Searching for continuous waves by line identification

M.A. Papa[1,4], P. Astone[2], S. Frasca[2,3], B.F. Schutz [1]

[1]*Max Planck Institut fuer Gravitationsphysik, Albert Einstein Institut, 1 Schlaatzweg, 14473 Potsdam, Germany*
[2]*I.N.F.N. - sez. Roma1, P.le A. Moro 2, 00185 Roma, Italy*
[3]*Dip. di Fisica, Univ. di Roma "La Sapienza", P.le A. Moro 2, 00185 Roma, Italy*
[4]*I.N.F.N. - sez. Roma2, Via della Ricerca Scientifica 1, 00133 Roma, Italy*

Abstract.
Results regarding the first stage of a hierarchical procedure to detect continuous gravitational signals with no a priori information on the signal, will be presented. In particular the procedure which we shall adopt to optimize the analysis will be described and examples of how it works in real cases will be shown.

1 The hierarchical procedure

The main issues related to the search of continuous gravitational wave signals, such as those expected from isolated neutron stars, have been examined by B.Schutz in his talk at this meeting, and we refer to his presentation to provide a general frame for this work (see also [**?**] and refs. therein).

The hierarchical procedure that we are studying [**?**] consists of two parts. The first one aims at learning about candidate signals which may be present in the data. The second uses this information to analyse more deeply the signals that have been spotted, concentrating on smaller frequency bands, thus on less data. Eventually this allows to gain the full sensitivity, that is the same spectral resolution that one would have got by performing a single FFT over the entire observation period. The real sensivity of the whole procedure depends, really, on the minimum signal one can spot with the first analysis: if a signal is not picked at that stage, it is simply lost. Here we intend to report about work in progress on optimizing this first analysis. The second stage is said to recover the full sensitivity in the sense that, if the signal is picked, it will be possible to extract information with the detail of the highest signal to noise ratio (SNR hereafter).

The whole procedure relies on the existence of a frequency domain data base (FDDB, hereafter) where the data are stored as FFTs performed over a suitable stretch of time along with all the necessary information on the status of operation of the detector. The procedure was conceived for the data of the resonant bar detectors of the Rome group, so the stretch of time is 0.66 hours which yields a frequency resolution of 0.42 mHz, which is greater than the maximum expected frequency variation due to the Doppler modulation because of the motion of the Earth around the Sun, at 1 kHz, which is 0.28 mHz. Such a

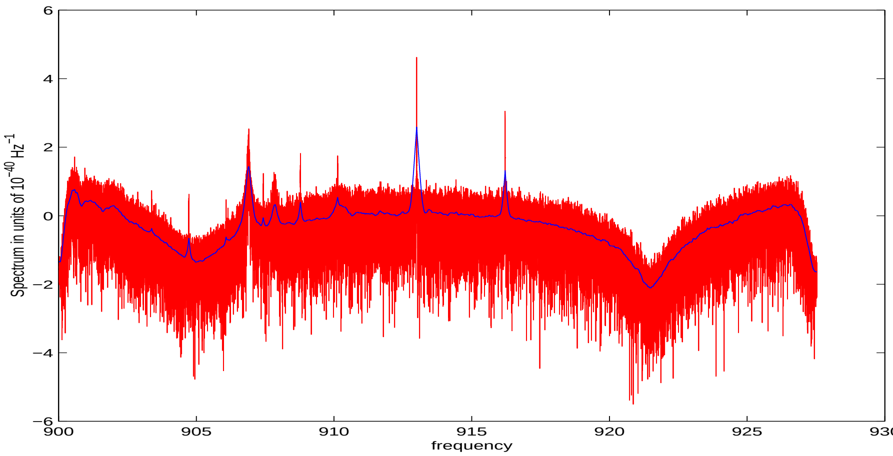

Figure 1: Spectrum and its average

FDDB may also be used in other types of searches, i.e. stochastic background, see P. Astone's talk at this meeting.

2 The first stage

In fig. **??** a spectrum computed from one of the FFTs of the FDDB is shown, and superimposed on it is its average computed with a Butterworth second order low pass filter with cut-off frequency $0.1 Hz$. More precisely fig.**??** shows the spectrum of a stochastic process that, by exciting the antenna, and in the absence of noise, would produce the noise spectrum that one observes at the output of the detector. The amplitudes are expressed in units of $10^{-40} Hz^{-1}$ so it is clear that in spectra of this kind, expected signals would be buried below the noise. In fact at the resonances the level of the noise would correspond to a monochromatic signal of amplitude h some 10^{-23}.

If a signal were present it would not show up as a significant peak in a single 0.66 hours spectrum, but it may be possible to identify it by examining more spectra since it may produce a peak in many of them. Thus, although *one* peak may not be significant its appearance in *many* spectra may be so.

So the first task consists in registering peaks. This is what we have called the "point registration procedure". It produces sets of time frequency points such as the ones show in fig. **??**, which refers to 55 spectra. The next step consists in searching such collections of points for patterns which may have originated from physical signals, possibly under rather loose constraints on what a signal should look like. This is the final part of the first stage procedure, and it has not not yet been developed. The main ideas for implementing it are essentially two:

one may adopt some kind of "matched filtering" scheme, by looking for well defined patterns that cover a suitably defined parameter space. Or one may use a heuristic approach, such as the ones implemented by chess games codes: rules are given that allow to give a score to any defined sequence of moves, which in our case means to any defined pattern. Patterns are picked as candidate when their score exceeds a predefined threshold.

3 Point registration

The procedure is being tested and optimized using 55 files ($\sim$ 36 hours) from the FDDB of the EXPLORER detector [?] in September 1991. From each of these files a spectrum Sp_i^j ($1 < j < 55$, $1 < i < 65536$) is produced (see fig. ??) containing 65536 bins. The distribution of the variables Sp_i^j for a fixed i is exponential but the expected value depends on the bin. In order to generate variables having the same distribution in every bin, each spectrum has been normalized to the average value over a neighbouring region by dividing bin per bin by the value of the averaged spectrum described previuosly and also shown in fig. ??. We shall work with this normalized spectrum. In addition to the demand that a point, to be stored, be a peak (a local maxima) one may also ask it exceeds some threshold. Some questions arise:

- how effective is this procedure at enhancing SNR, that is at selecting "signal-points" better than "noise-only-points" ?

- how does this effectiveness depend on signal amplitude (the procedure is non linear, thus it is not straightforward to assess performance versus signal amplitude) ?

- can it be optimized with the choice of a suitable threshold on the amplitude of the local maxima? and this question relates again to the previous one because the optimal threshold does depend on the minimum signal one wants to be able to detect.

In order to answer these questions we have started with the case of signals with constant frequency, i.e. sinusoids. The only free parameter is thus the amplitude of the sinusoid. The quantity that shall be studied is the "number of hits": i.e. how many signal points have been picked by the point registration procedure. It is necessary to compare the results one gets with the numbers one gets by chance. If the signals are sinusoids this is particularly easy because it consists in examining the occurence of straight lines in the time-frequency data. This study will provide answers to the questions outlined above since the number of hits does not depend on whether the pattern of points is alligned or not and it will set an upper bound to best performance that may be achieved with pattern tracing procedures.

Sinusoidal signals with different amplitudes $f \cdot A_{55}$ have been added (linearly) to the data. In order to have a good statistics for the same amplitude, sinusoids with different phases φ have been added: $f \cdot A_{55} \sin(2\pi\nu t + \varphi)$. A_{55} is the

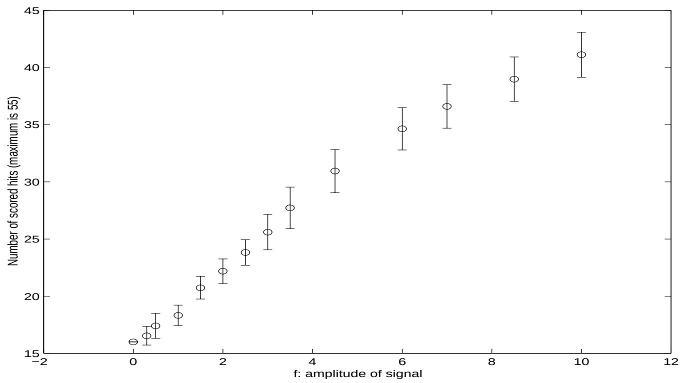

Figure 2: How many hits have been scored for signals of different amplitudes, f. Recall that the maximum number of hits is 55.

amplitude of a signal one would detect with SNR $\sim$ 1 if one added up the 55 power spectra. This corresponds to an amplitude SNR equal to $\frac{1}{\sqrt[4]{55}}$ in the single spectrum. With the point registration procedure one is not actually summing up the spectra (which hereafter we shall refer to as the "stacking method" - as opposed to "tracking"): the information from the 55 spectra is being pieced together in a different way; but one should expect that if this is done efficiently the gain be comparable - and in fact we will show it is. The reason for exploiting the point registration procedure is that, in general, the signal will not stay in a single bin, therefore to gain the maximum SNR one would have to sum spectra whose bins have been differently shifted, accordingly to some predefined source pattern. On the other hand we would like *to learn* from the data what the pattern is, thus to avoid any precise assumption on the signal[1].

Fig. **??** shows how many hits are to be expected (out of 55, which is the maximum possible value) for signals of different amplitudes with no constraint on the amplitude of the registered peak. What would one expect by chance ? This can be computed by considering the hit (when the point is registered) and the miss (when the point is not registered) as the outcomes of an experiment having only two possible results with probability p and $q = 1 - p$, respectively. It is then clear that the probability of getting n hits out of N trials by chance is easily computed by the binomial formula

$$p_{bin}(n, N) = \frac{N!}{n!(N-n)!} p^n q^{N-n}, \tag{1}$$

[1] Of course this will be the case if the pattern tracing procedure that one shall adopt is the heuristc one. In the other case summing shifted spectra or looking for defined patterns should be equivalent.

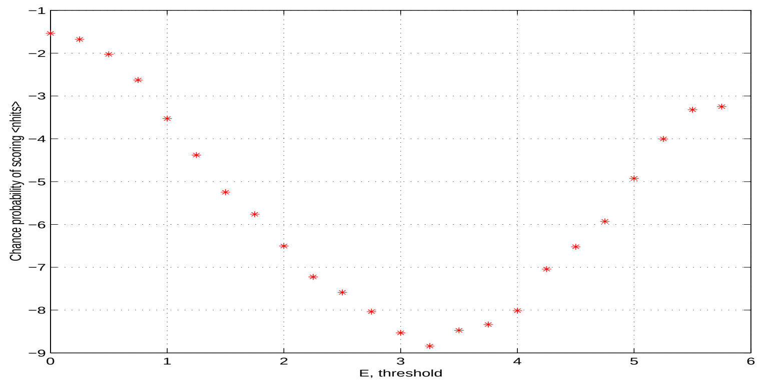

Figure 3: Chance probability versus threshold

with N=55, in this case.

For signals of different amplitudes it is possibile to choose an optimal threshold while selecting the peaks. This is done in the following way: having chosen a signal of a definite amplitude f one can compute how many peaks in the data where the signal is present are expected to lie above different thresholds, thr: $< nhits(thr) >$. On the other hand, by chosing different thresholds the probability per single extraction p varies: $p(thr)$. Thus for every threshold one can compute by eq. **??** the probability p_{bin} of getting by chance the number of hits one would score with that same threshold for the signal. The threshold that corresponds to the minimum value of this probability is the optimal threshold for a signal of amplitude f. In fig. **??** the behaviour of p_{bin} is plotted versus the threshold for a signal of amplitude $f = 3$. The optimal threshold is clearly observed at a threshold of $thr = 3.25$. The corresponding number of expected hits for the signal is 14.3. To assess the significance of this number one must look at the integral distribution of the average number of hits when only noise is present. It is then found that 14.3 hits are scored on the real data with a probability of $\simeq 7 \times 10^{-4}$, which is greater than that computed by the binomial formula because the data actually contains lines.

Is this optimization procedure effective ? In order to answer this question one should compare its performance with the performance of the "stacking method", which, as explained before, for monochromatic signals, is really the best one can do. On average the normalized spectrum in the signal bin yields a value of ~ 2.5. The probability of getting a value equal to or greater than that with noise only is $\sim 10^{-3}$. This is comparable with the performance of our procedure and actually shows than in this case the point registration method is even better.

As the signal becomes smaller the noise points that have to be registered in order to pick the signals ones, grow and the task of tracing them will be harder. What will eventually limit the smallest detectable signals is the complexity of the

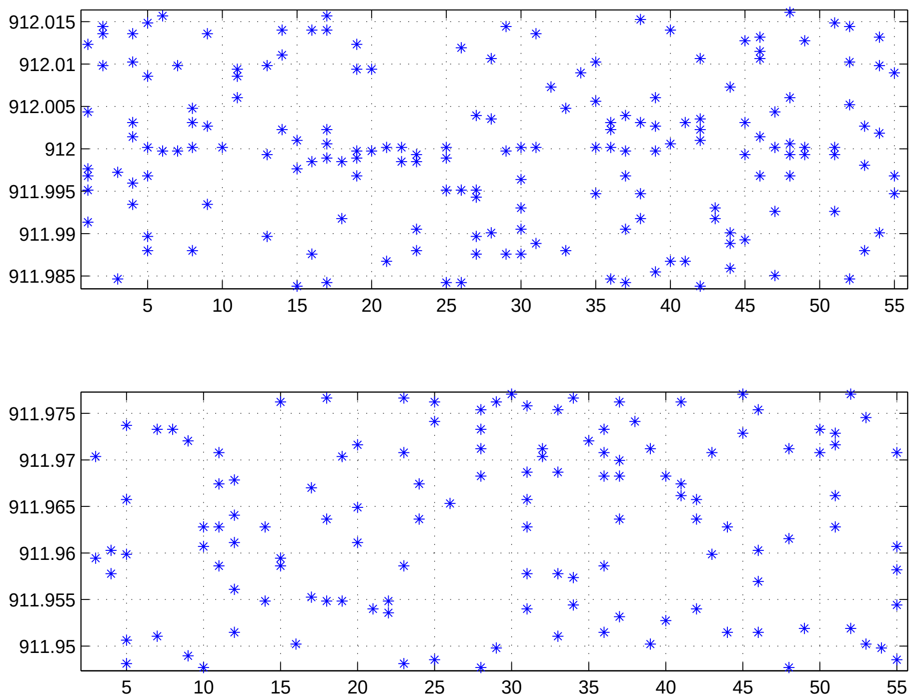

Figure 4: These are two zooms of the points registered by the point registration procedure optimized for a signal of amplitude f=3. Can you tell which is the one that contains the line pattern ?

time-frequency diagrams that will have to be searched. Optimizing at different signal amplitudes, which is what we have achieved with the method outlined above, is therefore crucial because it ensures the best environment for pattern tracing. Pattern tracing is the next problem we shall tackle.

References

[1] Brady P.R., Creighton T., Cutler C. & Schutz B.F., subm. to Phys. Rev D, available from http://xxx.lanl.gov/abs/gr-qc/970205

[2] Astone P., Frasca S., & M.A. Papa, 1997, in *Aspen Winter Conf.on gravitational wave detection*,January 1997, Aspen

[3] Astone P. et al., Phys. Rev. D, 47, (1993), p. 362