

Chapter 2

On Measuring the Gravitational-Wave Background Using Pulsar Timing Arrays

Science may be described as the art of systematic over-simplification.

Karl Popper

2.1 Introduction

At the time of this writing several large projects are being pursued in order to directly detect astrophysical gravitational waves. This chapter concerns a program to detect gravitational waves using pulsars as nearly-perfect Einstein clocks. The practical idea is to time a set of millisecond pulsars (called the “Pulsar Timing Array”, or PTA) over a number of years (Foster and Backer 1990). Some of the millisecond pulsars create pulse trains of exceptional regularity. By perturbing the space-time between a pulsar and the Earth, the gravitational waves (GWs) will cause extra deviations from the periodicity in the pulse arrival times (Estabrook and Wahlquist 1975; Sazhin 1978; Detweiler 1979). Thus from the measurements of these deviations (called “timing residuals”, or TR), one may measure the gravitational waves. Currently, several PTA project are operating around the globe. Firstly, at the Arecibo Radio Telescope in North-America several millisecond pulsars have been timed for a number of years. These observations have already been used to place interesting upper limits on the intensity of gravitational waves which are passing through the Galaxy (Kaspi et al. 1994; Lommen 2001). Together with the Green Bank Telescope, the Arecibo Radio

This chapter is adapted from: *R. van Haasteren et al.*

On measuring the gravitational-wave background using Pulsar Timing Arrays

MNRAS (2009) 395(2): 1005–1014

By permission of Oxford University Press on behalf of the Royal Astronomical Society.

Telescope will be used as an instrument of NANOGrav, the North American PTA. Secondly, the European PTA is being set up as an international collaboration between Great Britain, France, Netherlands, Germany, and Italy, and will use 5 European radio telescopes to monitor about 20 ms pulsars (Stappers et al. 2006). Finally, the Parkes PTA in Australia has been using the Parkes multi-beam radio-telescope to monitor 20 ms pulsars (Manchester 2006). Some of the Parkes and Arecibo data have also been used to place the most stringent limits on the GWB to date (Jenet et al. 2006).

One of the main astrophysical targets of the PTAs is the stochastic background of the gravitational waves (GWB). This GWB is thought to be generated by a large number of black-hole binaries which are thought to be located at the centres of galaxies (Begelman et al. 1980; Phinney 2001; Jaffe and Backer 2003; Wyithe and Loeb 2003; Sesana et al. 2008), by relic gravitational waves (Grishchuk 2005), or, more speculatively, by cusps in the cosmic-string loops (Damour and Vilenkin 2005). This chapter develops an inference method for the optimal PTA measurement of such a GWB.

The main difficulty of such a measurement is that not only Gravitational Waves create the pulsar timing residuals. Irregularities of the pulsar-beam rotation (called the “timing noise”), the receiver noise, the imprecision of local clocks, the polarisation calibration of the telescope (Britton 2000), and the variation in the refractive index of the interstellar medium all contribute significantly to the timing residuals, making it a challenge to separate these noise sources from the gravitational-wave signal. However, the GWB is expected to induce correlations between the timing residuals of different pulsars. These correlations are of a specific functional form [given by Eq. (2.9) below], which is different from those introduced by other noise sources (Hellings and Downs 1983). Jenet et al. (2005, hereafter J05) have invented a clever method which uses the uniqueness of the GWB-induced correlations to separate the GWB from other noise sources, and thus to measure the magnitude of the GWB. Their idea was to measure the timing residual correlations for all pairs of the PTA pulsars, and check how these correlations depend on the sky-angles between the pulsar pairs. J05 have derived a statistic which is sensitive to the functional form of the GWB-induced correlation; by measuring the value of this statistic one can infer the strength of the GWB. While J05 method appears robust, we believe that in its current form it does have some drawbacks, in particular:

- (1) The statistic used by J05 is non-linear and non-quadratic in the pulsar-timing residuals, which makes its statistical properties non-transparent.
- (2) The pulsar pairs with the high and low intrinsic timing noise make equal contributions to the J05 statistic, which is clearly not optimal.
- (3) The J05 statistic assumes that the timing residuals of all the PTA pulsars are measured during each observing run, which is generally not the case.
- (4) The J05 signal-to-noise analysis relies on the prior knowledge of the intrinsic timing noise, and there is no clean way to separate this timing noise from the GWB.

- (5) The prior spectral information on GWB is used for whitening the signal; however, there is no proof that this is an optimal procedure. The spectral slope of the GWB is not measured.

In this chapter we develop an inference method which addresses all of the problems outlined above. Our method is based on essentially the same idea as that of J05: we use the unique character of the GWB-induced correlations to measure the intensity of the GWB. The method we develop below is Bayesian, and by construction uses optimally all of the available information. Moreover, it deals correctly and efficiently with all systematic contributions to the timing residuals which have a known functional form, i.e. the quadratic pulsar spindowns, annual variations, one-time discontinuities (jumps) due to equipment change, etc. Many parameters of the timing model (the model popular pulsar timing packages use to generate TRs from pulsar arrival times) fall in this category.

The plan of this chapter is as follows. In the next section we review the theory of the GWB-generated timing residuals and introduce our model for other contributions to the timing residuals. In Sect. 2.3 we explain the principle of Bayesian analysis for GWB-measurement with a PTA, and we evaluate the Bayesian likelihood function. There we also show how to analytically marginalise over the contributions of known functional form but unknown amplitude (i.e., annual variations, quadratic residuals due to pulsar spindown, etc.). The details of this calculation are laid out in Appendix A of this chapter. Section 2.4 discusses the numerical integration technique which we use in our likelihood analysis: the Markov Chain Monte Carlo (MCMC). In Sect. 2.5 we show the analyses of mock PTA datasets. For each mock dataset, we construct the probability distribution for the intensity of the GWB, and demonstrate its consistency with the input mock data parameters. We study the sensitivity of our inference method for different PTA configurations, and investigate the dependence of the signal-to-noise ratio on the duration of the experiment, on redness and magnitude of the pulsar timing noise, and on the number of clocked pulsars. In Sect. 2.6 we summarise our results.

2.2 The Theory of GW-Generated Timing Residuals

2.2.1 Timing Residual Correlation

The measured millisecond-pulsar timing residuals contain contributions from several stochastic and deterministic processes. The latter include the gradual deceleration of the pulsar spin, resulting in a pulsar rotational period derivative which induces timing residuals varying quadratically with time (hereafter referred to as “quadratic spindown”), the annual variations due to the imperfect knowledge of the pulsar positions on the sky, the ephemeris variations caused by the known planets in the solar system, and the jumps due to equipment change (Manchester 2006). The stochastic component of the timing residuals will be caused by the receiver noise,

clock noise, intrinsic timing noise, the refractive index fluctuations in the interstellar medium, and, most importantly for us, the GWB. For the purposes of this chapter we restrict ourselves to considering the quadratic spindowns, intrinsic timing noise, and the GWB; other components can be similarly included, but we omit them for mathematical simplicity. In this case, the i^{th} timing residual of the a^{th} pulsar can be written as

$$\delta t_{ai} = \delta t_{ai}^{\text{GW}} + \delta t_{ai}^{\text{PN}} + Q_a(t_{ai}), \quad (2.1)$$

where $\delta t_{ai}^{\text{GW}}$ and $\delta t_{ai}^{\text{PN}}$ are caused by the GWB and the pulsar timing noise, respectively, and

$$Q_a(t_{ai}) = A_{a1} + A_{a2}t_{ai} + A_{a3}t_{ai}^2 \quad (2.2)$$

represent the quadratic spindown. One expects the timing noise from different pulsars to be uncorrelated, while the GWB will cause correlations in the timing residuals between different pulsars. Therefore, the information about GWB can be extracted by correlating the timing residual data between the different pulsars (J05). If one assumes that both GWB-generated residuals and the intrinsic timing noise are random Gaussian processes, then we can represent them by the $(n \times n)$ covariance matrices:

$$\begin{aligned} \langle \delta t_{ai}^{\text{GW}} \delta t_{bj}^{\text{GW}} \rangle &= C_{(ai)(bj)}^{\text{GW}} \\ \langle \delta t_{ai}^{\text{PN}} \delta t_{bj}^{\text{PN}} \rangle &= C_{(ai)(bj)}^{\text{PN}}, \end{aligned} \quad (2.3)$$

with the total covariance matrix given by

$$C_{(ai)(bj)} = C_{(ai)(bj)}^{\text{GW}} + C_{(ai)(bj)}^{\text{PN}}. \quad (2.4)$$

The timing residuals are then distributed as a multidimensional Gaussian:

$$P(\vec{\delta t}) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp \left[-\frac{1}{2} \sum_{(ai)(bj)} (\vec{\delta t}_{(ai)} - Q_a(t_{ai})) C_{(ai)(bj)}^{-1} (\vec{\delta t}_{(bj)} - Q_b(t_{bj})) \right], \quad (2.5)$$

where P denotes the probability distribution of the timing residuals. To be able to use Eq. (2.5) we must

- (1) be able to evaluate the GWB-induced covariance matrix from the theory, as a function of variables that parametrise the GWB spectrum, and
- (2) introduce well-motivated parametrisation of the pulsar timing noise. In this chapter, the spectral density of the stochastic GW background is taken to be a power law (Phinney 2001; Jaffe and Backer 2003; Wyithe and Loeb 2003; Maggiore 2000)

$$S_h = A^2 \left(\frac{f}{\text{yr}^{-1}} \right)^{-\gamma}, \quad (2.6)$$

where S_h represents the spectral density, A is the GW amplitude, f is the GW frequency, and γ is an exponent characterising the GWB spectrum. If the GWB is dominated by the supermassive black hole binaries, then $\gamma = 7/3$ (Phinney 2001). This definition is equivalent to the use of the characteristic strain as defined in Jenet et al. (2006):

$$h_c = A \left(\frac{f}{\text{yr}^{-1}} \right)^\alpha, \quad (2.7)$$

with $\gamma = 1 - 2\alpha$. The GWB-induced covariance matrix is then given by

$$C_{(ai)(bj)}^{\text{GW}} = \frac{A^2 \alpha_{ab}}{(2\pi)^2 f_L^{1+\gamma}} \left\{ \Gamma(-1-\gamma) \sin\left(\frac{-\pi\gamma}{2}\right) (f_L \tau)^{\gamma+1} - \sum_{n=0}^{\infty} (-1)^n \frac{(f_L \tau)^{2n}}{(2n)! (2n-1-\gamma)} \right\}. \quad (2.8)$$

Here α_{ab} is the geometric factor given by

$$\alpha_{ab} = \frac{3}{2} \frac{1 - \cos \theta_{ab}}{2} \ln \left(\frac{1 - \cos \theta_{ab}}{2} \right) - \frac{1}{4} \frac{1 - \cos \theta_{ab}}{2} + \frac{1}{2} + \frac{1}{2} \delta_{ab}, \quad (2.9)$$

where θ_{ab} is the angle between pulsar a and pulsar b (Hellings and Downs 1983), $\tau = 2\pi (t_{ai} - t_{bj})$, Γ is the gamma function, and f_L is the low cut-off frequency, chosen so that $1/f_L$ is much greater than the duration of the PTA operation. Introducing f_L is a mathematical necessity, since otherwise the GWB-induced correlation function would diverge. However, we show below that the low-frequency part of the GWB is indistinguishable from an extra spindown of all pulsars which we already correct for, and that our results do not depend on the choice of f_L provided that $f_L \tau \ll 1$.

The pulsar timing noise is assumed to be Gaussian, with a certain functional form of the power spectrum. The true profile of the millisecond pulsar timing noise spectrum is not well-known at present time. The timing residuals of the most precisely observed pulsars indicate that pulsar timing noise has a white and poorly-constrained red component (J. Verbiest and G. Hobbs, private communications).

For the purposes of this chapter we will always choose the spectra to be of the same functional form for all pulsars, but this is not an inherent limitation of the inference method. We consider 3 cases of pulsar timing noise spectra:

- (1) White (flat) spectra
- (2) Lorentzian spectra
- (3) Power-law spectra

Obviously, one could also consider a timing noise which is a superposition of these components; we do not do this at this exploratory stage. If we choose the pulsar timing

noise spectrum to be white, with an amplitude N_a , the resulting correlation matrix becomes:

$$C_{(ai)(bj)}^{\text{PN-white}} = N_a^2 \delta_{ab} \delta_{ij}. \quad (2.10)$$

The Lorentzian spectrum is a red spectrum with a typical frequency that determines the redness of the timing noise:

$$S_a(f) = \frac{N_a^2}{f_0 \left(1 + \left(\frac{f}{f_0} \right)^2 \right)}, \quad (2.11)$$

which yields the following correlation matrix:

$$C_{(ai)(bj)}^{\text{PN-lor}} = N_a^2 \delta_{ab} \exp(-f_0 \tau), \quad (2.12)$$

where f_0 is a typical frequency and N_a is the amplitude.

By using a power law spectral density with amplitude N_a and spectral index γ_a , one gets a timing-noise covariance matrix analogous to the one in Eq. (2.8):

$$C_{(ai)(bj)}^{\text{PN-pl}} = \frac{N_a^2 \delta_{ab}}{f_L^{\gamma_a-1}} \left\{ \Gamma(1 - \gamma_a) \sin\left(\frac{\pi \gamma_a}{2}\right) (f_L \tau)^{\gamma_a-1} - \sum_{n=0}^{\infty} (-1)^n \frac{(f_L \tau)^{2n}}{(2n)! (2n + 1 - \gamma_a)} \right\}. \quad (2.13)$$

2.3 Bayesian Approach

2.3.1 Basic Ideas

The method described in this report is based upon a Bayesian approach to the parameter inference. The general idea of the method is to (a) assume that the physical processes which produce the timing residuals can be characterised by several parameters, and (b) use Bayes theorem to derive from the measured data the probability distribution of the parameters of our interest. In our case, we assume that the timing residuals are created by

- (1) The GWB; we parametrise it by its amplitude A and slope γ , as in Eq. (2.6).
- (2) The intrinsic timing noise of the 20 monitored millisecond pulsars. We assume that the timing noise of each of the pulsars is the random Gaussian noise, with a variety of possible spectra described in the previous section. We shall refer to the variables parametrizing the timing noise spectral shape as $T N_a$.

- (3) The quadratic spindowns, parametrised for each of the pulsars by A_{a1} , A_{a2} , and A_{a3} , cf. Eq. (2.2).

With these assumptions, we shall write down below the expression for the probability distribution $P(\text{data}|\text{parameters})$ of the data, as a function of the parameters. By Bayes theorem, we can then compute the posterior distribution function; the probability distribution of the parameters given a certain dataset:

$$P(\text{parameters}|\text{data}) = P(\text{data}|\text{parameters}) \times \frac{P_0(\text{parameters})}{P(\text{data})}. \quad (2.14)$$

Here $P_0(\text{parameters})$ is the prior probability of the unknown parameters, which represents all our current knowledge about these parameters, and $P(\text{data})$ is the marginal likelihood, which we will use here as a normalisation factor to ensure that $P(A, \gamma, TN_a, A_{a1}, A_{a2}, A_{a3}|\text{data})$ integrates to unity over the parameter space. We note here that the marginal likelihood is in essence a goodness of fit measure that can be used for model selection. However, we will ignore the marginal likelihood in this chapter and postpone the model selection part of the data analysis to future work. For our purposes, we are only interested in A and γ , which means that we have to integrate $P(A, \gamma, TN_a, A_{a1}, A_{a2}, A_{a3}|\text{data})$ over all of the other parameters. Luckily, as we show below, for a uniform prior the integration over A_{a1} , A_{a2} , and A_{a3} can be performed analytically. This amounts to the *removal* of the quadratic spindown component to the pulsar data. We emphasise that this removal technique is quite general, and can be readily applied to unwanted signal of any known functional form (i.e., annual modulations, jumps, etc.—see Sect. 2.3.2), even if those parameters have already been fit for while calculating the timing residuals. The integration over TN_a must be performed numerically.

In this chapter we shall use MCMC simulation as a multi dimensional integration technique. Besides flat priors for most of the parameters, we will use slightly peaked priors for parameters which have non-normalisable likelihood functions. This ensures that the Markov Chain can converge.

In the rest of the chapter, we detail the implementation and tests of our inference method.

2.3.2 Removal of Quadratic Spindown and Other Systematic Signals of Known Functional Form

While this section is written with the PTA in mind, it may well be useful for other applications in pulsar astronomy. We thus begin with a fairly general discussion, and then make it more specific for the PTA case.

Consider a random Gaussian process δx_i^G with a covariance matrix $C(\sigma)$, which is contaminated by several systematic signals with known functional forms $f_p(t_i)$

but a-priori unknown amplitudes ξ_p . Here σ is a set of interesting parameters which we want to determine from the data δx . The resulting signal is given by

$$\delta x_i = \delta x_i^G + \sum_p \xi_p f_p(t_i), \quad (2.15)$$

or, in the vector form, by

$$\vec{\delta x} = \vec{\delta x}^G + F \vec{\xi}. \quad (2.16)$$

Here the components of the vectors $\vec{\delta x}$, $\vec{\delta x}^G$, and $\vec{\xi}$ are given by δx_i , δx_i^G , and ξ_p , respectively, and F is the non-square matrix with the elements $F_{ip} = f_p(t_i)$. Note that the dimensions of $\vec{\delta x}$ and $\vec{\xi}$ are different. The Bayesian probability distribution for the parameters is given by

$$P(\sigma, \vec{\xi} | \vec{\delta x}) = \frac{M}{\sqrt{\det C}} \exp \left[-\frac{1}{2} (\vec{\delta x} - F \vec{\xi}) C^{-1} (\vec{\delta x} - F \vec{\xi}) \right] \times P_0(\sigma, \vec{\xi}), \quad (2.17)$$

where P_0 is the prior probability and M is the normalisation. Since we are only interested in σ , we can integrate $P(\sigma, \vec{\xi} | \vec{\delta x})$ over the variables $\vec{\xi}$. This process is referred to as marginalisation; it can be done analytically if we assume a flat prior for $\vec{\xi}$ [i.e., if $P_0(\sigma, \vec{\xi})$ is $\vec{\xi}$ -independent], since ξ_p enter at most quadratically into the exponential above. After some straightforward mathematics which we have detailed in Appendix A of this chapter, we get

$$P(\sigma | \vec{\delta x}) = \frac{M'}{\sqrt{\det(C) \det(F^T C^{-1} F)}} \exp \left[-\frac{1}{2} \vec{\delta x} \cdot C' \vec{\delta x} \right], \quad (2.18)$$

where M' is the normalisation, and

$$C' = C^{-1} - C^{-1} F (F^T C^{-1} F)^{-1} F^T C^{-1}, \quad (2.19)$$

and the T -superscript stands for the transposed matrix. Equation (2.18) is one of the main equations of this work, since it provides a statistically rigorous way to remove (i.e., marginalise over) the unwanted systematic signals from random Gaussian processes. One can check directly that the above expression for $P(\sigma | \vec{\delta x})$ is insensitive to the values ξ_p of the amplitudes of the systematic signals in the Eq. (2.15).

We now apply this formalism to account for the quadratic spindowns in the PTA. As in Sect. 2.2, it will be convenient to use the 2-index notation for the timing residuals, δt_{ai} measured at the time t_{ai} , where a is the pulsar index, and i is the number of the timing residual measurement for pulsar a . The space of the spindown parameters A_{aj} , $j = 1, 2, 3$ has $3N$ dimensions, where N is the number of pulsars in the array.

In component language, we write

$$\delta t_{ai} = \delta t_{ai}^G + \sum_{b,j} F_{(ai)(bj)} A_{bj}, \quad (2.20)$$

where

$$F_{(ai)(bj)} = \delta_{ab} t_{ai}^{j-1}, \quad (2.21)$$

δt^G is the part of the timing residual due to a random Gaussian process (i.e., GWB, timing noise, etc.), and $j = 1, 2, 3$. The quantities $F_{(ai)(bj)}$ are components of the matrix operator, called the design matrix, which acts on the $3N$ -dimensional vector in the parameter space and produces a vector in the timing residual space. For example, for 20 pulsars, each with 250 timing residual observations, the matrix $F_{(ai)(bj)}$ has $20 \times 250 = 5000$ rows, each marked by 2 indices $a = 1, \dots, 20$, $i = 1, \dots, 250$, and $20 \times 3 = 60$ columns, each marked by 2 indices $b = 1, \dots, 20$, $j = 1, 2, 3$. Thus in the vector form, one can write Eq. (2.20) as

$$\vec{\delta t} = \vec{\delta t}^G + F \vec{A}, \quad (2.22)$$

which is identical to the Eq. (2.16). We thus can use Eq. (2.18) to remove the quadratic spindown contribution from the PTA data.

Although we only demonstrate this technique for quadratic spindown, this removal technique will be useful for treating other noise sources in the PTA. All sources of which the functional form is known (and therefore can be fit for, as most popular pulsar timing packages do) can be dealt with, i.e.

- (1) Annual variation of the timing residuals due to the imprecise knowledge of the pulsar position on the sky. The annual variation in each of the pulsars will be a predictable function of the associated 2 small angular errors (latitude and longitude). Thus our parameter space will expand by $2N$, but this will still keep the F matrix manageable.
- (2) Changes of equipment will introduce extra jumps, and must be taken into account. This is trivial to deal with using the techniques described above.
- (3) Some of the millisecond pulsars are in binaries, and their orbital motion must be subtracted. The errors one makes in these subtractions will affect the timing residuals. They can be parametrised and dealt with using the techniques of this section (we thank Jason Hessels for pointing this out).

2.3.3 Low-Frequency Cut-Off

All predictions for GWB spectrum show a steep power law $\propto f^{-\gamma}$, where for black-hole binaries $\gamma = 7/3$ (Phinney 2001). Physically, there is a low-frequency cut-off to the spectrum, due to the fact that black-hole binaries with periods greater than 1000

years shrink mostly due to the external friction (i.e., scattering of circum-binary stars or excitation of density waves in a circum-binary gas disc), and not to gravitational radiation. However, while the exact value of the low-frequency cut-off is poorly constrained, the PTA should not be sensitive to it since the duration of the currently planned experiments is much shorter than 1000 years. In this section, we show this formally by explicitly introducing the low-frequency cut-off and by demonstrating that our Bayesian probabilities are insensitive to its value.

Consider the expression in Eq. (2.8) for the GWB-generated correlation matrix for the timing residuals. This expression contains an integral of the form

$$I = \int_{f_L}^{\infty} \cos(f\tau) f^{-(\gamma+2)} df, \quad (2.23)$$

where $\tau = 2\pi(t_i - t_j)$. When the low-frequency cut-off is much smaller than the inverse of the experiment duration, i.e. when $f_L\tau \ll 1$, the integral above can be expanded as

$$I = B\tau^{\gamma+1} + \frac{1}{f_L^{\gamma+1}} \left\{ \frac{1}{(\gamma+1)} - \frac{(f_L\tau)^2}{2(\gamma-1)} + \mathcal{O}[(f_L\tau)^4] \right\}, \quad (2.24)$$

where

$$B = \Gamma(-1-\gamma) \sin\left(\frac{-\pi\gamma}{2}\right) \tau^{\gamma+1}. \quad (2.25)$$

In the expansion above we have assumed $1 < \gamma < 3$. The terms which contain f_L diverge when f_L goes to zero, and scale as τ^0 or τ^2 with respect to the time interval. We now show that these divergent terms get absorbed in the process of elimination of the quadratic spindowns.

Suppose that we add to the timing residuals of a pulsar a quadratic spindown term, $A_1 + A_2 t + A_3 t^2$. The spindown-removal procedure described in the previous section makes our results completely insensitive to this addition: A 's could be arbitrarily large but the measured GWB would still be the same. Clearly, the same is true if one treats A_1, A_2, A_3 not as fixed numbers, but as random variables drawn from some Gaussian distribution. The correlation introduced into the timing residuals by adding a *random* quadratic spindown is given by

$$\begin{aligned} \langle \delta t_i \delta t_j \rangle &= \langle A_1^2 \rangle + \langle A_1 A_2 \rangle (t_i + t_j) \\ &\quad + 2\langle A_2^2 \rangle t_i t_j + \langle A_1 A_3 \rangle (t_i^2 + t_j^2) \\ &\quad + \langle A_2 A_3 \rangle t_i t_j (t_i + t_j) + \langle A_3^2 \rangle t_i^2 t_j^2. \end{aligned} \quad (2.26)$$

The f_L -dependent part of Eq. (2.24) contains terms which scale as $t_i^2 + t_j^2$, $t_i t_j$, and ~ 1 , and thus have the same functional t_i, t_j dependence as some of the terms in Eq. (2.26). Since the terms in Eq. (2.26) can be made arbitrarily large, it is clear

that the terms corresponding to the low-frequency cutoff could be absorbed into the correlation function corresponding to the quadratic spindown with the stochastic coefficients. We have made this argument for the timing residuals from a single pulsar, but it is trivial to extend it to the case of multiple pulsars. Thus our results are not sensitive to the actual choice of the f_L so long as $f_L \tau \ll 1$; this is confirmed by direct numerical tests.

2.4 Numerical Integration Techniques

2.4.1 Metropolis Monte Carlo

The Bayesian probability distribution for the PTA is computed in multi dimensional parameter space, where all of the parameters except two characterise intrinsic pulsar timing noise and other potential interferences. To obtain meaningful information about the GWB, we need to integrate the probability distribution over all of the unwanted parameters. This is a challenging numerical task: a direct numerical integration over more than several parameters is prohibitively computationally expensive. Fortunately, numerical shortcuts do exist, and the most common among them is the Markov Chain Monte Carlo (MCMC) simulation. In a typical MCMC, a set of semi-random walkers sample the parameter space in a clever way, each generating a large number of sequential locations called a *chain* (Newman and Barkema 1999). After a sufficient number of steps, the density of points of the chain becomes proportional to the Bayesian posterior distribution. We found that for our purposes, the number of steps required for the chain convergence scales linearly with the number of dimensions of the parameter space; typically $\text{few} \times 10^4$ steps are required for reliable convergence. In this work we use the Metropolis (Newman and Barkema 1999) computational scheme for generating the chain, which can be used with an arbitrary distribution, the proposal distribution, for generating new locations of the chain. We use a Gaussian proposal distribution, centred at the current location in the parameter space. During an initial period, the burn-in period, the width of the proposal distribution in all dimensional directions is set to yield the asymptotically optimal acceptance rate of 23.4 % for Metropolis (Roberts et al. 1997). At the end of the MCMC simulation we check the convergence of the chain using a batched mean method (Efron 1979). We also calculate the global maximum likelihood value for all parameters using a conjugate directions search (Brent 1973). By carefully inspecting the marginalised posterior distributions we found that the posterior distribution only has a single peak for all the parameters and parametrisations we have considered so far. We therefore assume that we are not hindered by chains getting stuck in local maxima when assessing convergence.

2.4.2 Current MCMC Computational Cost

The greatest computational challenge in constructing the chain is the fast evaluation of the matrix C^{-1} in Eqs. (2.18) and (2.19). If 250 timing residuals are measured for each of the pulsars (50 weeks for 5 years), the size of the matrix C becomes (5000×5000) . We find it takes about 20 s to invert C and thus about 1.5 times as much to arrive to the next point in the chain. Therefore, for the required 10^5 chain points to get the convergent distribution, we need of order 1 month of the single-processor computational time. On a cluster this can be done in a couple of days. We emphasise that this is an order n^3 process. For matrices of (2000×2000) the calculation can be done overnight on a single modern workstation, but for $(10^4 \times 10^4)$ the calculation is already a serious challenge.

For the currently projected size of the datasets (Manchester 2006), the amount of timing residuals will most likely not exceed the 250 (Hobbs, private communications). Thus, the brute-force method presented here is not computationally expensive for the projected data volume over the next 5 years.

2.4.3 Choosing a Suitable Prior Distribution

For some models (e.g. the power law spectral density for pulsar timing noise) the likelihood function proves to be not normalisable. This would pose a serious problem in combination with uniform priors as the nuisance parameters then cannot be marginalised and the posterior cannot represent a probability distribution. This is a sign that our prior is improper. The prior distribution reflects our a priori knowledge of the model parameters, and since timing noise is currently not well enough understood to properly parametrise it, the natural choice of prior would be an uninformative one. However, a uniform prior is not integrable over all possible values, which makes the posterior distribution not normalisable. Studies by Verbiest et al. (2009) (see also Chap. 4 of this thesis) have suggested that the timing noise that is dominant in MSPs at current sensitivity levels is less “red” than the GWB. We therefore propose to use an integrable prior distribution in the form of a Lorentzian shaped profile, much broader than the likelihood distribution in the spectral index parameter:

$$P_0(\gamma_i) = \frac{\Delta_i}{\pi (\Delta_i^2 + \gamma_i^2)}, \quad (2.27)$$

where γ_i is the parameter for which we are construction a prior, and Δ_i is some typical width/value for this parameter.

As an example we show the likelihood function and the prior for the pulsar timing noise spectral index parameter of Eq. (2.13) in Fig. 2.1. The likelihood function seems to drop to zero for high γ_i , but it actually has a non-negligible value for all γ_i greater than the maximum likelihood value. The broadness of the prior is chosen such that

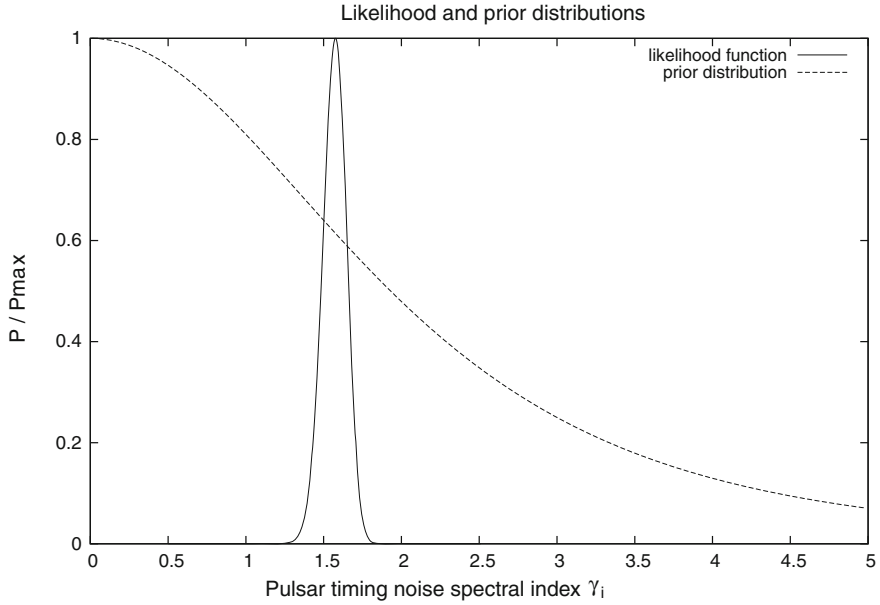


Fig. 2.1 The likelihood and prior distribution for a pulsar timing noise spectral index parameters γ_i . The *solid line* represents the likelihood function. It is sharply peaked and it looks as if it drops to zero for high γ_i . However, for high γ_i it will have a constant non-negligible value. The *dashed line* represents our chosen prior distribution. The prior is normalisable, and its application makes the posterior distribution normalisable as well. Figure from the MNRAS article van Haasteren et al. (2009)

it does not change the representation of the significant part of the likelihood in the posterior, but it does make sure that the posterior is normalisable.

2.4.4 Generating Mock Data

In order to generate mock data, we produce a realisation of the multi dimensional Gaussian process of Eq. (2.5), as follows. We rewrite Eq. (2.5) in a basis in which C is diagonal:

$$P(\vec{\delta}t) = \prod_{i=1}^n \frac{1}{\sqrt{\lambda_i}} \varphi\left(\frac{y_i}{\sqrt{\lambda_i}}\right), \quad (2.28)$$

where,

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (2.29)$$

Here λ_i are the eigenvalues of C , and

$$\vec{y} = T^{-1} \vec{\delta t}, \quad (2.30)$$

where T is the transformation matrix which diagonalises C :

$$(T^{-1}CT)_{ij} = \lambda_i \delta_{ij}. \quad (2.31)$$

Thus we follow the following steps:

- (1) Diagonalise matrix C , find T and λ_i .
- (2) Choose y_i from random Gaussian distributions of widths $\sqrt{\lambda_i}$.
- (3) Compute the timing residuals via Eq. (2.30).

It is then trivial to add deterministic processes, like quadratic spindown, to the simulated timing residuals.

2.5 Tests and Parameter Studies

We test our inference method by generating mock timing residuals for a number of millisecond pulsars which are positioned randomly in the sky. We found it convenient to parametrise the GWB spectrum by [cf. Eq. (2.6)]

$$S_h(f) = A^2 \left(\frac{f}{\text{yr}^{-1}} \right)^{-\gamma}. \quad (2.32)$$

Our mock timing residuals are a single realisation of GWB for some values of A and γ and the pulsar timing noise. Random quadratic spindown terms are added. We then perform several separate investigations as follows:

2.5.1 Single Dataset Tests

Our inference method is tested on several datasets in the following way:

The mock datasets were generated with parameters resembling an experiment of 20 pulsars, with observations approximately every 5 weeks for 5 years. The pulsar timing noise was set to an optimistic level of 100 ns each (rms timing residuals). In all cases the level of GWB has been set to $A = 10^{-15} \text{yr}^{1/2}$, with $\gamma = 7/3$. This level of GWB is an order of magnitude smaller than the most recent upper limits of this type (Jenet et al. 2006). We then analyse this mock data using the MCMC method. In Figs. 2.2, 2.3, and 2.4 we see examples of the joint A — γ probability

distribution, obtained by these analyses. For each dataset we also calculate the maximum likelihood value of all parameters using a conjugate directions search. The Bayesian inference method gives results consistent with the input parameters (i.e., they recover the amplitude and the slope of the GWB within measurement errors). This was observed in all our tests.

For all datasets we also calculated the Fisher information matrix, a matrix consisting of second-order derivatives to all parameters, at the maximum likelihood points. We can use this matrix to approximate the posterior by a multi dimensional Gaussian, since for some particular models this approximation is quite good. The Fisher information matrix can be calculated in a fraction of the time needed to perform a full MCMC analysis. For all datasets we have plotted the 1σ contour of the multi dimensional Gaussian approximation.

As an extra test, we have also used datasets generated by the popular pulsar timing package tempo2 (Hobbs et al. 2006) with a suitable GWB simulation plug-in (Hobbs et al. 2009). We were able to generate datasets with exactly the same parameters as with our own inference method, provided that the timing noise was white. We have

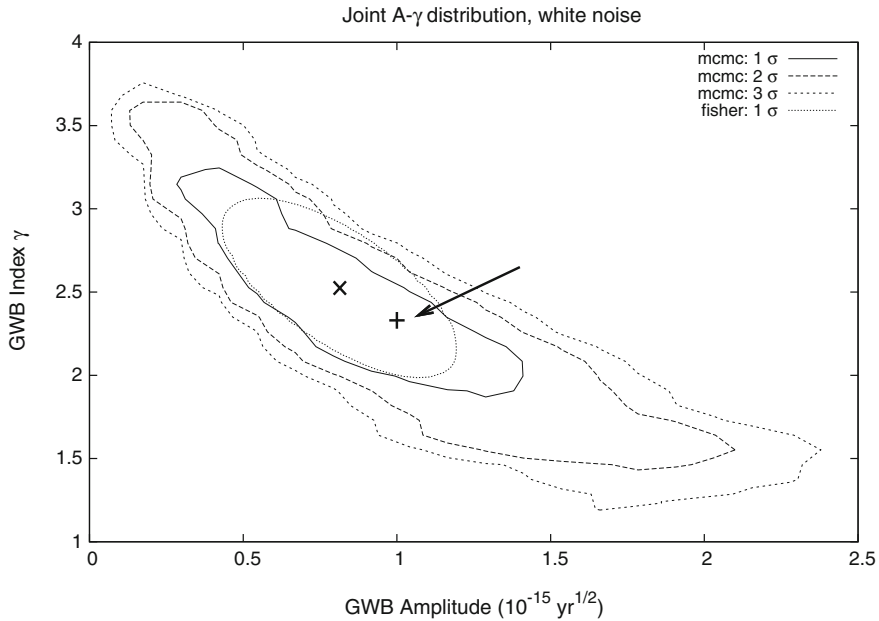


Fig. 2.2 The GW likelihood function (GW amplitude, GW slope vs. prob. density contours), determined with the MCMC method for a set of mock data with 20 pulsars, and 100 data points per pulsar approximately evenly distributed over 5 years. Each pulsar has a white timing noise of 100 ns. The true GW amplitude and slope are shown as a “+” with an *arrow*, and the maximum likelihood values are shown as “x”. The contours are in steps of σ , with the inner one at 1σ . The 1σ contour of the Gaussian approximation is also shown. Figure from the MNRAS article van Haasteren et al. (2009)

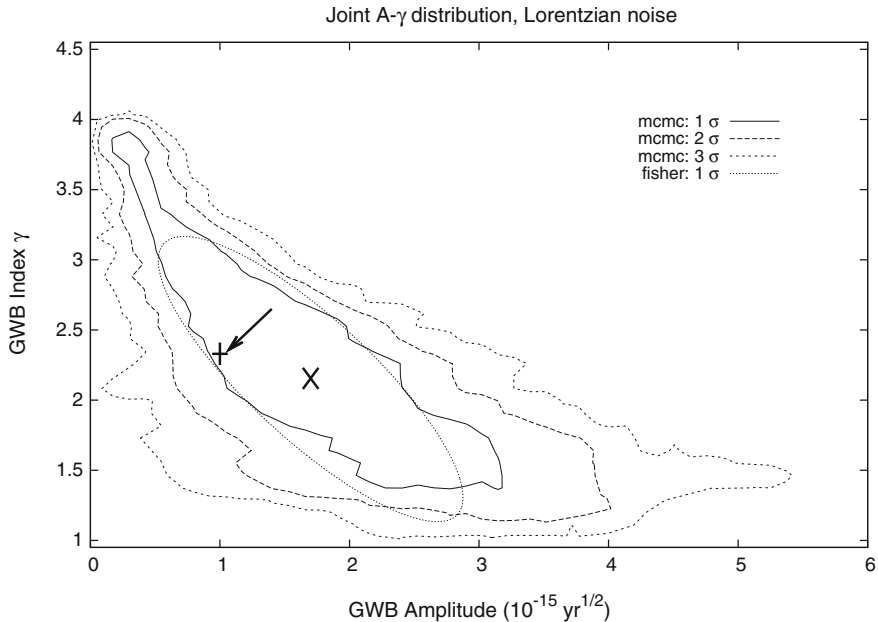


Fig. 2.3 Same as in Fig. 2.2, but the mock data is generated and analysed using Lorentzian timing noise. Overall timing noise amplitude and characteristic frequency f_0 are taken to be 100 ns and 1yr^{-1} for each pulsar. Figure from the MNRAS article van Haasteren et al. (2009)

confirmed that those datasets yield similar results when analysed with our inference method.

An important point is that the spectral form of the timing noise has a large impact on the detectability of the GWB. For a red Lorentzian pulsar timing noise there is far greater degeneracy between the spectral slope and amplitude in the timing residual data for the GWB than for white pulsar timing noise, and thus the overall signal-to-noise ratio is significantly reduced by the red component of the timing noise.

2.5.2 Multiple Datasets, Same Input Parameters

To estimate the robustness of our inference method, we also perform a maximum likelihood search on many datasets:

- (a) We generate a multitude of mock timing residual data for the same PTA configurations as in Sect. 2.5.1, with white timing noise.
- (b) For every one of these datasets we calculate the maximum likelihood parameters using the conjugate directions search. The ensemble of maximum likelihood

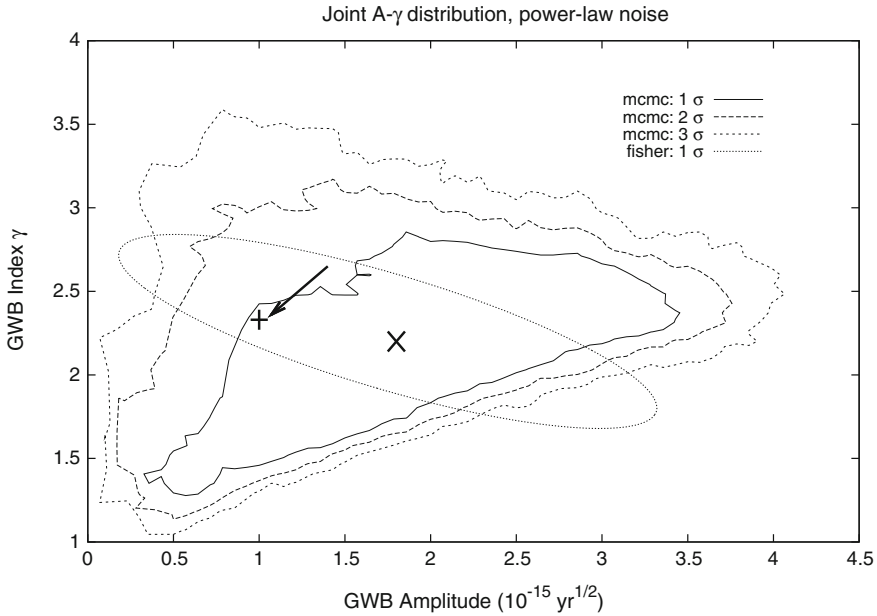


Fig. 2.4 Same as in Fig. 2.2, but the mock data is generated and analysed using power law timing noise. Overall timing noise amplitude and spectral index γ_i are taken to be 100 ns and 1.5 for each pulsar. For all γ_i , a prior distribution according to Eq. 2.27. Figure from the MNRAS article van Haasteren et al. (2009)

estimators for (A, γ) should be close to the true values used to generate the timing residuals.

The results of maximum likelihood search on many datasets is demonstrated in Fig. 2.5. The points are the maximum likelihood values for individual datasets. It can be seen that the points are distributed in a shape similar, but not identical, to Fig. 2.2: some points are quite far off from the input parameters. In order to test the validity of the results, we calculate the Fisher information matrix at the maximum likelihood points, and show the 1σ contour of the multidimensional Gaussian approximation based on the Fisher information matrix for three points. We see that the error contours do not exclude the true values at high confidence, even though the Fisher matrix does not yield a perfect representation of the error contours (the true posterior is not perfectly Gaussian), and we have a posteriori selected outliers for 2 of the 3 cases.

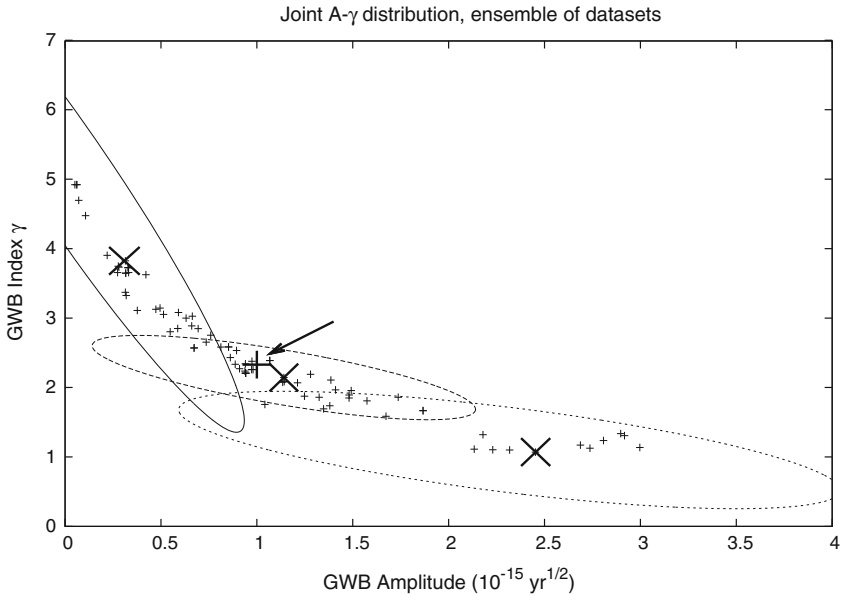


Fig. 2.5 The maximum likelihood values for an ensemble of realisations of mock datasets, all with the same model parameters: 100 ns white noise, 20 pulsars, and 100 data points per pulsar approximately evenly distributed over 5 years. The contours are confidence contours based on Fisher information matrix approximations of the likelihood function. Figure from the MNRAS article van Haasteren et al. (2009)

2.5.3 Parameter Studies

To test the accuracy of the Bayesian inference method, and to provide suggestions for optimal PTA configurations, we conduct some parameter studies on simplified sets of mock timing residuals:

- (a) We generate many sets of mock timing residuals for the simplified case of white pulsar timing noise spectra, all with the same white noise amplitude. The datasets are timing residuals for some number of millisecond pulsars which are positioned randomly in the sky. We parametrise the GWB by Eq. (2.32). We then generate many sets of timing residuals, varying several parameters [i.e., timing noise amplitude (assumed the same for all pulsars), duration of the experiment, and number of pulsars].
- (b) For each of the mock datasets we approximate the likelihood function by a Gaussian in the GWB amplitude A , with all other parameters fixed to their real value. We use A as a free parameter since it represents the strength of the GWB, and therefore the accuracy of A is a measure of the detectability. All other

parameters are fixed to keep the computational time low, but this does result in a higher signal to noise ratio than is obtainable with a full MCMC analysis.

- (c) For this Gaussian approximation, we calculate the ratio μ/σ as an estimate for the signal to noise ratio, where μ is the value of A at which the likelihood function maximises, and σ is the value of the standard deviation of the Gaussian approximation. Our results, represented as signal-to-noise contour plots for pairs of the input parameters, can be seen in Figs. 2.6, 2.7, 2.8, 2.9, 2.10, 2.11, and 2.12.

2.5.4 Comparison to Other Work

More than a decade ago, McHugh et al. (1996) used a Bayesian technique to produce upper limits on the GWB using pulsar timing. We found the presentation of this work rather difficult to follow. Nonetheless, it is clear that the analysis presented here is more general than that of McHugh et al.: we treat the whole pulsar array, and not just a single pulsar; we take into account the extreme redness of the noise and develop the formalism to treat the systematic errors like quadratic spindown.

Simultaneously with our work, a paper by Anholm et al. (2009) has appeared on the arxiv preprint service. Their approach was to construct a quadratic estimator, (dubbed

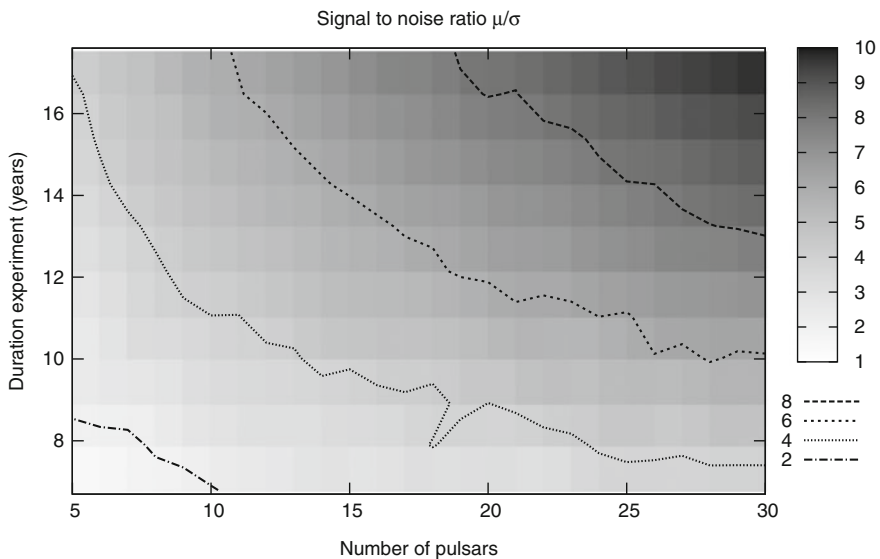


Fig. 2.6 Density plot of the signal to noise ratio μ/σ for different realisations of timing residuals. We have assumed monthly observations of pulsars with white timing noise of 100 ns each. The GWB amplitude has been set to $10^{-15} \text{ yr}^{1/2}$. Figure from the MNRAS article van Haasteren et al. (2009)

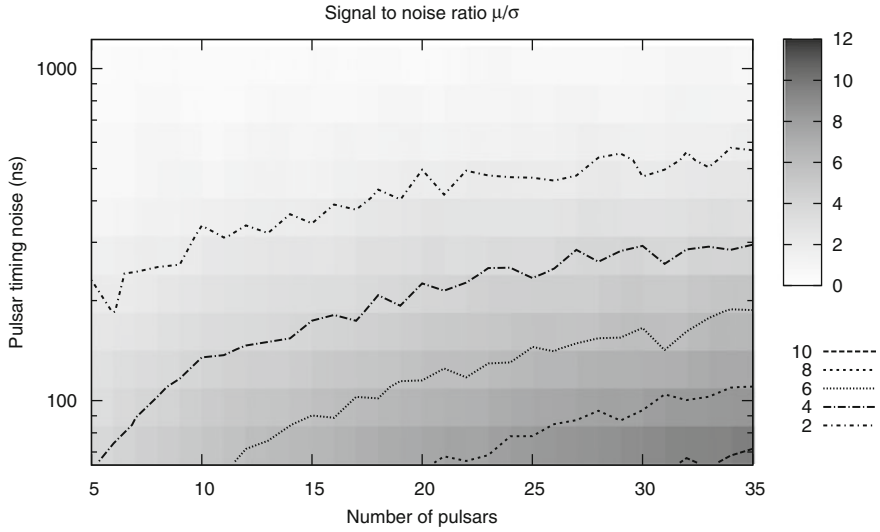


Fig. 2.7 Density plot of the signal to noise ratio μ/σ for different realisations of timing residuals. We have assumed 100 data points per pulsars, approximately evenly distributed over a period of 7.5 years. The GWB amplitude has been set to $10^{-15} \text{ yr}^{1/2}$. Figure from the MNRAS article van Haasteren et al. (2009)

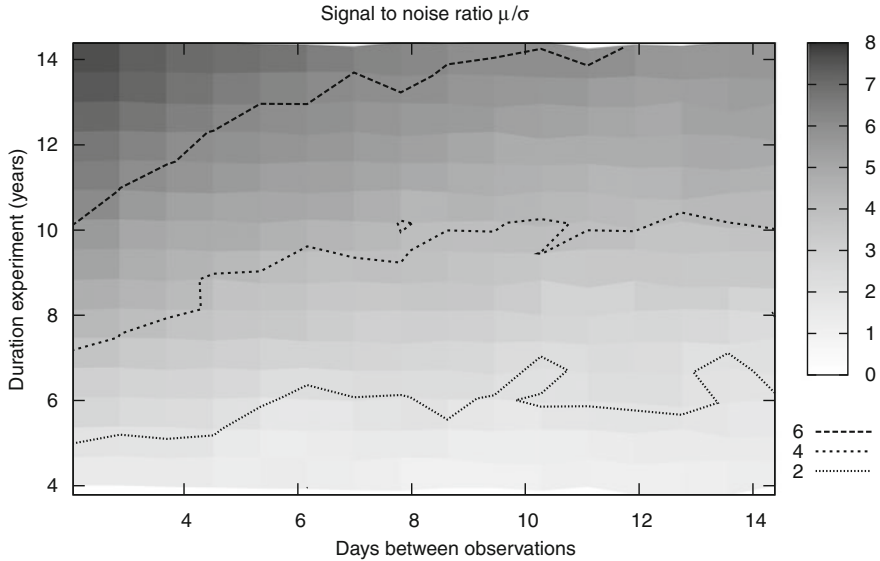


Fig. 2.8 Density plot of the signal to noise ratio μ/σ for different realisations of timing residuals. We have used a constant GWB amplitude of $10^{-15} \text{ yr}^{1/2}$ and 20 pulsars. Figure from the MNRAS article van Haasteren et al. (2009)

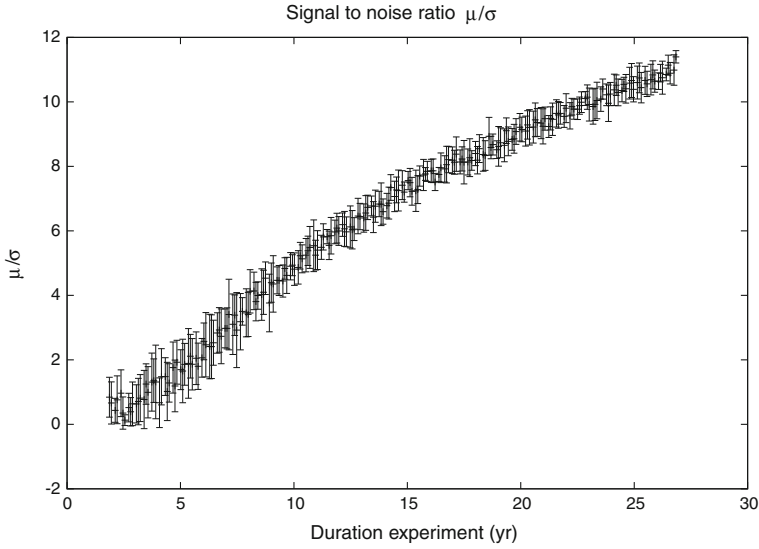


Fig. 2.9 Density plot of the signal to noise ratio $\frac{\mu}{\sigma}$ for different realisations of timing residuals. We have used 20 pulsars with white pulsar timing noise levels of 100 ns each, with monthly observations. The GWB amplitude has been set to $10^{-15} \text{ yr}^{1/2}$. The points and error bars are the mean and standard deviation of 10 realisations. Figure from the MNRAS article van Haasteren et al. (2009)

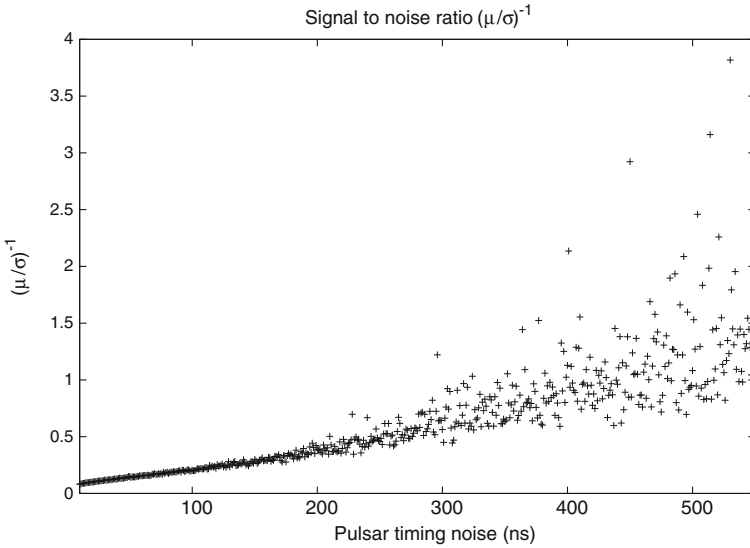


Fig. 2.10 Plot of one over the signal to noise ratio $(\mu/\sigma)^{-1}$ with respect to the pulsar timing noise for an experiment of 5 years, 20 pulsars and monthly observations. The GWB amplitude has been set to $10^{-15} \text{ yr}^{1/2}$. Figure from the MNRAS article van Haasteren et al. (2009)

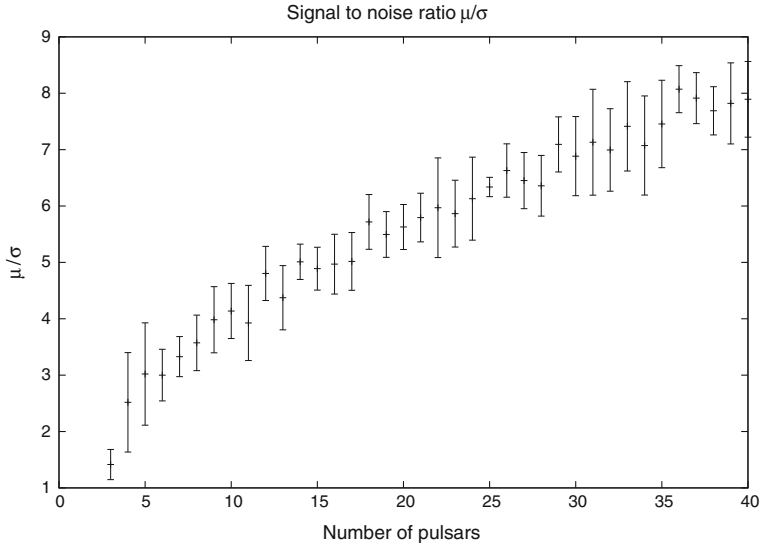


Fig. 2.11 Plot of the signal to noise ratio $\frac{\mu}{\sigma}$ with respect to the number of observed pulsars. The *white* timing noise of each pulsar has been set to 100 ns and the observations were taking every 2 months for a period of 7.5 years. The GWB amplitude has been set to $10^{-15} \text{ yr}^{1/2}$. The points and error bars are the mean and standard deviation of 10 realisations. Figure from the MNRAS article van Haasteren et al. (2009)

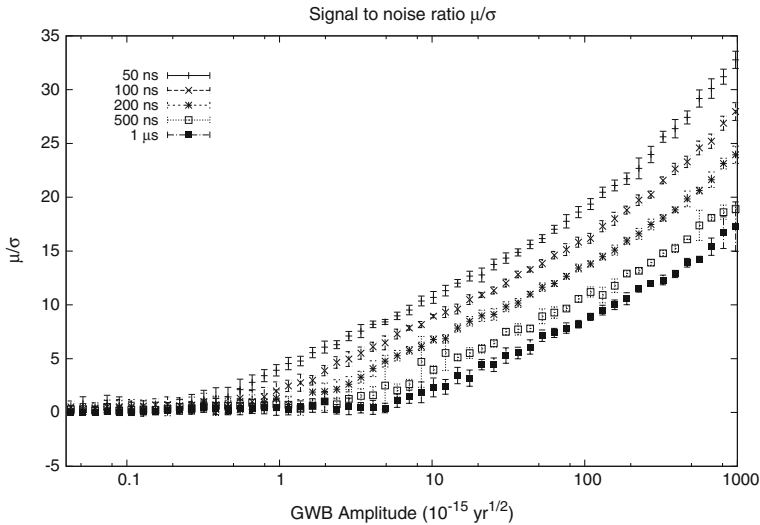


Fig. 2.12 Several plots of the signal to noise ratio $\frac{\mu}{\sigma}$ with respect to the level of the GWB amplitude. The number of pulsars was set at 20, with observations every two weeks for a period of 5 years. The pulsar noise levels were set at 50, 100, 200, 500, 1000 ns for the different plots. The points and error bars are the mean and standard deviation of 10 realisations. Figure from the MNRAS article van Haasteren et al. (2009)

the “optimal statistic”, written explicitly in the frequency domain), which aims to be optimally sensitive to the GWB. This improves on the original non-quadratic estimator of J05. However, a number of issues important for the pulsar timing experiment remained unaddressed, the most important among them the extreme redness of the GWB and the need to subtract consistently the quadratic spindown.

2.6 Conclusion

In this chapter we have introduced a practical Bayesian inference method for measuring the GWB using Pulsar Timing Arrays. Several attractive features of the inference method should make it useful to the PTA community:

- (1) the ability to simultaneously measure the amplitude and slope of GWB,
- (2) its ability to deal with unevenly sampled datasets, and
- (3) its ability to treat systematic contributions of known functional form. From a theoretical point of view, the inference method is guaranteed to extract information without loss of information, provided that our parametrisation of the timing noise is correct.

Test runs of our inference method have shown that the experiments signal-to-noise (S/N) ratio strongly decreases with the redness of the pulsar timing noise, and strongly increases with the duration of the PTA experiment. We have also charted the S/N dependence on the number of well-clocked pulsars and the level of their timing noise. These charts should be helpful in the design of the optimal strategy for future PTA observations.

Appendix A

In this Appendix we show explicitly how to perform marginalisation over the nuisance parameters $\vec{\xi}$ in Eq. (2.16), rewritten here for convenience:

$$P(\sigma, \vec{\xi} | \vec{\delta x}) = \frac{M}{\sqrt{\det C}} \exp \left[-\frac{1}{2} (\vec{\delta x} - F \vec{\xi}) C^{-1} (\vec{\delta x} - F \vec{\xi}) \right] P_0(\sigma, \vec{\xi}), \quad (2.33)$$

From here on we assume that P_0 is independent of $\vec{\xi}$ (a flat prior). All values are therefore equally likely for all elements of $\vec{\xi}$ prior to the observations. This assumption is also implicitly made in the frequentist approach when fitting for these kinds of parameters as is done in popular pulsar timing packages. We now perform the marginalisation:

$$P(\sigma | \vec{\delta x}) = \int P(\sigma, \vec{\xi} | \vec{\delta x}) d^m \xi, \quad (2.34)$$

where m is the dimensionality of $\vec{\xi}$. The idea now is to rewrite the exponent E of Eq. (2.33) in such a way that we can perform a Gaussian integral with respect to $\vec{\xi}$ (we have to get rid of the F in front of $\vec{\xi}$). Therefore, we will expand E and complete the square with respect to ξ :

$$\begin{aligned}
 E &= \left(\vec{\delta x} - F \vec{\xi} \right)^T C^{-1} \left(\vec{\delta x} - F \vec{\xi} \right) \\
 &= \vec{\delta x}^T C^{-1} \vec{\delta x} - 2 \vec{\xi}^T F^T C^{-1} \vec{\delta x} + \vec{\xi}^T F^T C^{-1} F \vec{\xi} \\
 &= \vec{\delta x}^T C^{-1} \vec{\delta x} + \left(\vec{\xi} - \vec{\chi} \right)^T F^T C^{-1} F \left(\vec{\xi} - \vec{\chi} \right) \\
 &\quad - \vec{\chi}^T F^T C^{-1} F \vec{\chi},
 \end{aligned} \tag{2.35}$$

where we have used the substitution:

$$\vec{\chi} = \left(F^T C^{-1} F \right)^{-1} F^T C^{-1} \vec{\delta x}. \tag{2.36}$$

Using this, we can write the $\vec{\xi}$ dependent part of the integral of Eq. (2.34) as a multi dimensional Gaussian integral:

$$\begin{aligned}
 I &= \int \exp \left(\frac{-1}{2} \left(\vec{\xi} - \vec{\chi} \right)^T F^T C^{-1} F \left(\vec{\xi} - \vec{\chi} \right) \right) d^m \xi \\
 &= (2\pi)^m \det \left(F^T C^{-1} F \right)^{-1}.
 \end{aligned} \tag{2.37}$$

From this it follows that:

$$P(\sigma | \vec{\delta x}) = \frac{M'}{\sqrt{\det(C) \det(F^T C^{-1} F)}} \exp \left[-\frac{1}{2} \vec{\delta x} \cdot C' \vec{\delta x} \right] \tag{2.38}$$

where we have absorbed all constant terms in the normalisation constant M' , and where we have used:

$$C' = C^{-1} - C^{-1} F (F^T C^{-1} F)^{-1} F^T C^{-1}. \tag{2.39}$$

References

- Anholm M., Ballmer S., Creighton J. D. E., Price L. R., Siemens X., 2009, *Phys. Rev. D*, 79, 8
- Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307
- Brent R., 1973, *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey
- Britton M. C., 2000, *ApJ*, 532, 1240
- Damour T., Vilenkin A., 2005, *Phys. Rev. D*, 71, 063510
- Detweiler S., 1979, *ApJ*, 234, 1100
- Efron B., 1979, *Ann. of Stat.*, 7, 1
- Estabrook F., Wahlquist H., 1975, *Gen. Relativ. Gravitation*, 6, 439
- Foster R., Backer D., 1990, *ApJ*, 361, 300
- Grishchuk L. P., 2005, *Uspekhi Fizicheskikh Nauk*, 48, 1235
- Hellings R., Downs G., 1983, *ApJ*, 265, L39
- Hobbs G. B., Edwards R. T., Manchester R. N., 2006, *MNRAS*, 369, 655
- Hobbs G. B., Bailes M., Bhat N. D. R., Burke-Spolaor S., Champion D. J., Coles W., Hotan A., Jenet F., et al. 2009, *Publications of the Astronomical Society of Australia*, 26, 103
- Jaffe A., Backer D., 2003, *ApJ*, 583, 616
- Jenet F., Hobbs G., Lee K., Manchester R., 2005, *ApJ*, 625, L123
- Jenet F., Hobbs G., van Straten W., Manchester R., Bailes M., Verbiest J., Edwards R., Hotan A., Sarkissian J., Ord S., 2006, *ApJ*, 653, 1571
- Kaspi V. M., Taylor J. H., Ryba M. F., 1994, *ApJ*, 428, 713
- Lommen A. N., 2001, PhD thesis, UC Berkeley
- Maggiore M., 2000, *Phys. Rep.*, 331, 283
- Manchester R. N., 2006, *Chinese Journal of Astronomy and Astrophysics Supplement*, 6, 139
- McHugh M. P., Zalamansky G., Vernotte F., Lantz E., 1996, *Phys. Rev. D*, 54, 5993
- Newman M., Barkema G., 1999, *Monte Carlo Methods in Statistical Physics*. Oxford University Press Inc., pp 31–86
- Phinney E. S., 2001, eprint arXiv:astro-ph/0108028v1
- Roberts G., Gelman A., Gilks W., 1997, *Ann. of Appl. Prob.*, 7, 110
- Sazhin M. V., 1978, *Soviet Astronomy*, 22, 36
- Sesana A., Vecchio A., Colacino C. N., 2008, *MNRAS*, 390, 192
- Stappers B. W., Kramer M., Lyne A. G., D'Amico N., Jessner A., 2006, *Chinese Journal of Astronomy and Astrophysics Supplement*, 6, 298
- van Haasteren R., Levin Y., McDonald P., Lu T., 2009, *MNRAS*, 395, 1005
- Verbiest J. P. W., Bailes M., Coles W. A., Hobbs G. B., van Straten W., Champion D. J., Jenet F. A., Manchester R. N., et al., 2009, *MNRAS*, 400, 951
- Wyithe J., Loeb A., 2003, *ApJ*, 595, 614

Gravitational Wave Detection and Data Analysis for
Pulsar Timing Arrays

van Haasteren, R.

2014, XI, 139 p. 40 illus. in color., Hardcover

ISBN: 978-3-642-39598-7