

A χ^2 veto for continuous gravitational wave searches

Llucia Sancho de la Jordana¹ and Alicia M Sintes^{1,2}

¹ Departament de Física, Universitat de les Illes Balears, Cra. Valldemossa Km. 7.5,
E-07122 Palma de Mallorca, Spain

² Max-Planck-Institut für Gravitationsphysik, Albert Einstein Institut, Am Mühlenberg 1,
D-14476 Golm, Germany

E-mail: llucia.sancho@uib.es and sintes@aei.mpg.de

Received 7 April 2008, in final form 15 May 2008

Published 2 September 2008

Online at stacks.iop.org/CQG/25/184014

Abstract

χ^2 vetoes are commonly used in searching for gravitational waves, in particular for broadband signals, but they can also be applied to narrowband continuous wave signals, such as those expected from rapidly rotating neutron stars. In this paper, we present a χ^2 veto adapted to the Hough transform searches for continuous gravitational wave signals; we characterize the χ^2 significance plane for different frequency bands and discuss the expected performance of this veto in LIGO analysis.

PACS numbers: 04.80.Nn, 07.05.Kf, 95.55.Ym, 97.60.Gb

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Continuous gravitational wave signals emitted by neutron stars in our galaxy are among the targets of on-going searches for gravitational waves using GEO600 [1, 2], LIGO [3, 4] and VIRGO [5] data. Examples of such searches include targeting known radio pulsars [6–8] and the low-mass x-ray binary system Scorpius X-1 [9, 10], as well as all-sky surveys for unknown rotating neutron stars [9, 11, 12]. The first type of searches typically use matched filtering techniques and are not very computationally expensive. The second type of searches look for as yet undiscovered sources. This involves searching over large parameter space volumes and turns out to be computationally limited, as the number of templates that must be searched over increase rapidly with the observation time. The ultimate goal for wide parameter searches for continuous signals over large data sets is to employ hierarchical schemes which alternate coherent and semi-coherent techniques [13–18], as those currently employed by Einstein@Home [19], a distributed-computing effort that uses the idle CPU time of computers across the world.

The Hough transform [20, 21] is an example of a semi-coherent method that can be used to select candidates in parameter space to be followed up. Results of the Hough transform to search the entire sky have been reported in [11, 12]. In those papers, the Hough transform was used to search for cumulative excess power from a hypothetical periodic signal by examining successive spectral estimates based on short Fourier transforms (SFTs) of the calibrated detector data, taking into account the Doppler frequency shift due to the motion of the detector with respect to the solar system barycenter and the intrinsic frequency evolution of the source. Two flavors of the Hough transform have been developed and employed for different searches, the ‘standard Hough’ [11, 20] and the ‘weighted Hough’ [12, 21]. In the ‘standard Hough’, the cumulative excess power is computed as the sum of binary zeroes and ones, where a SFT contributes unity if the power exceeds a normalized threshold, and in the ‘weighted Hough’ the contribution of the SFTs is weighted according to the noise and detector antenna pattern to maximize the signal-to-noise (SNR) ratio.

In all hierarchical methods it is crucial that the selection of candidates done by the semi-coherent stage is as effective as possible, since it determines the final sensitivity of the full pipeline. For this reason the development of veto and/or coincidence tests is very important in order to reduce the number of false alarms. In this paper, we present a χ^2 veto adapted to the Hough transform searches for continuous wave signals and we discuss the expected performance of this veto in LIGO analysis.

χ^2 discriminators are commonly used in searching for gravitational wave signals. In particular, for binary inspiral searches a χ^2 time–frequency discriminator is used as a veto for the output of matched filter, by analyzing the output of different frequency bands. This χ^2 test [22, 23] was specifically constructed for broadband signals, but it can be modified for signals that are narrowband as the continuous wave signals expected from rapidly rotating neutron stars. For these continuous wave signals that will be observed over periods of several months, in order to build up the sufficient SNR, we propose to split the data into several chunks, analyze each chunk separately and construct a χ^2 statistic by combining the partial results. This statistic would then be able to discriminate if the SNR accumulates along the different chunks in a way that is consistent with the properties of the signal and the detector noise.

The remainder of the paper is organized as follows: section 2 briefly summarizes the Hough transform and its statistical properties. Section 3 derives a χ^2 discriminator for different implementations of the Hough transform. Section 4 characterizes the χ^2 significance plane and discusses its application using LIGO data. Section 5 concludes with a summary of the results.

2. The Hough transform method

The Hough transform is a well-known method for pattern recognition that has been applied to the search for continuous gravitational waves. In this case, the Hough transform is used to find a signal whose frequency evolution fits the pattern produced by the Doppler shift and the spin-down in the time–frequency plane of the data. Further details can be found in [20], here we only give a brief summary and statistical properties.

The starting point for the Hough transform are N SFTs. Each of these SFTs is digitized by setting a threshold ρ_{th} on the normalized power

$$\rho_k = \frac{2|\tilde{x}_k|^2}{T_{\text{coh}}S_n(f_k)}. \quad (1)$$

Here \tilde{x}_k is the discrete Fourier transform of the data, the frequency index k corresponds to a physical frequency of $f_k = k/T_{\text{coh}}$, $S_n(f_k)$ is the single-sided power spectral density of the

detector noise and T_{coh} is the time baseline of the SFT. The k th frequency bin is selected if $\rho_k \geq \rho_{\text{th}}$, and rejected otherwise. In this way, each SFT is replaced by a collection of zeros and ones called a peak gram. The probability that a frequency bin is selected is $q = e^{-\rho_{\text{th}}}$ for Gaussian noise and η , given by

$$\eta = q \left\{ 1 + \frac{\rho_{\text{th}}}{2} \lambda_k + \mathcal{O}(\lambda_k^2) \right\} \quad (2)$$

in the presence of a signal. λ_k is the signal-to-noise ratio within a single SFT, and for the case when there is no mismatch between the signal and the template:

$$\lambda_k = \frac{4|\tilde{h}(f_k)|^2}{T_{\text{coh}}S_n(f_k)} \quad (3)$$

with $\tilde{h}(f)$ being the Fourier transform of the signal $h(t)$.

The Hough transform is used to map points from the time–frequency plane of our data (understood as a sequence of peak grams) into the space of the source parameters. Each point in parameter space corresponds to a pattern in the time–frequency plane, and the Hough number count n is the weighted sum of the ones and zeros of the different peak grams along this curve. For the ‘weighted Hough’ this sum is computed as

$$n = \sum_{i=1}^N w_i n_i, \quad (4)$$

where n_i is either 0 or 1 depending on where the power crosses the threshold and the weights are normalized according to

$$\sum_{i=1}^N w_i = N. \quad (5)$$

When all $w_i = 1$ we obtain the ‘standard Hough’. The ‘weighted Hough’ can improve the sensitivity of the search taking into account the possible non-stationarities of the detector noise and the amplitude modulation due to the motion of the detector, and allow for multi-interferometer searches.

For large values of N , the number count distribution n can be considered a continuous variable and well approximated by a Gaussian distribution:

$$p(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n - \langle n \rangle)^2}{2\sigma^2}\right). \quad (6)$$

In the absence of a signal, the mean and variance are

$$\langle n \rangle = Nq \quad \text{and} \quad \sigma^2 = q(1 - q) \sum_{i=1}^N w_i^2, \quad (7)$$

and in the presence of a signal, the mean and variance of n become

$$\langle n \rangle = qN + \frac{q\rho_{\text{th}}}{2} \sum_{i=1}^N w_i \lambda_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^N w_i^2 \eta_i (1 - \eta_i). \quad (8)$$

Because of the weights σ varies for different sky locations we should not compare number counts directly but the significance of a number count. The significance s of the observed number counts n is defined as

$$s = \frac{n - \langle n \rangle}{\sigma}, \quad (9)$$

where $\langle n \rangle$ and σ are the expected mean and standard deviation for pure noise. Furthermore, one can see that setting a threshold at a given false alarm rate α is equivalent to set a threshold at a certain significance [21]

$$s_{\text{th}} = \sqrt{2} \operatorname{erfc}^{-1}(2\alpha). \quad (10)$$

3. The χ^2 veto

In this section we derive a χ^2 discriminator for the different implementations of the Hough transform. The idea is to split the data into p non-overlapping chunks, each of them containing a certain number of SFTs $\{N_1, N_2, \dots, N_p\}$, such that

$$\sum_{j=1}^p N_j = N, \quad (11)$$

and analyze them separately, obtaining the Hough number count n_j which, for the same pattern across the different chunks, would then satisfy

$$\sum_{j=1}^p n_j = n, \quad (12)$$

where n is the total number count for a given point in parameter space. The χ^2 statistic will look along the different chunks to see if the SNR accumulates in a way that is consistent with the properties of the signal and the detector noise. Small values of χ^2 are consistent with the hypothesis that the observed SNR (or more precisely the significance) arose from a detector output which was a linear combination of Gaussian noise and the continuous wave signal. Large values of χ^2 indicate either the signal did not match the template or that the detector noise was non-Gaussian.

3.1. The standard Hough

In the simplest case in which all weights are set to unity, it is easy to see that the expected values of the number counts are

$$\langle n \rangle = N\eta, \quad \sigma_n^2 = N\eta(1 - \eta), \quad (13)$$

$$\langle n_j \rangle = N_j\eta = N_j \frac{\langle n \rangle}{N}, \quad \sigma_{n_j}^2 = N_j\eta(1 - \eta), \quad (14)$$

where n is the total measured number count and η is the probability of selecting a peak in the presence of a signal.

Consider the p quantities defined by

$$\Delta n_j \equiv n_j - \frac{N_j}{N}n. \quad (15)$$

With this definition, it holds true that

$$\langle \Delta n_j \rangle = 0, \quad \sum_{j=1}^p \Delta n_j = 0, \quad \langle n_j n \rangle = \frac{N_j}{N} \langle n^2 \rangle, \quad (16)$$

and the expectation value of the square of Δn_j is

$$\langle (\Delta n_j)^2 \rangle = \left(1 - \frac{N_j}{N}\right) N_j \eta (1 - \eta). \quad (17)$$

Therefore we can define the χ^2 discriminator statistic by

$$\chi^2(n_1, \dots, n_p) = \sum_{j=1}^p \frac{(\Delta n_j)^2}{\sigma_{n_j}^2} = \sum_{j=1}^p \frac{(n_j - nN_j/N)^2}{N_j \eta (1 - \eta)}, \quad (18)$$

and expected value of χ^2 is

$$\langle \chi^2 \rangle = \sum_{j=1}^p \left(1 - \frac{N_j}{N} \right) = p - 1. \quad (19)$$

Equation (18) has a χ^2 distribution with $p - 1$ degrees of freedom. To implement this discriminator, we need to measure, for each point in parameter space, the total number count n , the partial number counts n_j and assume a constant value of $\eta = n/N$.

3.2. The weighted Hough

We can generalize the previous result (18) for the weighted Hough transform. Let I_j be the set of SFT indices for each different p , the mean and variance of the number count become

$$\langle n_j \rangle = \sum_{i \in I_j} w_i \eta_i, \quad \langle n \rangle = \sum_{j=1}^p \langle n_j \rangle, \quad \sigma_{n_j}^2 = \sum_{i \in I_j} w_i^2 \eta_i (1 - \eta_i), \quad (20)$$

and we can define

$$\Delta n_j \equiv n_j - n \frac{\sum_{i \in I_j} w_i \eta_i}{\sum_{i=1}^N w_i \eta_i}, \quad (21)$$

so that $\langle \Delta n_j \rangle = 0$, $\sum_{j=1}^p \Delta n_j = 0$. The χ^2 discriminator would now be:

$$\chi^2 = \sum_{j=1}^p \frac{(\Delta n_j)^2}{\sigma_{n_j}^2} = \sum_{j=1}^p \frac{(n_j - n(\sum_{i \in I_j} w_i \eta_i) / (\sum_{i=1}^N w_i \eta_i))^2}{\sum_{i \in I_j} w_i^2 \eta_i (1 - \eta_i)}. \quad (22)$$

In a given search, we can compute $\sum_{i \in I_j} w_i$, $\sum_{i \in I_j} w_i^2$ for each of the p chunks. The problem for implementing this discriminator (22) is that the different η_i values cannot be measured from the data itself because they depend on the exact SNR for a single SFT as defined in equations (2) and (3) and not just its averaged value. For this reason we propose to approximate in equation (22) $\eta_i \rightarrow \eta^*$, where $\eta^* = n/N$. Thus

$$\chi^2 \approx \sum_{j=1}^p \frac{(n_j - n(\sum_{i \in I_j} w_i) / N)^2}{\eta^* (1 - \eta^*) \sum_{i \in I_j} w_i^2}. \quad (23)$$

4. Application of the χ^2 veto on the LIGO fourth science run

We use as a playground the SFT data produced during LIGO's 29.5 day fourth science run (S4) and analyze it by means of the 'weighted Hough' scheme, combining the data of the three LIGO detectors (H1, H2, L1). The SFTs were generated directly from the calibrated data stream, using 30 min intervals of contiguous data for which the interferometer is operating in what is known as science mode, having in the end 1004 SFTs from H1, 1063 SFTs from H2 and 899 from L1. An all-sky search for periodic gravitational waves in the frequency range 50–1000 Hz and with the frequency's time derivative in the range -1×10^{-8} to 0 Hz s^{-1} using the S4 data is reported in [12]. In that paper, three different semi-coherent methods of

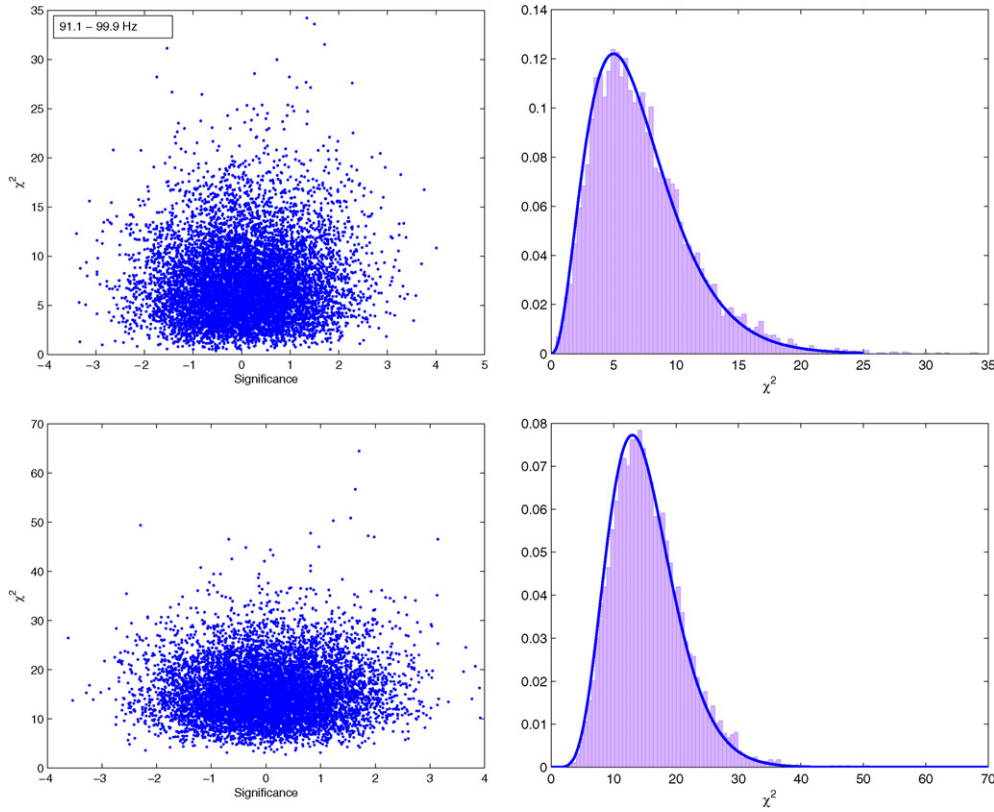


Figure 1. χ^2 versus significance obtained for the 90–100 Hz band and comparison of the χ^2 values obtained with a χ^2 distribution with $p - 1$ degrees of freedom. The top figures correspond to $p = 8$ and the bottom ones to $p = 16$.

transforming and summing strain power from SFT data have been used. In [12], to identify the most interesting subset in parameter space, a fixed threshold of SNR or significance of 7 was applied for all three searches, and candidates were analyzed by means of a simple coincident test. The paper reported no evidence of periodic gravitational radiation.

To characterize the χ^2 significance plane in order to discriminate between instrumental noise and real signals, we first study some small frequency bands of the S4 data. The main purpose is to verify that in those bands free of large spectral disturbances, the measured χ^2 distribution is consistent with what we would expect in the case of Gaussian noise. For this we split the data into $p = 8$ or $p = 16$ segments. It is reasonable to split the data into segments with a similar relative contribution to the total number count. Therefore, we split the SFTs in such a way that the sum of weights into each block satisfies $\sum_{i \in I_j} \omega_i \approx N/p$. In figure 1 we show the distributions in the 90–100 Hz frequency band. The results agree very well with the expected theoretical distribution. This test was carefully done avoiding the 1 Hz comb present in the data.

The next step is to characterize the χ^2 significance plane in the presence of signals. If there was no mismatch between the signals and the templates, and if we could compute exactly the χ^2 value given by equation (22) instead of (23), we would obtain the same χ^2 distribution as for the Gaussian noise case only. But in a real search, templates are placed on

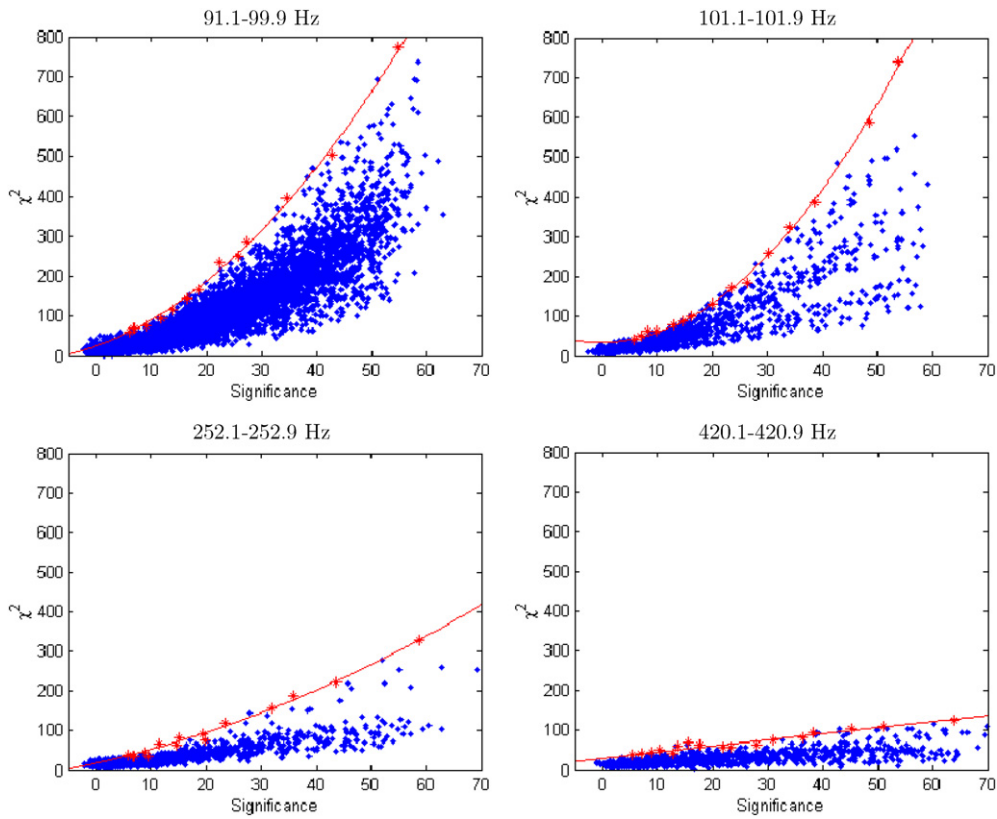


Figure 2. χ^2 versus significance for software injected pulsar signals. We represent the results for four different frequency bands of 0.8 Hz free of spectral disturbances (avoiding the 1 Hz comb) using $p = 16$, together with the best-fitted quadratic curve of the envelope of the points obtained.

a grid and due to the mismatch there is a dependency of the χ^2 values with the significance [23]. For this reason we select 22 frequency bands between 50 and 1000 Hz free of spectral disturbances and we analyze them by means of Monte Carlo software injections. For each of these bands we inject at least 10 000 artificial signals of different amplitudes, frequencies, inclination angles and sky locations. We have checked that 10 000 signals were enough for our purposes. More extensive analyses were performed in the 91–100 Hz band, in which we injected up to 100 000 signals without observing any considerable impact in the final result. In figure 2 we represent the results obtained for four of these different bands. The Monte Carlo injections have been analyzed with no mismatch and also with a small mismatch between the signals and the templates, using the same grid separation that was employed in the S4 search [12]. It is worth mentioning that this grid was not based on a metric approach that would guarantee a maximum given mismatch at any point in parameter space, but it was uniformly spaced in frequency and frequency derivative and used an almost isotropic grid, but frequency dependent, in the sky. As a consequence of this choice, together with the fact that the S4 run was shorter than a month, the mismatch depends on the sky location and the frequency.

For each of the analyzed bands we find a veto curve. The way we proceed is the following: we first sort the points with respect to the significance, we group them in sets containing equal number of points (typically 50) and we pick that with the largest χ^2 value (the highlighted

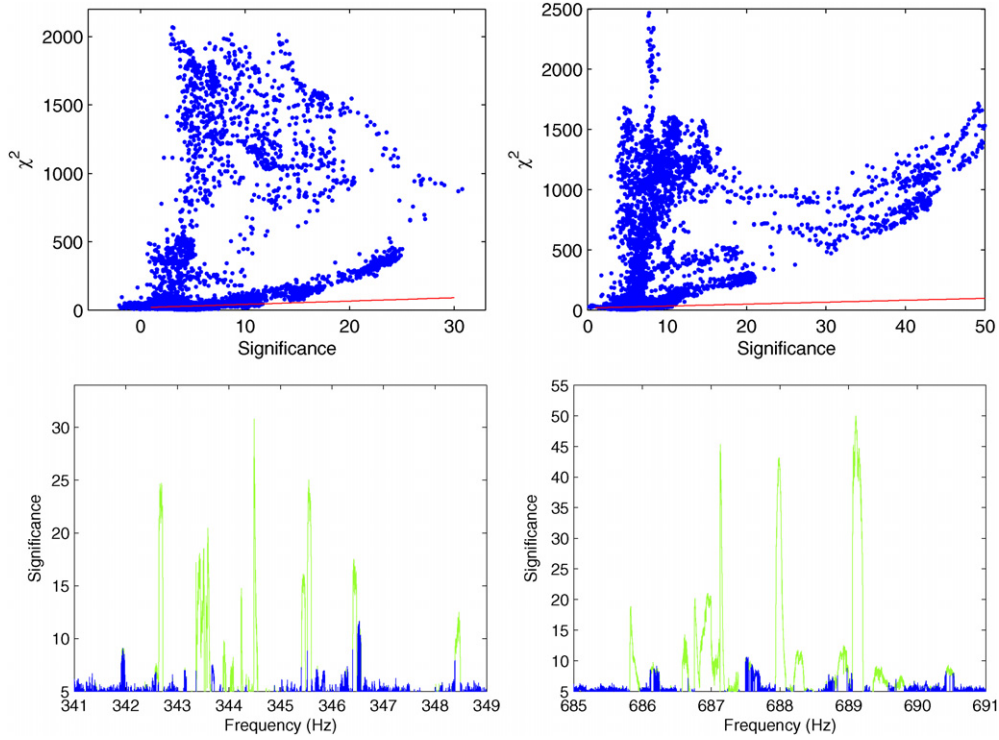


Figure 3. Veto of the violin modes by the χ^2 using $p = 16$. In the top panels there is the χ^2 significance plane and the solid line corresponds to the veto curve at an average frequency in each case. The bottom panels show the significance versus frequency before and after applying the χ^2 veto. The test successfully vetoes the strongest lines related to the violin modes while it fails to veto others, e.g. the line at 342 Hz. The dots below the veto curve with small values of significance are consistent with the expected distribution in the case of Gaussian noise.

stars in figure 2). With the set of selected points, we fit a curve to the upper contour of the χ^2 significance planes. From the set of these curves we deduce empirically the best parameters of a quadratic curve $\chi^2 = As^2 + Bs + C$ valid for any frequency between 50 and 1000 Hz and for significance values greater than 5. For the S4 data and using $p = 16$, these parameters are:

$$A = \begin{cases} -4.229 \times 10^{-4} f + 0.1274 & 50 < f < 300 \\ 0 & 300 < f < 1000, \end{cases}$$

$$B = \frac{562.6}{f} + 0.7873,$$

$$C = 18.666.$$

We want to point out that this veto curve, in this case, was only valid for significance values greater than 5. In a search, one will set a threshold on the significance equal or higher than this value (a threshold of 7 was used for selecting triggers in the S4 search in [12]), and for each trigger one will compare if the χ^2 value is above or below the veto curve.

Using this veto curve we have analyzed the whole S4 data and here we show details of some frequency bands. This χ^2 discriminator is able to veto all the violin modes present in the data and many other narrow instrumental lines, as it is shown in figure 3. In this figure

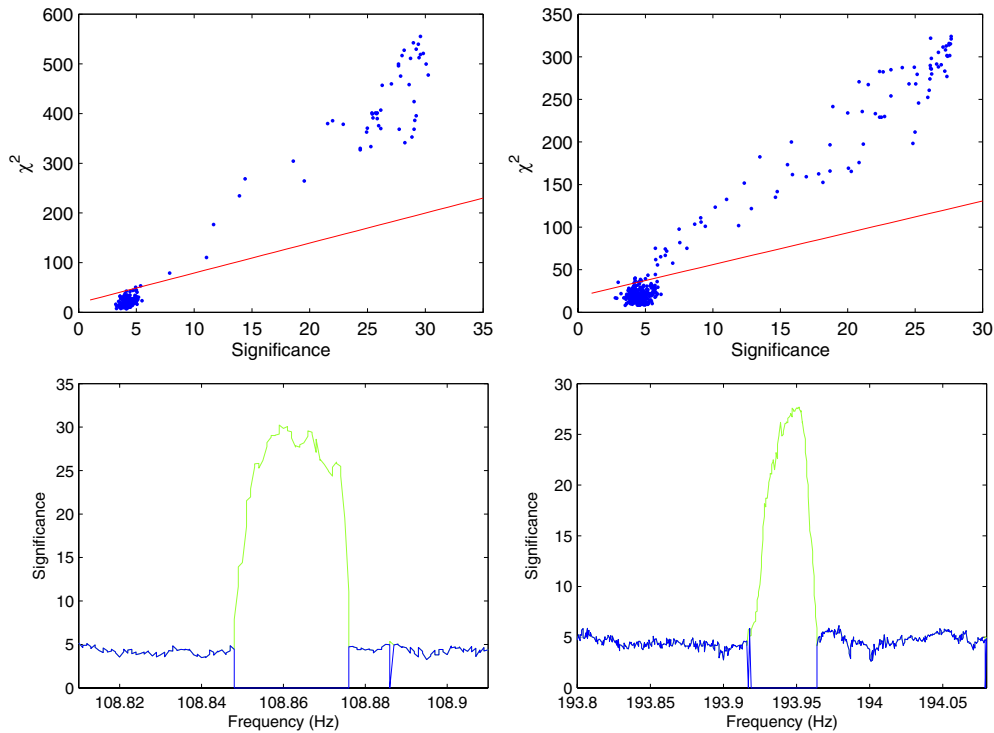


Figure 4. Same as figure 3 for the frequency bands that contain the hardware injected Pulsar 3 (left) and Pulsar 8 (right). In the top panels, the dots on top of the solid line correspond to the hardware injected signals that are vetoed. The dots below the curve, with low values of significance, are consistent with the expected distribution in the Gaussian noise case and correspond to the frequencies not affected by the presence of the hardware injections. The χ^2 test is able to veto these signals because they were injected in the data in an intermittent way and therefore they did not behave like the signals we are looking for.

one can see how the strong lines are clearly vetoed while this test failed to veto others, e.g. the line at 342 Hz.

During the S4 run ten artificial pulsar signals were hardware injected in the data in an intermittent way. Four of these pulsars: Pulsar2, Pulsar3, Pulsar8 and Pulsar9 were strong enough to be detected by the multi-interferometer Hough search (see [12] for further details). Because these signals are not continuously present in the data, the χ^2 test is able to veto them as shown in figure 4. Of course, this does not happen if we analyze only the data segments when the injections took place. In those cases, the χ^2 significance plane is consistent with that obtained for software injected signals. It is worth mentioning that this test failed to veto all the 60 Hz line harmonics.

5. Conclusions

In this paper, we have presented a new χ^2 veto adapted to the Hough transform search for continuous gravitational wave signals and discussed the performance of this veto using the data from LIGO fourth science run. The implementation of this veto is very simple and does not imply a considerable increase in computational cost. We foresee its usage in future

searches performed by the LIGO and VIRGO Scientific Collaboration. The χ^2 veto presented here is adapted to the Hough transform but it could be generalized for other semi-coherent techniques such as PowerFlux or StackSlide, and also when the Hough transform starts with the maximum detection statistic (known as \mathcal{F} -Statistics [9, 20]) rather than SFT power as the input data.

Acknowledgments

We would like to thank the LIGO Scientific Collaboration for many useful discussions and for providing the data. We also acknowledge the support of the Max-Planck Society, the Spanish Ministerio de Educación y Ciencia Research Projects FPA-2007-60220, HA2007-0042, CSD207-00042 and the Govern de les Illes Balears, Conselleria d'Economia, Hisenda i Innovació. We are also grateful to the Albert Einstein Institute for hospitality where this work was initiated. The analysis was performed with the aid of the Merlin cluster of the Albert Einstein Institute. This document has been assigned LIGO Laboratory Document No. LIGO-P080030-00-Z.

References

- [1] Willke B *et al* 2004 *Class. Quantum Grav.* **21** S417
- [2] Grote H *et al* 2005 *Class. Quantum Grav.* **22** S193
- [3] Abramovici A *et al* 1992 *Science* **256** 325
- [4] Barish B and Weiss R 1999 *Phys. Today* **52** 44
- [5] Caron B *et al* 1997 *Nucl. Phys. B* **54** 167
- [6] Abbott B *et al* (The LIGO Scientific Collaboration) 2004 *Phys. Rev. D* **69** 082004
- [7] Abbott B *et al* (The LIGO Scientific Collaboration) 2005 *Phys. Rev. Lett.* **94** 181103
- [8] Abbott B *et al* (The LIGO Scientific Collaboration) 2007 *Phys. Rev. D* **76** 042001
- [9] Abbott B *et al* (The LIGO Scientific Collaboration) 2007 *Phys. Rev. D* **76** 082001
- [10] Abbott B *et al* (The LIGO Scientific Collaboration) 2007 *Phys. Rev. D* **76** 082003
- [11] Abbott B *et al* (The LIGO Scientific Collaboration) 2005 *Phys. Rev. D* **72** 102004
- [12] Abbott B *et al* (The LIGO Scientific Collaboration) 2008 *Phys. Rev. D* **77** 022001
- [13] Brady P R and Creighton T 2000 *Phys. Rev. D* **61** 082001
- [14] Papa M A, Schutz B F and Sintes A M 2001 *Gravitational Waves: A Challenge to Theoretical Astrophysics (ICTP Lecture Notes Series vol 3)* ed V Ferrari, J C Miller and L Rezzolla (Trieste: ICTP) p 431
- [15] Cutler C, Gholami I and Krishnan B 2005 *Phys. Rev. D* **72** 042004
- [16] Frasca S 2000 *Int. J. Mod. Phys. D* **9** 369
- [17] Frasca S, Astone P and Palomba C 2005 *Class. Quantum Grav.* **22** S1013
- [18] Frasca S and Palomba C 2004 *Class. Quantum Grav.* **21** S1645
- [19] <http://einstein.phys.uwm.edu>
- [20] Krishnan B, Sintes A M, Papa M A, Schutz B F, Frasca S and Palomba C 2004 *Phys. Rev. D* **70** 082001
- [21] Krishnan B and Sintes A M 2007 *Hough Search with Improved Sensitivity* LIGO Technical Document T070124 (Available at <http://admdbsrv.ligo.caltech.edu/dcc/>)
- [22] Allen B 2000 *GRASP: A Data Analysis Package for Gravitational Wave Detection* version 1.9.8, p 180 (Available at <http://www.lsc-group.phys.uwm.edu/~ballen/grasp-distribution/>)
- [23] Allen B 2005 *Phys. Rev. D* **71** 062001