Search of a gravitational wave background in timing residuals of
PSR 1937 + 21: minimal model and upper limits on $\Omega_{gr}$

Gilles Zalamansky,¹ Christian Robert,² François Vernotte¹ and François Taris³

¹Besançon Observatory, BP 1615, 25010 Besançon, France*
²CREST, INSEE, timbre J340, 73673 Paris Cedex 14, France
³BNM-LPT, Paris Observatory, 75014 Paris, France

Accepted 1997 February 26. Received 1997 February 20; in original form 1996 August 5

ABSTRACT
Although many phenomena can account for pulsar timing residuals, previous upper limits on the amplitude $\Omega_{gr}$ of a gravitational wave background were all obtained with the same decomposition into a measurement noise with known variance and a gravitational noise parametrized by $\Omega_{gr}$. The justification of this minimal model (MM) is its ability to provide, on average over an infinite set of measurements, the highest upper limits on $\Omega_{gr}$ compared with other models. Keeping with this model, and using free-access Arecibo data concerning PSR 1937 + 21, we derive a set of 157 estimators of $\Omega_{gr}$ and obtain the 95 per cent confidence upper limit of $1.7 \times 10^{-7}$ using a Bayesian method. We also show evidence of a third term in the residuals (97 per cent confidence), which questions the adequacy of the MM.


1 INTRODUCTION
Cosmic strings which were possibly produced in the early Universe (Kibble 1976; Vilenkin & Shellard 1994) may have emitted a highly energetic gravitational wave background (GWB) in the range $10^{-8}, 10^{-6}$ Hz. The relevant parameter to describe this homogeneous and isotropic background is its energy density per unit volume and its logarithmic frequency in units of closure density, $\Omega(f) = \rho_{cr}^{-1}(3 \ln \rho / \dot{\rho})$.

Assuming a scale-invariant evolution (in horizon units) of a cosmic string network, Yachaspati & Vilenkin (1985) derived an analytic expression for $\Omega(f)$ which predicts

\[ \Omega(f) \approx \Omega_{gr} \approx 10^{-5} \]

independently of $f$. This was derived soon after the discovery of the millisecond pulsar PSR 1937 + 21 (Baker, Kulkarni & Taylor 1983), the high stability of which enables us to set upper limits on $\Omega_{gr}$ when used as a clock in a Doppler effect detection experiment. It is of crucial interest in cosmology to set limits on $\Omega_{gr}$, since cosmic strings could have introduced the primordial density contrast that eventually grew into the large-scale structures of the Universe (Hogan & Rees 1984; Vilenkin & Shellard 1994). Upper limits on $\Omega_{gr}$ are also interesting in time metrology (Taylor 1991). Although several authors (see Caldwell & Allen 1992 and Caldwell, Battye & Shellard 1996 for a detailed review) do not assume $\Omega(f)$ to be flat, we will keep with equation (1) in this paper, in order not to depart from the choices made by Stinebring, Kaspi, Taylor and Ryba (Stinebring et al. 1990; Kaspi, Taylor & Ryba 1994, KTR). The procedure we develop here can be easily extended to any parametrized $\Omega(f)$. We restricted ourselves to PSR1937 + 21 and we used the free Web access files 1937ao.92r (Kaspi et al. 1994).

2 PSR 1937 AND THE GRAVITATIONAL WAVE BACKGROUND
An ideal Doppler effect detector of gravitational waves consists of a perfectly stable clock $\mathcal{C}$ and a receiver $\mathcal{R}$ at a distance $L$, both freely falling with null reciprocal speed in a flat metric, clocks at $\mathcal{R}$ and $\mathcal{C}$ measuring proper times. Calling $T_{\text{e}}$ the time of emission of a given pulse, $T_{\text{s}}$ its arrival time at $\mathcal{R}$, $T_{\text{e}} = T_{\text{s}} + L/c$ the expected arrival time under the assumption of no perturbation, and $R = T_{\text{e}} - T_{\text{s}}$ the timing residual, it can be shown (Maschke 1979, 1982; Maschke & Grishchuk 1980; Bertotti, Carr & Recs 1983) that if this residual is because of a GWB, its spectral density

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System
In a linear approximation, this least-squares fit is equivalent to cancelling the projection of \( p \) over a subset \( H' \) of \( \mathbb{R}^n \) generated by the partial derivatives of (weighted) \( T_e \) and \( T_z \), with respect to \( p_0 \) and \( p_7 \) (Blandford et al. 1984). The dimension of \( H' \) is 8 with models (2) and (3). [We did not include \( \partial T_e / \partial D \) into \( H' \), since file 1973a-92r only refers to frequency \( f_0 = 2300 \) MHz. Had it contained data corresponding to frequency \( f_0 = 1400 \) MHz, we would have included \( \partial T_e / \partial D \) into \( H' \), with \( u_0 = -1 \) if \( f(T_e) = f_1 \) and, inversely for \( u_4 \).]

3 THE MINIMAL MODEL

We now write a general model for the redshift

\[ \rho = w + g + \rho_{\text{par}} + \rho' \]  
\[ w = \text{weighted measurement noise, with covariance matrix } \mathbf{S}_w \text{ (identity matrix).} \]
\[ \mathbf{g} = \text{weighted gravitational term, with covariance matrix} \]
\[ \mathbf{V}_{\mathbf{g}} = \frac{1}{\sigma_g} \int_{-\infty}^{\infty} e^{-|y|/2} dy + \Omega \mathbf{G}. \]  
\[ \rho_{\text{par}} = \text{errors in } p_0 \text{ and } p_6. \]  
\[ \rho' = \text{all other terms.} \]

Let \( H_0^* \) denote the subset of \( \mathbb{R}^n \) orthogonal to \( H' \). We avoid all estimators of \( \Omega_k \) to be biased by \( \rho_{\text{par}} \), if we make them only depend on the projection \( \rho' \) of \( \rho \) over \( H_0^* \).

The integral in (6) diverges at 0, but the inverse time travel \( 1/L \) is a low-frequency cut-off. The three largest eigenvalues of \( \mathbf{G} \) go to infinity when this cut-off goes to zero, whereas the others are only slightly affected by this. Since \( f_0 \) is imprecisely known, wisdom advises us to include corresponding eigenvectors in \( H' \). Fortunately they already belong to it, since they coincide with the partial derivatives of (weighted) \( T_e \) with respect to \( \{N, v, \nu\} \) (Blandford et al. 1984).

3.1 Construction of estimators

Kaspi et al. (1994) and before them Stonebring, Taylor, Ryba and Romani (Stonebring et al. 1990) (STR) used a minimal model (MM) for the residual, in which \( \rho' = 0 \). This is supposed to yield the highest possible upper limits on \( \Omega \), allowed by the timing of PSR 1937. The probability distribution of \( \rho \) in the frame of the MM is (Borovkov 1987; Papoulis 1991)

\[ p(\rho) = \frac{1}{(2\pi)^{n/2} \sqrt{\det V}} \exp \left[ -\frac{1}{2} (\rho - \rho_{\text{par}}) V^{-1} (\rho - \rho_{\text{par}}) \right] \]  

with

\[ V = \mathbf{S}_w + \mathbf{g}^2 + \mathbf{S}_w + \Omega \mathbf{G}. \]

A general treatment would consist of estimating the probability distribution of \( (p_0, p_6, \Omega) \) given \( p_6 \) using a Bayesian method, and deducing the marginal distribution for \( \Omega \), by integrating over \( p_6 \) and \( p_7 \). This implies an eight-dimensional integration, and the necessity of a precise choice of \( f_0 \). The projection of \( \rho \) over \( H' \) greatly simplifies the task, since \( \Omega \) is the only parameter in the distribution of \( p_6 \).
In file 1937ao.92r, N = 165. Using a Gram–Schmidt procedure we built an orthonormal basis \(h_1, \ldots, h_N\) of \(\mathbb{R}\) with the first 157 generating \(H^*\) and the last eight \(H''\). To derive this basis we used the DE200 ephemerides and the rough model (3), assuming that the differences between (3) and the true model used by TEMPO only introduces a negligible angle between our \(H^*\) and the correct one.

Let \(\eta_1, \ldots, \eta_N\) be the components of \(\rho\) over \(h_1, \ldots, h_N\). In matricial formalism, with \(A^T\) meaning the transpose of a matrix \(A\), this reads \(\rho = A^T \eta\), with \(A_{ij} = \rho_{ij}\) [\(\eta\) = canonical basis of \(H^*\), \(\rho = (\rho_1, \ldots, \rho_N)\)].

Let \(P\) denote the rectangular \(N \times N\) matrix, \((N = 157, P = \delta_{ij})\) which performs a projection onto \(H^*\). A straightforward calculation yields the covariance matrix \(\langle \eta_1, \eta_1' \rangle\), \(\langle \eta_1, \eta_1'' \rangle\), \(\langle \eta_1, \eta_2' \rangle = P \cdot \rho\).

\[
\langle \eta_1, \eta_1' \rangle = \rho \Rightarrow \langle \eta_1, \eta_1'' \rangle = \rho \Rightarrow \langle \eta_1, \eta_1'' \rangle = \rho \Rightarrow \langle \eta_1, \eta_1'' \rangle = \rho
\]

\[
\langle \eta_1, \eta_1'' \rangle = P \cdot \rho \Rightarrow \langle \eta_1, \eta_1'' \rangle = P \cdot \rho
\]

We now diagonalize \(\rho\). We call \(w_1, \ldots, w_N\) its eigenvectors, \(\lambda_1, \ldots, \lambda_N\) its eigenvalues, and \(q = (q_1, \ldots, q_N)\) the components of \(\rho\) over \(w_1, \ldots, w_N\). We get the probability distribution of \(q\)

\[
p(q) = \frac{1}{\sqrt{2\pi}} \frac{1}{\lambda^2} \exp \left( -\frac{q^2}{2\lambda^2} \right)
\]

which is the most simplified form of (7) we could hope to derive. We deduce \(q\) from \(\rho\) with the matricial product \(q = \rho^T \cdot \rho\).

In Fig. 1(a) we plot the eigenvalues \(\lambda_1, \ldots, \lambda_{157}\) and in Fig. 1(b) the first eigenvector \(w_1\), which has some features of the vector \(p_1\) used by STR and KTR (Decker & Boynton 1982; Decker 1984) to derive their estimators \(S_1, S_2, S_3\), \(S_4\), \(S_5\), the averages of \(S_i\), over one half, one eighth of the data span. We plot in Fig. 2(a) the components \(q_1, \ldots, q_{157}\), and in Fig. 2(b) the subset \(q_{25}, \ldots, q_{157}\), whose variance is too large for those numbers to be \(\mathcal{N}(0, 1)\) distributed, as we will now show.

3.2 Discussing the minimal model

Unless \(\Omega\) is large compared with 1, which is unlikely according to (Vachaspati & Vilenkin 1985; Bennett & Bouchet 1988; Caldwell & Allen 1992; Vilenkin & Shellard 1994; Caldwell et al. 1995), white noise dominates the distribution of all \(q_1\) for which \(\lambda_1\) is less than 1. The last 130 numbers \((q_{158}, \ldots, q_{157})\) are in such a situation, since \(\lambda_{158}, \ldots, \lambda_{157}\) decrease from 0.015 to \(10^{-8}\). If we assume them to be Gaussian and standardized, then \(S = q_{25}^2 + \cdots + q_{157}^2\) should be chi-square distributed with 130 degrees of freedom, which is close to \(\mathcal{N}(130, 260)\). The number \(S = 130, 260\) should then be \(\mathcal{N}(0, 1)\) distributed. The value we obtain, \(s = 2.17\), makes this hypothesis quite unlikely: \(s\) has probability 0.97 to be smaller under standard normal distribution. In other words the MM is not compatible with the data on PSR 1937 at the 97 per cent confidence level. Instead of writing \(q_1 = q_{157}^2 + q_{157}^2\), \(\Var(q_{157}) = 1\), \(\Var(q_{157}) = \lambda_{157}\Omega\), where \(q_{157}^2\) and \(q_{157}^2\) are the images of \(w\) and the projection \(\mathcal{H}\) and the rotation \(\mathcal{H}\), we should write \(q_1 = q_{157}^2 + q_{157}^2 + q_{157}^2\), \(q_{157}^2\) unknown.

On the average over an infinite set of measurements \(q_1\), yields \(\mathbb{E}[q_1^2] = \mathbb{E}[q_{157}^2] + \mathbb{E}[q_{157}^2] + 2\mathbb{E}[q_{157} q_{157}^2]\), \(\mathbb{E}[q_{157}^2]\) with \(\mathbb{E}[q_{157} q_{157}^2] = \sim q_{157} q_{157}^2 + q_{157} q_{157}^2\), where \(q_{157}^2\) is unknown.

4 BAYES THEOREM AND UPPER LIMITS ON \(\Omega\)

Bayes theorem (Bayes 1763) tells us that the probability distribution of a measurement \(x\) given parameter(s) \(\theta\), and \(x\), is not enough to get a probability distribution for \(\theta\). One also needs an a priori distribution (the 'prior') \(P(\theta)\), reflecting his or her degree of belief that \(\theta\) should belong to given regions of the parameter space. The posterior distribution is then

\[
E[q_1^2] = 1 + \lambda_{157}\Omega + 2E[q_{157}^2]
\]

\[
E[q_1^2] = 1 + \lambda_{157}\Omega + 2E[q_{157}^2]
\]

\[
E[q_1^2] = 1 + \lambda_{157}\Omega + 2E[q_{157}^2]
\]

\[
E[q_1^2] = 1 + \lambda_{157}\Omega + 2E[q_{157}^2]
\]

\[
E[q_1^2] = 1 + \lambda_{157}\Omega + 2E[q_{157}^2]
\]

\[
E[q_1^2] = 1 + \lambda_{157}\Omega + 2E[q_{157}^2]
\]
The everlasting debate is: which prior should we choose when we know nothing about $\theta$? Two arguments favor Jeffreys prior:

$$\pi_{\text{Jeffreys}}(\theta) = \sqrt{\mathcal{F}(\theta)},$$

where $\mathcal{F}(\theta)$ is the determinant of Fisher's information matrix, $\mathcal{F}(\theta) = \text{Cov} \left( \ln L / \partial \theta_i, \ln L / \partial \theta_j \right)$, $L$ being the likelihood. These arguments are as follows:

1. The $\chi^2$ distance between two distributions $p_1 = p(x|\theta_1)$ and $p_2 = p(x|\theta_2)$, with $\Delta \theta = \theta_1 - \theta_2$,

$$d^2 = \left[ \frac{p_1(x) - p_2(x)}{p_1(x)} \right]^2 p_1(x) \, dx$$

is $\Delta \theta \cdot \mathcal{F}(\theta_1) \cdot \Delta \theta + O(\|\Delta \theta\|^2)$ (Borovkov 1987; Monfort 1996). $d^2$ can be considered as the average discriminating power of the measurement. In the parametrization $\phi(\theta)$ for which $\mathcal{F}(\phi)$ is constant [which is equivalent to (11) in a general parametrization (Robert 1994; Schervish 1995)], this discriminating power is the same for all $\phi$. It is then sensible to choose a prior which respects this uniformity: $\pi(\phi) = 1$ [which is also Laplace's prior (Laplace 1960)].

2. When one performs a large number $n$ of identical and independent measurements, the maximum likelihood estimator (MLE) distribution converges to a Gaussian one, with mean $\theta$ and covariance matrix $[n \mathcal{F}(\theta)]^{-1}$ (Schervish 1995). With $\mathcal{F}(\phi) = 1$ the MLE converges to the true value at uniform speed. See Bernardo (Bernardo 1979) and Robert (Robert 1994) for a more involved discussion.

It is straightforward to check that $\pi_{\text{Jeffreys}}(\theta)$ is uniform when $\theta$ is a translation parameter [$p(x|\theta) = p(x - \theta)$], that it is $1/\theta$ when $\theta$ is a scaling one [$p(x|\theta) = p(x/\theta)$], that it is invariant under reparametrization: $\pi_{\text{Jeffreys}}(\phi) = \pi_{\text{Jeffreys}}(\theta) |J(\theta)|$, $J(\theta)$ being the Jacobian determinant of the transformation, and that when the components of $x$ are independent $\pi_{\text{Jeffreys}}(\theta)$ is unchanged under reordering. Any of those conditions, had it not been fulfilled, would have ruled out Jeffreys prior as a sensible choice.

In our case, all $Q_1, \ldots, Q_{157}$ with $Q_2 = q_i^2$, are chi-square distributed:

$$p(Q|\Omega) = \frac{1}{\Gamma(1+\lambda_i \Omega_i)Q^\lambda_i} e^{-\Omega_i / 2},$$

Fisher information is

$$\mathcal{F}(\Omega_i) = \sum_{i=1}^{157} \frac{\lambda_i^2}{(1+\lambda_i \Omega_i^2)},$$

and the prior we choose is

$$\pi(\Omega) = \sqrt{\mathcal{F}(\Omega)} \cdot Y(\Omega),$$

where $Y$ is the step function, equal to 1 over $[0, \infty]$ and 0 over $]-\infty, 0]$. With (14), (13) and (10) we obtained the probability density function and the corresponding cumulative density function for $\Omega$. We plot them in Fig. 3(a) and (b). The upper limit at the 95 per cent confidence level we find for $\Omega$ is $1.7 \times 10^{-7}$, which is slightly more stringent than Stinebring's value of $4 \times 10^{-7}$ (Stinebring et al. 1990).

4.1 PSR 1855 + 09

Using Stinebring's estimators (Stinebring et al. 1990) $S_1, S_2, S_3, S_4$ and their assumed probability distributions (Kaspi et al. 1994) (conditional to $\Omega_i$), the same Bayesian analysis (Mchugh et al. 1996) yields the 95 per cent confidence upper limit $0.93 \times 10^{-6}$ on $\Omega_i$. However, the construction of those estimators is not clearly detailed in papers Stinebring et al. (1990) and Kaspi et al. (1994), nor is the choice of two different clock models for both pulsars: least-squares of fit of TEMPO removes the cubic trend in timing residuals of PSR 1855 + 09 (Byba & Taylor 1991), whereas it only removes the quadratic one in those of PSR 1937 + 21. This may explain why the values of $S_1$ and $S_2$ are so small compared with the assumed level of white noise (Kaspi et al. 1994). We thus restricted ourselves to PSR 1937 + 21.

4.2 Remark: tests of hypotheses

Consider only one estimator $Q$, the null hypothesis $H(\omega) = H_\omega \geq \omega$ and the 'alternative', $H_\omega = 0$, where $[0, Q]$ is taken as rejection region for $H(\omega)$. The probability of rejecting $H(\omega)$ whereas it is true, is $\alpha(Q) = p([0, Q]|H_\omega)$ (type I error), and the probability of accepting it whereas it is false (type II error) is $\beta(Q) = p([Q, \infty]|H_\omega = 0)$. Confusing $1 - \alpha(Q)$ with the probability for $\Omega$ to be less than $\omega$ is misleading here, for the following three reasons. First, we would find a non-zero probability for $\Omega$ to be less than $\omega$ with this method. Secondly, if the $\sigma_i, \ldots, \sigma_q$ grow to infinity, $Q$ and $\hat{Q}$ tend to 0, and so does $\alpha(Q)$, for any $\omega$, including very small values. In other words, the less reliable the measurements, the tighter the constraint on $\Omega_i$, which is of course unacceptable! Thirdly, estimation and decision are two different problems: before taking a decision, one wants to estimate the probability to take the wrong one, $\alpha$ or $\beta$.
which is different from the probability for the parameter to be in a given region. This latter is related with \( r(\theta) \), and this distribution is the starting point of Bayesian methods. Two authors (Thorsett & Dewey 1996) who recently used a Neyman Pearson test to set limits on \( \Omega_c \), obtained, we believe (McHugh et al. 1996), inadequate results.

5 CONCLUSION

An upper limit around \( 10^{-7} \) rules out the simplest models of structure formations based on cosmic strings (Turok 1986), yet not all of them (Vilenkin & Shellard 1994). We wanted to underline the model-dependence of this result, and the incompatibility of the generally assumed minimal model with the available data. Alternative models should be studied at this stage, a task we could not undertake in this paper.

REFERENCES

Bayes T., 1763, Phil. Trans. Roy. Soc., 53, 370
Hogan C. J., Rees M. J., 1984, Nat, 301, 109
Kibble W., 1976, J. Phys A, 9, 1387
Laplace P. S., 1800, Oeuvres Completes, Gauthier-Villars, Paris