum_{i \neq j} H(1/\eta).
\]

Given \(N\) pulsars, the sum \(M = \sum_{i \neq j}\) is the number of independent pulsar pairs. To evaluate the quality of the detector, we need the expectation for the detection significance \((S)\), which is

\[
\langle S \rangle \sim \sqrt{M} \left[ 1 + \sum_{i \neq j} \langle (1/A + 1/B) \rangle \right]^{-1/2},
\]

where

\[
1/A = \frac{1}{M} \sum_{kk'} \left[ \left(1 + H(1/\eta)\right) \gamma_{kk'}^2 \right] + \left( \eta_{kk'} + \eta_{kk'}' \right) \gamma_{kk'} + \left( \eta_{kk'} + \eta_{kk'}' \right) \gamma_{kk'}
\]

and

\[
1/B = \frac{1}{M} \left( \eta_{w} + \eta_{w}' + \eta_{w} \eta_{w} + \eta_{w} \eta_{w} + \eta_{w} \eta_{w} \right),
\]

and \(\eta\) denotes the ratio between the power of noise components and the power of the GW-induced signal, which are

\[
\eta_{kk'} = \frac{\sigma_{kk'}^2}{\sigma_{kk'}^2},
\]

\[
\eta_{w} = \frac{\sigma_{w}^2}{\sigma_{w}^2}.
\]

If the noise level \(\sigma_{w}\) is identical for all pulsars and there is no red noise component \(\sigma_{w} = \sigma_{w}'\) and \(\sigma_{w} = \sigma_{w}' = 0\) for all the \(i, j, \ldots, N\), equation (15) reduces to the result found by Jenet et al. (2005) and Verbiest et al. (2009).

The \(1/A\) terms are independent of telescope integration time, because they only contain the RMS level of the GW-induced signal and the red noise, both of which are independent of telescope integration time. The observing schedule, the plan of allocating telescope time to each pulsar, only changes terms of \(1/B\), which depend on the ratio between signal and noise amplitude. By optimizing the observing schedule, we can reduce the summation of \(1/B\), such that the GW-detection significance is maximized. We present the technique to maximize the GW-detection significance in the next section.

## 3 OPTIMIZATION OF THE GW-DETECTION SIGNIFICANCE UNDER OBSERVING CONSTRAINTS

Pulsar timing observations are usually conducted as a series of successive observing sessions. In each session, multiple pulsars are observed using either one or multiple telescopes. There are basically two ways to use multiple telescopes. The simple way is to use each telescope independently, combine the TOA data from each telescope, remove the time jumps between each data set and form a single TOA data set (Janssen et al. 2008). The other way, as in the LEAP project, is to use telescopes simultaneously to form a phased array and then calculate the TOAs from the phased-array data. Due to such different methods of using multiple telescopes, the optimization techniques differ from one another. We answer the following questions in this section: (i) What are the variables to optimize? (ii) What are the constraints on the optimization? (iii) How do we perform the optimization?

### 3.1 The objective, variables and constraints of the optimization

Given a fixed amount of telescope time, we can adjust the amount of observing time allocated to each pulsar. Increasing the observing time for one pulsar will reduce its timing measurement error, but will increase the timing error for other pulsars. Naturally, for the purpose of detecting GWs, the optimization objective is to maximize the expected GW-detection significance \((S)\), while the constraint is the total amount of telescope time for the project.

Generally, for a timing project using \(N_{\text{tel}}\) telescopes to observe \(N\) pulsars, we need \(2N_{\text{tel}} + 3N\) input parameters to characterize the whole timing project. \(2N_{\text{tel}}\) parameters are used to characterize \(N_{\text{tel}}\) telescopes, where each telescope is quantified by the gain \((G)\) and the total available telescope time \((r)\). \(3N\) parameters are used to characterize the timing behaviour of \(N\) pulsars. The red noise level \(\sigma_{r}\) and spectral index \(\beta\) are assumed to be pulsar intrinsic. The observed white noise RMS levels \(\sigma_{w}\) depend on telescope gain, telescope time allocation to the pulsar and other parameters intrinsic to the pulsar (e.g. flux, pulse width, and so on).

For single-telescope cases, we can encapsulate the dependence of the white noise level on the schedule into the normalized white noise level \(\sigma_{0}\) and pulse jitter noise level \(\sigma_{j}\) (Foster & Backer 1990;
\[ \sigma_w = \left( \sigma_0^2 G^{-2} + \sigma_j^2 \right)^{1/2} \left( \frac{\tau}{1 h} \right)^{-1/2}, \tag{21} \]

where \( \sigma_w \) is the measured RMS level for the white noise component of the \( i \)th pulsar, \( \tau \) is the telescope time being used for the \( i \)th pulsar per observing session and \( G \) is the telescope gain. The normalized noise level \( \sigma_0 \) and the \( \sigma_j \) are used to characterize the radiometer noise and the pulse jitter noise of the pulsar. On the one hand, if there is no pulse jitter noise, \( \sigma_0 \) will be the observed RMS level of the white noise for a 1 h observation using a telescope with unit gain \( G = 1 \). On the other hand, if we have a telescope with infinite gain, the pulsar timing accuracy will be limited by the pulse jitter and \( \sigma_j \) will be the observed RMS level of white noise for a 1 h observation. If phased array observations are performed using multiple telescopes (as done in the LEAP project), the situation is identical to the case of a single telescope, and we use the effective gain for the array to determine the RMS level of noise.

For multiple incoherent telescopes, the gain of \( N_{\text{tel}} \) telescopes can be summarized by a vector \( G_\nu \), where \( \nu = 1, \ldots, N_{\text{tel}} \) and the \( \nu \)th component \( G_\nu \) is the gain for the \( \nu \)th telescope. In a similar fashion, the available telescope time of each telescope is summarized by the vector \( \tau_\nu \). The definition of the observing schedule becomes more complex, since we need to specify the observation time for each pulsar using each telescope. Furthermore, the schedule should also include information on the telescope availability, e.g., certain pulsars may not be visible to some telescopes due to geographical reasons. In this paper, we use the resource allocation matrix \( O \) to describe the telescope availability, where the \( i \)th row and \( \nu \)th column element \( O_{i,\nu} \) indicates whether we use the \( \nu \)th telescope to observe the \( i \)th pulsar, i.e., \( O_{i,\nu} = 1 \), if the \( \nu \)th telescope observes the \( i \)th pulsar and \( O_{i,\nu} = 0 \) otherwise. The telescope time allocation is described by another matrix, the schedule matrix \( P \), where the \( i \)th row \( \nu \)th column element \( P_{i,\nu} \) is the time allocated for the \( \nu \)th telescope observing the \( i \)th pulsar.

With the schedule matrix \( P \), the equivalent RMS level of the white noise in the combined data is

\[ \sigma_w = \left[ \sigma_0^2 \left( \sum_{\nu=1}^{N_{\text{tel}}} G_\nu^2 P_{i,\nu} O_{i,\nu} \right)^{-1} + \sigma_j^2 \left( \sum_{\nu=1}^{N_{\text{tel}}} P_{i,\nu} O_{i,\nu} \right)^{-1} \right]^{1/2}. \tag{22} \]

The main reason to introduce the resource allocation matrix \( O \) to take care of the complexity of telescope availability, and the telescope time can be treated in an identical way independent of the availability. For a single telescope or a phased array, the optimization constraint is

\[ \tau = \sum_{i=1}^{N} \tau_i, \tag{23} \]

i.e. the total telescope time is pre-fixed to be \( \tau \). For multiple incoherent telescopes, the above constraint is applied to each telescope individually, i.e. the constraints specify the available observation time \( \tau_\nu \) for each telescope, which gives

\[ \tau_\nu = \sum_{i=1}^{N} P_{i,\nu} O_{i,\nu}. \tag{24} \]

With the observing schedule (i.e. the vector \( \tau \) for a single telescope or the matrices \( O \) and \( P \), for incoherent telescopes), one can use equations (21) and (22) to determine the white noise level, and then determine the GW-detection significance as explained in Section 2.3. The optimization of the observing schedule means that we choose an appropriate \( \tau \) or \( P \) such that the expected GW-detection significance \( S \) is maximized under the constraint of equation (23) or (24).

From equation (15), the maximization of \( S \) is equivalent to the minimization of the term \( \sum_{i-j \text{ pairs}} (\sum_{\mu} \sigma_i^2 + \sigma_{\mu}^2) \). Because the terms of \( (\sum_{\mu} \sigma_i^2 + \sigma_{\mu}^2) \) are independent of the telescope time, the maximization of \( S \) by adjusting the observing schedule is thus equivalent to minimizing the objective function \( \mathcal{L} \) defined as

\[ \mathcal{L} = \sum_{i-j \text{ pairs}} \frac{1}{\mathcal{B}}. \tag{25} \]

### 3.2 Optimizing a single telescope

Our technique to optimize the single-telescope schedule takes two steps. The first step is to convert the constrained optimization problem to a constraint-free version, and the second step is to solve the constraint-free optimization. For comparison, we have also developed an alternative semi-analytical iterative technique in Appendix C.

To remove the constraints, we transform variables \( \tau \) to a new set of variables \( \theta_{\mu} \), where \( \mu = 1, \ldots, N-1 \), and the transformation is

\[
\begin{align*}
\theta_1 &= \arccos \left( \frac{\sum_{\nu=1}^{N_{\text{tel}}} G_\nu^2 P_{i,\nu} O_{i,\nu}}{\sum_{\nu=1}^{N_{\text{tel}}} \sum_{\nu'=1}^{N_{\text{tel}}} P_{i,\nu} O_{i,\nu} \sigma_j^2} \right), \\
\theta_2 &= \arccos \left( \frac{\sum_{\nu=1}^{N_{\text{tel}}} G_\nu^2 P_{i,\nu} O_{i,\nu} \sigma_i^2}{\sum_{\nu=1}^{N_{\text{tel}}} \sum_{\nu'=1}^{N_{\text{tel}}} P_{i,\nu} O_{i,\nu} \sigma_j^2} \right), \\
& \quad \vdots \\
\theta_{N-1} &= \arccos \left( \frac{\sum_{\nu=1}^{N_{\text{tel}}} G_\nu^2 P_{i,\nu} O_{i,\nu} \sigma_i^2}{\sum_{\nu=1}^{N_{\text{tel}}} \sum_{\nu'=1}^{N_{\text{tel}}} P_{i,\nu} O_{i,\nu} \sigma_j^2} \right), \\
\theta_N &= \arccos \left( \frac{\sum_{\nu=1}^{N_{\text{tel}}} \sum_{\nu'=1}^{N_{\text{tel}}} P_{i,\nu} O_{i,\nu} \sigma_j^2}{\sum_{\nu=1}^{N_{\text{tel}}} \sum_{\nu'=1}^{N_{\text{tel}}} P_{i,\nu} O_{i,\nu} \sigma_j^2} \right).
\end{align*}
\tag{26}
\]

of which the inverse transformation is

\[
\begin{align*}
\tau_1 &= \cos^2 \theta_1 \\
\tau_2 &= \sin^2 \theta_1 \cos^2 \theta_2 \\
\tau_3 &= \sin^2 \theta_1 \sin^2 \theta_2 \cos^2 \theta_3 \\
& \quad \vdots \\
\tau_{N-1} &= \prod_{\mu=1}^{N-2} \sin^2 \theta_\mu \cos^2 \theta_{\mu+1} \\
\tau_N &= \prod_{\mu=1}^{N-1} \sin^2 \theta_\mu \tag{27}
\end{align*}
\]

Here it is implicitly assumed that \( \tau \geq 0 \), the \( \prod_i \) and \( \prod_\mu \) are the serial products using the index \( i \) and \( \mu \), respectively. Equations (26) and (27) are, in fact, the transformation between an \( N \)-dimensional Cartesian coordinate system and its corresponding hyperspherical coordinate system. The constraint in the new spherical coordinate system corresponds to fixing the radius of the hypersphere (fixing \( \tau \)), and all angular coordinates \( \theta_\mu \) are free variables for which one can specify any value for \( \theta_\mu \) without breaking the constraint, i.e. after the above coordinate transformation, the objective \( \mathcal{L} \) becomes a function of a new variable \( \theta_\mu \); the constraint equation (23) is automatically satisfied for any \( \theta_\mu \). We can then find the minimum of \( \mathcal{L} \) using numerical methods for constraint-free problems. In this paper, the downhill simplex method (Nelder & Mead 1965) is adopted.
which has, usually, better global converging behaviour than other methods (Kramer et al. 1994). Once the optimal \( \theta_{\nu} \) are found, we transform them back to \( \tau \) using equation (27), which yields the optimal single-telescope schedule. In practical situations, one also needs to add telescope slewing time, observing time for calibration sources and other necessary auxiliary time to top of this telescope time schedule to get the final schedule.

### 3.3 Optimizing multiple incoherent telescopes

Similarly to Section 3.2, the optimal observing schedule for multiple incoherent telescopes is calculated by minimizing \( L \), although the constraints are slightly more complex here. In this section, we present the optimizing technique and discuss later the relation between the optimization of multiple telescopes and single-telescope optimization.

From the multiple-telescope constraint equation (24), we can see that the constraints are applied to each telescope individually. Thus, the generalization of the method presented in the previous section is straightforward by applying the transformation equation (26) to each telescope separately. Take telescope 1 as an example. The first column of matrix \( P \), the \( P_{\nu,1} \), is the observing schedule for the first telescope. We can transform those components of \( P_{\nu,1} \) indicated by \( O_{\nu} = 1 \) using equation (26) to remove the constraint of the first telescope. Similarly, by applying the transformation to the other columns of matrix \( P \) successively, one can remove all the constraints. With the new constraint-free variables, we use the downhill simplex method to find the optimization. We then transform back to \( P_{\nu,1} \), which is the optimization schedule.

In the optimization algorithm, we treat the resource allocation matrix \( O \) as input knowledge. A question naturally arises as to whether one can find a better observing schedule for the same telescopes with the same amount of telescope time but with a different \( O \), i.e. whether one can add pulsars to or remove pulsars from schedules of certain telescopes to increase the detection significance? The configuration with all \( O_{\nu} = 1 \) allows one to use any telescope to observe any pulsar, i.e. allows one to adjust the schedule with the maximal degrees of freedom. In this way, the optimization schedule of configuration \( O_{\nu} = 1 \) leads to the highest GW-detection significance. If any schedule has the same detection significance as the optimal schedule with all \( O_{\nu} = 1 \), we call such schedule 'global optimal'. To determine whether the global optimization is achievable, we first investigate the case when the pulse jitter noise can be ignored and then discuss situations where the pulse jitter noise becomes important.

When the pulse jitter noise is neglected, there is a close relation between the single-telescope optimization and the multiple-telescope optimization. In fact, under certain conditions, the optimization for multiple telescopes is equivalent to the single-telescope optimization. To see this, we replace the variables \( O_{\nu} \) and \( P_{\nu,1} \) in equation (22) by the effective telescope time \( \tau_{e} \) as follows:

\[
\tau_{e} = \sum_{i=1}^{N_{\nu}} \tau_{e,0} G_{\nu}^{2} P_{\nu,0}/O_{\nu,0}.
\] (28)

After ignoring \( \sigma_{\nu} \), equation (22) becomes

\[
\tau_{e} = \sigma_{\nu,0}^{1/2} \tau_{e,0}^{-1/2},
\] (29)

and the constrain equation (24) reduces to a single constraint

\[
\sum_{i=1}^{N} \tau_{e,i} = \tau_{e,0} \equiv \sum_{i=1}^{N_{\nu}} \tau_{i} G_{\nu}^{2}.
\] (30)

By comparing equations (29) and (30) with equations (23) and (21), one can see that the optimization for multiple telescopes is very similar to the single-telescope optimization. In fact, if there exists a unique solution of \( P_{\nu} \) to equation (28), which satisfies each individual constraint of equation (24), the multiple-telescope and single-telescope optimization are mathematically identical.

The differences between multiple-telescope and single-telescope optimization lie in the difference between the constraint equations (24) and (30). For the single-telescope case, only one constraint (equation (23)) is involved. This is very different from the case of multiple telescopes, where \( N_{\nu} \) constraints (equation 24) are present. The variable substitution in equation (28) combines all \( N_{\nu} \) constraints (equation 24) and forms a single constraint (equation 30), where the constraint of each individual telescope is ignored. Whether global optimization is achievable is, now, equivalent to whether one can find a solution to \( P_{\nu} \) for equation (28), while satisfying all individual constraints (equation 24).

Generally, when solving \( P_{\nu} \) from equations (28) and (24), one meets three types of situations: a unique solution, multiple solutions and no solution. For the case of multiple solutions, there are multiple choices for the optimal schedule. All these configurations are identical in the sense that they give the same GW-detection significance. The case of no solution can only arise when constraints for some telescopes cannot be met, i.e. some of the pulsars need more time than the telescopes can give, while some of the pulsars have more telescope time than should be assigned. Take the case with two telescopes and two pulsars as an example, where the gain of the two telescopes are the same, two pulsars have identical \( \sigma_{\nu} \) and \( \tau_{e,0} \) as \( [ \nu \nu ] \), i.e. only the second telescope can observe the second pulsar. Since the two pulsars have the same \( \sigma_{\nu} \), the total effective telescope time should then also be the same for the optimal schedule (i.e. \( \tau_{e,0} = \tau_{e,0} \)). However, if the first telescope does not have enough time, one gets \( \tau_{e,0} < \tau_{e,0} \) and the global optimal solution is not achievable for such configurations.

The case for which there is no solution to equation (28) is due to an improper choice of telescopes. Most of the time the matrix \( O \) is determined by the sky coverage of the telescopes. Thus, if no solution can be found, one needs to seek telescopes with the appropriate sky coverage or extend the time for telescopes with the sky coverage. In order to identify such a no-solution situation we show in Appendix D that a solution to equation (28) exists, if the following conditions are satisfied:

\[
\tau_{e} \leq \sum_{i=1}^{N_{\nu}} G_{\nu}^{2} P_{\nu,0}/O_{\nu,0}, \text{ for any index of pulsar } \nu.
\] (31)

One can identify which pulsar in equation (31) fails. These failures indicate the corresponding elements of the resource allocation matrix to be adjusted.

We now consider the situation where the pulse jitter noise becomes important. Clearly, if we cannot ignore the pulse jitter noise, the effective telescope time is no longer linearly dependent on the telescope time \( P_{\nu} \) as in equation (28) and we do not have a simple method to check if the global optimization is achieved. However, we can still set all the \( O_{\nu} = 1 \) find the global optimal strategy and compare it with the optimal strategy for the input \( O_{\nu} \) to check if the global optimization is achieved.

In Fig. 1, we give four examples for the optimization of multiple telescopes. In these examples, two telescopes are used to observe two pulsars, where the pulse noise parameters and the telescope parameters are specified in Table 1. These four examples are given as follows: (i) 'Case A', two identical pulsars are observed with
The optimal schedule for pulsar timing array observations

Figure 1. Four examples to illustrate the optimization of multiple telescopes. Here, we plot $L$, defined in equation (25), as a function of the telescope time. The smaller the $L$, the better the observing schedule. In all examples, two pulsars are observed with two telescopes. The telescope configurations and pulsar noise parameters for each case are given in Table 1. The $x$-axis and $y$-axis are for the $\tau_1$, i.e. the time for the first telescope observing the first pulsar, and $\tau_2$, i.e. the time for the second telescope observing the second pulsar, respectively.

Table 1. The telescope configurations and pulsar noise parameters for each case given in Fig. 1. $\sigma_0$ and $\sigma_j$ take the unit of second, $\tau$ takes the unit of hour and the gain $G$ takes an arbitrary fiducial unit as explained in the main text.

| Case  | Radiometer noise | Jitter noise | Observation time | Telescope gain | Resource allocation matrix |
|-------|------------------|--------------|------------------|----------------|--------------------------|
| A     | $10^{-7}$        | 0            | 1                | 1              | $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ |
| B     | $10^{-7}$        | 0            | 1                | 2              | $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ |
| C     | $10^{-7}$        | $10^{-7}$    | 1.5              | 1              | $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ |
| D     | $10^{-7}$        | 0            | 1                | 2              | $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ |

As indicated in Fig. 1, as long as we assign the same amount of effective telescope time $\tau_e$ for two pulsars, it is the optimal observing schedule. (ii) ‘Case B’, telescope 1 has twice the gain compared to telescope 2. Similar to ‘case A’, the global optimal schedule is achievable and one can exchange telescope time. But due to the gain differences, the telescope time exchange should be weighted by the gain, such that the same amount of effective telescope time is assigned to the two identical pulsars. (iii) ‘Case C’, where telescope 1 has more time (1.5 h) available compared to telescope 2 (1 h) and only telescope 2 can be used to observe the second pulsar (as indicated by the matrix $O_\nu$). For this case, the total telescope time is the same as in ‘case A’, but we do not achieve the same low level of $L$ as in ‘case A’, i.e. global optimization is not achievable. This is due to the constraint from matrix $O_\nu$, which prevents us from reaching the global optimization as discussed. (iv) ‘Case D’, where pulsar 1 is affected by pulse jitter noise. In
this case, one does not have the freedom to exchange telescope time and optimization suggests that the low-gain telescope (telescope 2) should spend more time on the pulse jitter affected pulsar (pulsar 1).

4 RESULTS

As examples, we use pulsar properties measured from data of the PPTA, EPTA and NANOGrav to show the potential benefit of optimizing the observing schedule. The parameters of the pulsars are given in Table 2.

Fig. 2 shows the comparison between the GW-detection significance (5) for an unoptimized and an optimized PTA with the same two to three times weaker than unoptimized arrays depending on the curve, the optimized arrays are able to detect a GW background the GW-detection significance. Evaluating from the rising edge of the curve, the optimized arrays are able to detect a GW background and a good approximation for the optimal schedule is

\[ t^* = \tau \left( \frac{\sqrt{Q}}{\sum_{j=1, \ldots, N} \sqrt{Q}} \right), \]

(32)

We also note that the larger the red noise level, the less we gain from optimizing the observing schedule. When the amplitude of intrinsic red noises is large, the \( IA \) dominates the detection significance as shown in equation (15), and thus the detection is no longer sensitive to the schedule optimization, which only affects the \( JB \) terms. In fact, only if the pulsar noise level is sensitive to the observing schedule (i.e. when red noise does not dominate), the optimization will be effective. This conclusion applies to any type of GW detector.

In the optimization process, one always uses a numerical technique to determine the optimal observation plan. It is, nevertheless, worth finding a rule of thumb to determine an ‘approximate’ optimal strategy. We prove in Appendix C that, for strong GW cases, the optimal observing schedule weakly depends on the amplitude of the GW background and a good approximation for the optimal schedule is

\[ t^* = \tau \left( \frac{\sqrt{Q}}{\sum_{j=1, \ldots, N} \sqrt{Q}} \right), \]

Table 2. Parameters for PPTA, EPTA and NANOGrav pulsars taken from Hobbs et al. (2010). We assume that all the white noise is due to the radiometer noise and pulse jitter noise can be ignored.

| PSR J | P (ms) | \( P_b \) (d) | S1400 (mJy) | Array | PPTA \( \sigma_0 \) (\( \mu \)s) | EPTA \( \sigma_0 \) (\( \mu \)s) | NANOGrav \( \sigma_0 \) (\( \mu \)s) |
|-------|-------|---------------|-------------|-------|----------------|----------------|----------------|
| J0030+0451 | 4.87 | – | 0.6 | EPTA, NANOGrav | – | 0.54 | 0.31 |
| J0218+4232 | 2.32 | 2.03 | 0.9 | NANOGrav | – | – | 4.81 |
| J0437−4715 | 5.76 | 5.74 | 142.0 | PPTA | – | 0.03 | – |
| J0613−0200 | 3.06 | 1.20 | 1.4 | PPTA, EPTA, NANOGrav | 0.71 | 0.45 | 0.50 |
| J0621+1002 | 28.85 | 8.32 | 1.9 | EPTA | – | 9.58 | – |
| J0711−6830 | 5.49 | – | 1.6 | PPTA | 1.32 | – | – |
| J0751+1807 | 3.48 | 0.3 | 3.2 | EPTA | – | 0.78 | – |
| J0900−3144 | 11.1 | 18.7 | 3.8 | EPTA | – | 1.55 | – |
| J1012+5307 | 5.26 | 0.60 | 3.0 | EPTA, NANOGrav | – | 0.32 | 0.61 |
| J1022+1001 | 16.45 | 7.81 | 3.0 | PPTA, EPTA | 0.37 | 0.48 | – |
| J1024−0719 | 5.16 | – | 0.7 | PPTA, EPTA | 0.43 | 0.25 | – |
| J1045−4509 | 7.47 | 4.08 | 3.0 | PPTA | 2.68 | – | – |
| J1455−3330 | 7.99 | 76.17 | 1.2 | EPTA, NANOGrav | – | 3.83 | 1.60 |
| J1600−3053 | 3.60 | 14.35 | 3.2 | EPTA, PPTA | 0.32 | 0.23 | – |
| J1603−7202 | 14.84 | 6.31 | 3.0 | PPTA | – | – | – |
| J1640+2224 | 3.16 | 175.46 | 2.0 | EPTA, NANOGrav | – | 0.45 | 0.19 |
| J1643−1224 | 4.62 | 147.02 | 4.8 | EPTA, NANOGrav | 0.57 | 0.56 | 0.53 |
| J1713+0747 | 4.57 | 67.83 | 8.0 | PPTA, EPTA, NANOGrav | 0.15 | 0.07 | 0.04 |
| J1730−2304 | 8.12 | – | 4.0 | PPTA, EPTA | 0.83 | 1.01 | – |
| J1732−5049 | 5.31 | 5.26 | – | PPTA | 1.74 | – | – |
| J1738+0333 | 5.85 | 0.35 | – | NANOGrav | – | – | 0.24 |
| J1741+1351 | 3.75 | 16.34 | – | NANOGrav | – | – | 0.19 |
| J1744−1134 | 0.08 | – | 3.0 | PPTA, EPTA, NANOGrav | 0.21 | 0.14 | 0.14 |
| J1751−2857 | 3.91 | 110.7 | 0.06 | EPTA | – | 0.90 | – |
| J1824−2452 | 3.05 | – | 0.2 | PPTA, EPTA | 0.39 | 0.24 | – |
| J1853+1303 | 4.09 | 115.65 | 0.4 | NANOGrav | – | – | 0.17 |
| J1857+0943 | 5.37 | 12.33 | 5.0 | PPTA, EPTA, NANOGrav | 0.82 | 0.44 | 0.25 |
| J1909−3744 | 2.95 | 1.53 | 3.0 | PPTA, EPTA, NANOGrav | 0.19 | 0.04 | 0.15 |
| J1910+1256 | 4.98 | 58.47 | 0.5 | EPTA, NANOGrav | – | 0.99 | 0.17 |
| J1918−0642 | 7.65 | 10.91 | – | EPTA, NANOGrav | – | 0.87 | 1.08 |
| J1939+2134 | 1.56 | – | 10.0 | PPTA, EPTA, NANOGrav | 0.11 | 0.02 | 0.03 |
| J1955+2908 | 6.13 | 117.35 | 1.1 | NANOGrav | – | – | 0.18 |
| J2019+2425 | 3.94 | 76.51 | – | NANOGrav | – | – | 0.66 |
| J2124−3358 | 4.93 | – | 1.6 | PPTA | 1.52 | – | – |
| J2129−5721 | 3.73 | 6.63 | 1.4 | PPTA | 0.87 | – | – |
| J2145−0750 | 16.05 | 6.84 | 8.0 | PPTA, EPTA, NANOGrav | 0.86 | 0.40 | 1.37 |
| J2317−1439 | 3.44 | 2.46 | 4.0 | NANOGrav | – | 0.81 | 0.25 |
The optimal schedule for pulsar timing array observations

Figure 2. The GW background detection significance for PPTA, EPTA and NANOGrav with 5 yr biweekly (one session every two weeks) observations. The white noise levels $\sigma_0$ are taken from Table 2 and the red noise levels $\sigma_i$ are given above each panel. The $x$-axis shows the characteristic strain $A_0$ of the GW background with spectral index of $\alpha = -2/3$. The $y$-axis shows the expected detection significance $\langle S \rangle$. The solid lines, dashed lines and dash--dotted lines are for PPTA, EPTA and NANOGrav, respectively. The thick lines are the optimized cases, while the thin lines are the unoptimized versions. Here, the constraint is the observation time, i.e. for each project, the total amount of observing time of each session is fixed to be 20 h. If the red noise level is zero, the optimized array is able to detect two to three times weaker GW signals compared to its unoptimized version depending on the pulsar population.

where $iQ$ is defined as

$$iQ = \sigma_0^2 G^{-2} + \sigma_i^2.$$  

(33)

Here $iQ$, noise parameter, defines the noisiness of the $i$th pulsar observed using a telescope of gain $G$.

By comparing the numerical optimization and the result from equation (32) in Fig. 3, we show that the optimal schedule is insensitive to the GW amplitude and can be well approximated by equation (32), when the amplitude of the GW is large. For the case of a weak GW background, the optimal schedule for most of the pulsars is still close to equation (32), but the optimal schedule for noisy pulsars (pulsars with larger $iQ$) starts to deviate from the analytic approximation.

There are a few more general conclusions on improving the timing accuracy independent of GW-detection algorithms. In order to improve timing accuracy, one can use telescopes with higher effective gain or increase telescope time. Increasing the telescope gain reduces $\sigma_w$ and increasing telescope time reduces both $\sigma_w$ and $\sigma_i$. The example of case ‘D’ in Fig. 1 shows that one should use a low-gain telescope on those pulsars with a larger jitter noise level, while using a high-gain telescope on pulsars with larger radiometer noise. This is further supported by the results in Fig. 4, which shows that increasing the telescope time is more effective than using a high-gain telescope for pulse jitter noise dominant pulsars.

Besides providing the optimal schedule to detect a GW background, our technique answers the question about the optimal number of pulsars one should observe in a PTA with a given amount of telescope time. The number of pulsars in a PTA has two effects on the GW-detection significance. First, from equation (15), one can see that the significance increases as $\langle S \rangle \propto \sqrt{M} \sim N$. Secondly, since the telescope time is limited, observing more pulsars increases the TOA noise level, i.e. $\sigma_w^2 \propto N^{-1}$ given a fixed amount of telescope time. When $N$ becomes large, the two effects mentioned above cancel each other out, and the detection significance becomes insensitive to $N$. In general, when the number of pulsars $(N)$ is small, the detection significance increases with $N$. If all pulsars have the same noise level, the detection significance saturates for large $N$, where the saturation level is mainly determined by the available telescope time. In practice, when trying to include more pulsars in a timing array, pulsars with a higher noise level will be inevitably included, such that the detection significance will decrease for any GW-detection algorithm. In this way, observing more pulsars does not necessarily help detecting the GW background, unless one gets more telescope time. Given the telescope time, the number of pulsars, at which the GW-detection significance achieves its maximum, is the optimal number of pulsars one should use in the PTA. We propose the following algorithm to determine the best sample of pulsars to observe.

(i) From a group of to-be-observed pulsars, choose the two pulsars with smallest noise levels, then optimize the schedule and compute the GW-detection significance.

---

4 Here, telescopes with higher effective gain $G$ can be telescopes with larger collection area, lower system temperature, wider bandwidth and so on.
case, we conclude that it is important, even for a high-gain \( \sigma \) and telescope gain are input parameters. The optimal telescope time for each pulsar as a function of its noise level is shown in Figure 3. The x-axis shows the 1 h pulsar noise level \( \sigma_0 \). The y-axis shows the optimal telescope time for the pulsar per observing session. The EPTA pulsars are used in this demonstration, where the total telescope time is 24 h, i.e. the average telescope time for each pulsar is 1 h. We indicate \( \sigma_0 \) for each pulsar using a vertical line on the top of the figure. The dot–dashed line, dashed line, dotted line and solid line are for optimization results with \( \sigma = 10^{-12} \), \( 10^{-13} \), \( 2 \times 10^{-13} \) and \( 10^{-14} \), respectively. The results of equation (32) overlap with the dot–dashed line. For GWs with amplitude between \( 10^{-12} \) and \( 10^{-13} \), the optimal schedules are very close to each other. For the weaker GW cases, e.g. \( A_0 = 10^{-14} \), the optimal schedule starts to deviate from the approximation equation (32). Such a deviation is mainly due to the pulsars with a high noise level. Since these noisy pulsars will not contribute significantly to the GW signal detection, the optimal algorithm starts to reduce its observing time. For most of the pulsars, the optimal schedule is still close to the result from equation (32) for multiple telescopes. We also examine the links between the multiple- and single-telescope optimization. We investigate the optimal number of pulsars to observe for a given PTA, where the algorithm to construct the optimal group of pulsars from candidates is also included. In our optimization, the total telescope time \( \tau \) and \( \tau_r \) are input parameters that need to be determined before optimizing the schedule. When defining a PTA project, one can start with a reasonable amount of telescope time, say 20 h per each session/telescope, optimize the schedule, calculate the detection significance \( S \), check if the detection is sensitive enough to the predicted GW background and then adjust the input total telescope time accordingly, i.e. increase total telescope time, if the PTA is not sensitive enough. In practice, the detailed numerical values for the optimal observing strategy are still to be determined, due to the lack of measurements for the necessary pulsar noise parameters. To determine the realistic optimization strategy, these parameters are critical. A detailed investigation on the individual timing properties of potential PTA pulsars is highly necessary, from which further observations will benefit.

One may encounter the situation, in which the optimal strategy requires longer observing time for certain pulsars per session than their transit time. If this happens, one has to split one observing session into several successive sessions. The GW-detection significance is not significantly impaired by session splitting, since the GW-induced timing signal has a very red spectrum and we are detecting its low-frequency component. However, the session splitting can lead to inconveniences in practical arrangements. A better way to avoid such a situation is to construct the PTA using telescopes
with enough geographical coverage, which can be one of the driving reasons for the IPTA project.

The optimization in this paper is designed to maximize the GW-detection significance. Our optimization is built for the correlation detector proposed by Jenet et al. (2005). As shown in Appendix C, our optimization, which is very close to minimizing the total PTA noise power, can be different from optimization for other types of GW detectors e.g. frequentist detectors (Verbiest et al. 2009; Yardley et al. 2011) or the Bayesian detector (van Haasteren et al. 2009, 2011). In fact, as already shown by Burt, Lommen & Finn (2011), one can get a different optimal observation schedule, when focusing on single-source detection. We have ignored the information of historical non-overlapping data in our algorithms, because the non-overlapping data have very limited contributions to the cross-power of GW signals, although these data can be important to constrain the upper limit of GW amplitude (van Haasteren et al. 2011). Furthermore, the PTA offers an opportunity to investigate much broader topics, e.g., interstellar medium effects, time metrology and so on. This paper, thus, by no means claims that our objective function is the only one we should pursue. However, our basic framework of optimization, constraints and related numerical techniques will be the same for other detectors. It is straightforward to generalize our method to these detectors. For Bayesian detectors, there is no analytical expression for the detection significance at present. The Bayesian detectors are usually computationally expensive, which makes the optimization difficult at this stage.

Fig. 2 shows that increasing the levels of red noise does not decrease the saturation level of the detection significance, i.e. the detection significance at large GW amplitude. This seems to contradict the conclusion of Jenet et al. (2005) and Verbiest et al. (2009), where the red spectrum of the GW limits the saturation level. As shown in equation (15), the term limiting the saturation level is due to the term $(1 + H(i\theta))vy_k^2$ in $\langle A \rangle$, which is the non-zero correlation between the GW signals of two different pulsars. Since the red intrinsic noise of pulsars, unlike the GW-induced signal, are uncorrelated, it is natural that they do not limit the saturation level.

In this paper, we use a phenomenological model to describe the noise component, which is a superposition of a telescope-time-dependent white noise component and another red noise component independent of telescope time. The reason for using such a phenomenological model is to use the observational information and to introduce minimal theoretical assumptions. Our white noise term contains both the radiometer noise and the pulse jitter noise. Although the radiometer noise is the current bottleneck in the time-amplitude of most MSPs ($\sigma_j G^{-1} \gg \sigma_i$), the pulse jitter noise (Liu et al. 2011) can be a potential limitation for future single-dish high-gain telescopes. Similarly, red noise can be another limitation for the long-term timing accuracy. Detailed studies on the pulse jitter noise and red noise properties and related mitigation algorithms will be useful for the future prospects of GW detection.

We assumed that the noise sources are not correlated among pulsars. This may be valid for all the noise of astrophysical origin, although the clock error can be an identical noise among all pulsars (Hogan & Rees 1984; Foster & Backer 1990; Manchester 1994; Tinto 2011). The clock error may also introduce correlations in TOAs from different telescopes, since observatory clocks are usually synchronized. However, thanks to the red spectrum of most clock errors (Riehle 2004), one can completely remove the noise using simultaneous differential measurements (Tinto 2011). Furthermore, as we argue in Appendix E, such a common noise source can be significantly suppressed by post-processing, if each session is compact within the time-scale of days. In this way, our assumption is justified that different noise sources are not correlated among pulsars.

Our optimal observation strategy is for allocating telescope time among pulsars. This only affects the white noise related part ($\langle B \rangle$ terms) in GW-detection significance (equation (15). In fact, one can also specify the epoch of each session to minimize the effect of the red noise related part ($\langle A \rangle$ terms). Since these $\langle A \rangle$ terms can be more effectively reduced using whitening techniques (Jenet et al. 2005, 2006), we do not optimize the epoch of each session to avoid sampling artefacts and reduce the complexities in observation management. The discussions for the frequency of observation sessions are omitted, since it is not bounded in terms of optimization, i.e. observing more frequently sessions simply increase the sensitivity.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the ERC Advanced Grant ‘LEAP’, Grant Agreement Number 227947 (PI: Michael Kramer). We thank Nobert Wex for reading through the manuscript and his useful discussions. We also thank the anonymous referee for the important and helpful comments.

REFERENCES

Burt B. J., Lommen A. N., Finn L. S., 2011, ApJ, 730, 17
Cordes J. M., Shannon R. M., 2010, preprint (arXiv:10103785)
Ferdman R. D. et al., 2010, Class. Quantum Grav., 27, 084014
Foster R. S., Backer D. C., 1990, ApJ, 361, 300
Helling R. W., Downs G. S., 1983, ApJ, 265, L39
Hobbs G. B. et al., 2009, Publ. Astron. Soc. Australia, 26, 103
Hobbs G. G., 2010, Class. Quantum Grav., 27, 084013
Hogan C. J., Rees M. J., 1984, Nat, 311, 109
Jaffe A. H., Backer D. C., 2003, ApJ, 583, 616
Janssen G. H. et al., 2008, A&A, 490, 753
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
Jenet F. A. et al., 2009, preprint (arXiv:09091058)
Kramer M., Stappers B., 2010, in ISKAF2010 Science Meeting, LOFAR, LEAP and Beyond: Using Next Generation Telescopes for Pulsar Astrophysics. Assen, the Netherlands, p. 10
Kramer M., Stappers B., 2010, Proceeding of Science, ISKAF2010, 2010, 034
Lazaridis K., 2010, in Tsiganos K., Hatzidimitriou D., Matsakos T., eds, ASP Conf. Ser. Vol. 424, High Precision Milisecond Pulsar Timing with EPTA. Astron. Soc. Pac., San Francisco, p.394
Lee K. J., Jenet F. A., Price R. H., 2008, ApJ, 685, 1304
Lee K., Jenet F. A., Price R. H., Wez N., Kramer M., 2010, ApJ, 722, 1589
Liu K., Verbist J. P. W., Kramer M., Stappers B., van Staten W., Cordes J. M., 2011, MNRAS, 417, 2916
Liu K., Keane E. F., Lee K. J., Kramer M., Cordes J. M., Purver M. B., 2012, MNRAS, 420, 361
Manchester R. N., 1944, PASA, 11.1
Manchester R. N., 2008, in Bassa C., Wang Z., Cumming A., Kaspi V. M., eds, AIP Conf. Ser. Vol. 983, The Parkes Pulsar Timing Array Project. Am. Inst. Phys., New York, p.584
Nan R., Wang Q., Zhu L., Zhu W., Jin C., Gan H., 2006, Chinese J. Astron. Astrophys. Suppl., 6, 020000
Nelder J. A., Mead R., 1965, Comput. J., 7, 303
Riehle F., 2004, Frequency Standards: Basics and Applications. Wiley-VCH, Weinheim
Rodin A. E., 2011, Astron. Rep., 55, 132
Sesana A., Haardt F., Madau P., Volonteri M., 2004, ApJ, 611, 623
Shannon R. M., Cordes J. M., 2010, ApJ, 725, 1607
Smits R., Kramer M., Stappers B., Lorimer D. R., Cordes J., Faulkner A., 2009a, A&A, 493, 1161

© 2012 The Authors, MNRAS 423, 2642–2655
Monthly Notices of the Royal Astronomical Society © 2012 RAS
In this section, we calculate the cross-correlation for a Gaussian random signal with a power-law spectrum. We discuss two main issues arising in the calculation, the effect of polynomial fitting on the cross-correlation and the effects of spectral leakage. For a stationary continuous-in-time random signal $s(t)$, which has a prescribed power spectrum of $S_0(f)$, the well-known Wiener–Khinchin theorem states that the autocorrelation $(s(t)s(t))$ is the Fourier transform of the power spectral density, i.e.

$$\langle s(t)s(t) \rangle = \int_0^\infty S_0(f)e^{2\pi if(t-t)}\,df.$$  \hfill (A1)

However, due to the fitting of polynomials, the direct application of the above to the pulsar timing problem needs revision. In a practical pulsar timing data reduction pipeline, one usually uses a least-squares polynomial fitting to extract the pulsar parameters. Such a fit is not a stationary process, which prevents us from calculating the correlation directly using equation (A1). In this paper, correlations are calculated via numerical simulations, which take the following steps: (i) generate a series of the sampled signal using individual frequency components as described in Lee et al. (2008). (ii) Fit the signal with a polynomial to simulate the effects of fitting the pulsar rotation frequency and its derivative. (iii) Calculate the required cross-correlation. (iv) Repeat steps (i), (ii), (iii) and average the cross-correlation, until the required precision is attained. In this paper, the correlations are calculated to a relative error of 1 per cent. From the numerical results in Fig. A1, the polynomial fitting clearly breaks the stationarity of signals.

The other issue is the spectral leakage. A red noise signal with a steep spectrum introduces leakage from low-frequency components within the frequency range $[f, f + df]$ is $f^{-\beta/2}df^{1/2}$. The waveform of this component will be sinusoidal, i.e.

$$s(t) \simeq f^{-\beta/2}f^{1/2}\exp(2\pi f_0t).$$

At the low-frequency limit, where $f \ll 1$, $s(t)$ can be approximated using Taylor series:

$$s(t) \sim f^{-\beta/2}f^{1/2}\exp(2\pi f_0t) \simeq f^{1/2}\sum_{l=0}^{\infty} \frac{(2\pi t)^l}{l!}.$$  \hfill (A2)

If we fit the waveform with an $n$-degree polynomial, we effectively remove the leading terms up to $f^{-1+n/2-\beta/2}$. Clearly, if $n + 3/2 - \beta/2 < 0$, the signal amplitude goes to infinity, when $f \to 0$. Thus, there will be no low-frequency cut-off in the spectrum, even for data with a finite length. In this case, the low-frequency components are always dominant. To guarantee that the low-frequency cut-off rises naturally (regularize the signal), we need to fit the time series to a polynomial with degree $n > \beta/2 - 3/2$. For example, the SMBH...
where \(C(t_1 - t_2) = \langle s(t_1)s(t_2) \rangle \). For the case of a power-law spectrum, we have
\[
C(t) = S_0 T^{-\beta-1} \frac{\Gamma(1-\beta)}{\beta-1} F_2 \left( \frac{1-\beta}{2}, \frac{3-\beta}{2}; -\pi^2 f_{0, \text{L}}^2 t^2 \right) + T^{-\beta-1} (2\pi)^{-\beta} \Gamma(1-\beta) \sin \left( \frac{\pi\beta}{2} \right) |t|^{\beta-1}.
\]
(\ref{eq:powerlaw_C}
where \(\tau\) is the dimensionless time in units of \(T\), \(F_2(\cdot)\) is the generalized hypergeometric function (see also van Haasteren et al. 2011 for the series presentation) and \(\Gamma(\cdot)\) is the Gamma function.

Integrating equation (A8), one reads
\[
\sigma^2 = S_0 T^{-\beta-1 - 2} (1+n) \left[ \sum_{k=n+1}^{\infty} \frac{(2\pi)^{2k}(1+n) \Gamma(k) \sin(\pi \beta)}{(1+2k-\beta) \Gamma(2k-\beta) \Gamma(k-n) \Gamma(2k+n)} + 2^{\beta-2} (1+n) \beta^{-1} \frac{\Gamma(\frac{3}{2}n-\beta) \Gamma(\frac{1}{2}+\beta) \Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}+2n-\beta) \Gamma(1+\beta)} \right].
\]
(A10)

Clearly, if \(2n+3 \geq \beta\), we have
\[
\lim_{f_{0, \text{L}} \rightarrow 0} \sigma^2 = S_0 T^{-\beta-1-2} (1+n) \beta^{-1} \frac{\Gamma(\frac{3}{2}n-\beta) \Gamma(\frac{1}{2}+\beta) \Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}+2n-\beta) \Gamma(1+\beta)},
\]
(A11)
i.e. the RMS \(\sigma^2\) is regularized to be a finite value for \(f_{0, \text{L}} \rightarrow 0\), which confirms our previous estimation. Usually pulsar spin and spin down are subtracted from the data. This corresponds to the case of \(n = 2\). For GW-induced signal, we have \(\beta = 13/3\) and \(S_0 = h^2(\gamma r^{-1})(12\pi^2)^{-1} \gamma^{-4/3}\), which gives
\[
\sigma^2 = h^2(\gamma r^{-1}) \frac{3^3(2\pi)^{10/3} \Gamma(\frac{3}{2})}{2 \cdot 5 \cdot 7 \cdot 11 \cdot 13} \gamma^{-4/3}.
\]
(A12)

This is identical to the results of van Haasteren & Levin (2012).

It should also be noted that for such a case (\(\beta = 13/3\) and \(S_0 = h^2(\gamma r^{-1})(12\pi^2)^{-1} \gamma^{-4/3}\) equation (A4) becomes
\[
\sigma^2 \gtrsim 2.55 \times 10^{-3} h^2 T^{10/3} \gamma^{4/3}.
\]
(A13)

In this way, estimating the RMS value of the fitted signal using equation (A4) is accurate enough for practical purposes, given data length is adopted as the effective cut-off frequency, i.e. to use \(f_{0, \text{L}} = 1\).

**APPENDIX B: GW-DETECTION SIGNIFICANCE**

In this section, we calculate the GW-detection significance. The expected value for the detection significance \(\langle \Sigma_c \rangle\) depends on \(\Sigma_c\) and \(\langle \gamma \rangle\) as shown in equation (12). We determine the \(\langle \gamma \rangle\) first. From equation (7), we have
\[
\langle \gamma \rangle = \frac{1}{m} \sum_{k=1}^{m} \langle R_k R_k \rangle = \sigma_q^2 H(\langle \gamma \rangle).
\]
(B1)

To determine \(\Sigma_c\), we need \(\langle \gamma^2 \rangle\), which is calculated in a similar fashion such that
\[
\langle \gamma^2 \rangle = \frac{1}{m^2} \sum_{k=1}^{m} \sum_{k=1}^{m} \langle R_k R_k R_k R_k \rangle.
\]
(B2)

After using the correlation relation (3), equation (6) and performing the Wick expansion (Zee 2010) to calculate higher momentum, we have
\[
\left( \frac{1}{m^2} \sum_{k=1}^{m} \sum_{k=1}^{m} \langle R_k R_k R_k R_k \rangle \right) = \sigma_q^4 \left( \langle I A \rangle + \langle I B \rangle + H(\langle \theta \rangle)^2 \right).
\]
(B3)

The \(\Sigma_c\) is then
\[
\Sigma_c = \sqrt{\frac{1}{M} \sum_{i-j \text{ pairs}} \left( \langle \gamma \rangle^2 - \langle \gamma \rangle^2 \right)} = \sigma_q^2 \sqrt{\Sigma_0^2 + \frac{1}{M} \sum_{i-j \text{ pairs}} \langle \gamma \rangle^2 + \langle \gamma \rangle^2}.
\]
(B4)

with which one can derive equation (15).

**APPENDIX C: OPTIMIZATION USING LAGRANGIAN MULTIPLIER**

In Section 3, we described the optimization technique using variable transformation and numerical optimization. In this section, we introduce another method of solving the constrained optimization problem directly. As an example, we present the technique for the single-telescope situation, which is also readily generalized.

The optimization problem for the single-telescope case is to search the minimal value of \(L = \sum_{i-j \text{ pairs}} \langle \gamma \rangle \) under the constraint that \(\tau = \sum_{i=1}^{n} \tau_i\). Using a Lagrangian multiplier \(\lambda\), we can recast the optimization problem to optimize \(L'\), where
\[
L' = L + \lambda \left( \tau - \sum_{i=1}^{n} \tau_i \right)
\]
(C1)

\[
= \sum_{i-j \text{ pairs}} \frac{k i q}{\tau_i} \frac{k i q}{\tau_i} + \frac{i q}{\tau_i} + \lambda \left( \tau - \sum_{i=1}^{n} \tau_i \right).
\]
(C2)

where \(\langle \gamma \rangle = (\sigma_q^2 G^{-2} + \langle \sigma_q^2 \rangle) e^{-2}\) and \(k = 1 + \sigma_q^2 \langle \gamma \rangle\). The minimization of \(L'\) can be found by
\[
\frac{\partial L'}{\partial \tau_i} = 0
\]
(C3)

and
\[
\frac{\partial L'}{\partial \lambda} = 0,
\]
(C4)

which give
\[
\frac{q}{\tau_i} \sum_{j=1}^{n} \frac{k' q}{\tau_j} + \frac{k q}{\tau_i} + \sum_{j=1}^{n} \frac{k q}{\tau_j} = \frac{\lambda}{2} = 0
\]
(C5)

and
\[
\sum_{i=1}^{n} \tau_i = 0.
\]
(C6)

It is easy to check that equations (C5) and (C6) can be solved using the following steps.

(1) Guess an initial value for \(\tau\).
(2) Update $\tau$ with a new value using
\[ \tau \to \tau = \frac{\sqrt{q} \sum_{j=1}^{n} (k + q/\tau)}{\sum_{i=1}^{n} \sqrt{q} \sum_{j=1}^{n} (k + q/\tau)} \to \tau, \tag{C7} \]

(3) Repeat step 2, until the required precision is achieved.

One can monitor the change of $\tau$ for each iteration until it converges to the necessary precision. The initial value for the iteration is determined from the strong-signal limit, i.e. $\tau \to 0$, where the iteration equation (C7) reduces to a solution
\[ \tau = \frac{\sqrt{q} \sum_{j=1}^{n} (k)}{\sum_{i=1}^{n} \sqrt{q} \sum_{j=1}^{n} (k)} + O(q). \tag{C8} \]

It is worth noting that, in fact, the iteration process (equation C7) will not change the results very much from the initial value equation (C8). This is due to
\[ \tau \to \tau = \frac{\sqrt{q} \sum_{j=1}^{n} (k)}{\sum_{i=1}^{n} \sqrt{q} \sum_{j=1}^{n} (k)} \to \tau. \tag{C9} \]

where $\tau = \sigma^2 G^{-2} + \sigma^2$. Clearly, the initial value equation (C8) we use is already a good approximation to the optimal observation strategy.

**APPENDIX D: THE CONDITION OF EXISTING SOLUTIONS TO $P_v$**

In this section, we investigate the conditions to be satisfied such that the following equations have solution $P_v$, given $\tau_v$, $G_v$, $\delta O_v$, and $\tau_v$.

\[
\begin{align*}
\tau_v &= \sum_{i=1}^{N} \tau_i \delta O_v, \tag{D1} \\
\tau_v &= \sum_{i=1}^{N} \tau_i \delta O_v, \tag{D2} \\
P_v &\geq 0, \tag{D3}
\end{align*}
\]

where $i = 1, \ldots, N$, $\nu = 1, \ldots, N_{\nu}$ and $\nu$ satisfies equation (28).

We want to prove that equations (D1), (D2) and (D3) have solutions $P_v$, if and only if, for any $i$, the following condition is satisfied,
\[ \tau_v \leq \sum_{i=1}^{N_{\nu}} G_{\nu}^2 \tau_i \delta O_v. \tag{D4} \]

The proof contains two steps as follows: (i) the ‘only if’ part. If $P_v$ is the solution, then
\[ \tau_v = \sum_{i=1}^{N_{\nu}} G_{\nu}^2 \tau_i \delta O_v \leq \sum_{i=1}^{N_{\nu}} G_{\nu}^2 \tau_i \delta O_v, \tag{D5} \]

where the second step is due to the constraint of equations (D2) and (D3). (ii) The ‘if’ part. It is easy to note that, under the constraints of equation (D2), any linear combination of the column vectors, e.g., $\beta_i = \sum_{j=1}^{N_\nu} \nu_{ij} \beta_j \delta O_j$, belongs to a hyperplane $S$, because $\sum_{i=1}^{N} \beta_i = \sum_{i=1}^{N} \nu_i \tau_i \delta O_v$ in $S$. Due to equation (D3), only part of the hyperplane $S$ is accessible to the vector $\beta$, i.e. $0 \leq \beta_i \leq \sum_{j=1}^{N_\nu} c_j \nu_{ij} = S$. Clearly, if the vector $\tau_v$ is in the accessible region $A$, there exists a solution to equation (D1) which satisfies both equations (D2) and (D3). From equation (28), we know $0 \leq \tau_v \leq \sum_{i=1}^{N} G_{\nu}^2 \tau_i \delta O_v$; thus, the vector $\tau_v$ belongs to the accessible region $A$, and the solution exists.

A graphical illustration for the condition is given in Fig. D1, where we choose $O = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$.

**APPENDIX E: COMMON NOISE MITIGATION**

Clock errors and other common noise sources are harmful to PTA observations. Tinto (2011) has shown that a simultaneous differential measurement of a pulsar TOA completely removes the clock error. Instead of the requirement for simultaneous observations of multiple pulsars, we argue in this section that one can still remove most of the clock errors in post-processing without simultaneity, if each observation session is compact enough.

Suppose the pulsar timing signal contains an identical (clock) noise $n_c(t)$ with a red spectrum. A subtraction between the timing signals of two different pulsars at two close epochs will remove most power of the identical noise component. One can check this by looking at the residuals $[i.e., n_c(t) - n_c(t + \Delta)]$ of the identical noise after the subtraction. It is easy to show that the power spectrum of the residual $S_p(f)$ becomes
\[ S_p(f) = S_n(f) [1 - \cos(2\pi f \Delta)], \tag{E1} \]

where $S_n(f)$ is the noise spectrum of $n_c$. $\Delta$ is the time difference between the two epochs and is thus roughly the time span of one observing session. The clock noise $S_p(f)$ dominates at low frequencies.
with time-scales of 10 yr; thus, we can check the residual at such frequencies, where the residual power spectrum becomes

\[ S_n(f) \simeq 1.5 \times 10^{-6} \left( \frac{f}{10\text{yr}^{-1}} \right)^2 \left( \frac{\Delta}{1\text{d}} \right)^2 S_c(f). \]  

(E2)

Thus, if we keep each observation session compact within a few days, the clock error can still be significantly (factor of $\sim 10^6$) suppressed by post-processing, and we do not need to worry about such common noise for planning an observing schedule at this stage.