Mon. Not. R. Astron. Soc. **301,** 729–744 (1998)

# All-sky search algorithms for monochromatic signals in resonant bar gravitational wave detector data

J. A. Lobo* and M. Montero

*Departament de Física Fonamental, Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain*

Accepted 1998 July 29. Received 1998 July 1; in original form 1998 April 16

**ABSTRACT**

In this paper we design and develop several filtering strategies for the analysis of data generated by a resonant bar gravitational wave (GW) antenna, with the goal of assessing the presence (or absence) therein of long-duration monochromatic GW signals, as well as the eventual amplitude and frequency of the signals, within the sensitivity band of the detector. Such signals are most likely generated in the fast rotation of slightly asymmetric spinning stars. We develop practical procedures, together with a study of their statistical properties, which will provide us with useful information on the performance of each technique. The selection of candidate events will then be established according to threshold-crossing probabilities, based on the Neyman–Pearson criterion. In particular, it will be shown that our approach, based on phase estimation, presents a better signal-to-noise ratio than does pure spectral analysis, the most common approach.

**Key words:** gravitation – instrumentation: detectors – methods: data analysis – methods: statistical.

## 1 INTRODUCTION

It is generally believed that the most intense gravitational waves (GWs) arriving at the Earth from remote sources in the Universe correspond to very short duration (∼1 ms) bursts, generated in the explosion of a supernova (Thorne 1987), or in gamma-ray bursters (Roland, Frossati & Teyssier 1994). Since their very first origins, cylindrical bar GW antennas have been applied to the detection of this sort of event (Weber 1969), and the more modern cryogenic bars have also been used for this purpose, with considerably enhanced sensitivities (Astone et al. 1991; Astone et al. 1993; Hamilton et al. 1994): the long decay times of the oscillations of the bar make it well suited for the measurement of impulsive, short duration signals (Gibbons & Hawking 1971; Astone, Bonifazi & Pallottino 1990).

It so happens, however, that some cylindrical GW antennas have been in continuous operation for many consecutive months, even years. This is the case, for example, with the Explorer detector, owned and operated by the Ricerche Onde Gravitazionali (ROG) group in Rome (Italy) and installed within the CERN premises in Geneva, Switzerland (Astone et al. 1993). Long-term operation naturally provides the appropriate background for a search of *monochromatic* signals in the detector data, as the requisite long integration times become available.

Monochromatic signals are most probably generated by the rotation of asymmetric stars, such as a pulsar or a neutron star. The intensity of the GWs depends strongly on the amount of asymmetry of the source, and this is in turn dependent on its equation of state (Bonazzola & Gourgoulhon 1996). Reasonably optimistic upper bounds on typical star parameters give an extremely weak signal estimation of $h \sim 10^{-27}$ (Thorne 1987), which must be seen against a noisy background. Clearly, long integration times are required to reveal this kind of signal.

A systematic search for such a signal must face a practical difficulty which derives from the fact that the signal is received in the antenna Doppler-shifted as a result of the daily and yearly motions of the Earth – in addition to possible internal motions within the source if it is, e.g., in a binary system. Fourier analysis of long stretches of data results in high-frequency resolution (Kay 1990), thence in signal spread across several spectrum bins if it is Doppler shifted. This can naturally cause a significant reduction in the post-filter signal-to-noise ratio. The problem is easily overcome if the source position in the sky is known (or assumed) ahead of time by means of suitable corrections based on ephemeris calculations. Analyses of this type exist in the literature: traces of a pulsar in the centre of the supernova SN1987A were sought in four days of data generated by the 30-m Garching interferometer in 1989 March (Niebauer et al. 1993), and Frasca & La Posta studied almost four years of data generated by the room-temperature bar detector GEOGRAV in search of periodic signals from the Galactic Centre and the Large Magellanic Cloud (Frasca & La Posta 1991). More recently, Mauceli (Mauceli 1997) has looked for monochromatic GW signals coming

* E-mail: lobo@ffn.ub.es (JAL)

© 1998 RAS

from the region of Tuc 47 and from the Galactic Centre in three months of data generated by the cryogenic detector ALLEGRO at Louisiana State University.

A different strategy must of course be used for an all-sky search. The philosophy of the procedure put forward by Frasca & La Posta (Frasca & La Posta 1991) consists of the construction of a large *bank of spectra*, taken over shorter stretches of data such that the frequency resolution in each individual spectrum be sufficiently low that daily Doppler-shifted signals fit in a *single* spectral bin. Suitable comparison and averaging are thereafter applied to the spectra in order to draw conclusions about the intensity and/or bounds on signals. Astone et al. (Astone et al. 1997a; Astone 1998) have looked at one year (1991) of data taken by the above-mentioned Explorer detector in order to perform an all-sky search for monochromatic sources of GWs. Their method is based upon local maxima identification in a bank of spectra, followed by *close up* analyses of frequency peaks, looking for evidence of Doppler-shift patterns across the duration of the entire data set.

In this paper we design and develop algorithms for the analysis of data generated by a resonant bar detector of GWs, in search of monochromatic signals within the sensitive frequency band of the system. We are also interested in an all-sky search, but adopt a different point of view. Rather than scanning a bank of spectra, we propose to use a *matched filter* technique to estimate both frequency *and* phase of candidate signals. We then set a threshold, using the Neyman–Pearson criterion, to select those events which have a given probability of crossing it as a consequence of pure random noise fluctuations. We have tested our methods in simulations with real Explorer detector data from 1991, and have seen that they perform very satisfactorily. We plan to apply our methods to the massive processing of long stretches of data from the same antenna in a future paper, in order to provide complementary analyses to the procedures and methods already reported in Astone et al. (1997b) and Astone (1998).

The article is structured as follows. In Section 2 we present a few technical generalities and set the basic conventions of notation. Section 3 is devoted to a detailed study of a situation in which the signal has a frequency exactly equal to one of those in the discrete Fourier spectrum of the data (Lobo & Montero 1997); this corresponds to an idealized situation the consideration of which is methodologically useful, as it allows us to determine the phase of the signal, and to investigate the statistical properties of the filter output; it also characterizes the main guidelines for the more realistic study in subsequent sections. In Section 4 the method is illustrated with a signal artificially added to real detector data, which includes the estimation of the noise spectral density in the presence of such a signal. In Section 5 we address the real case, in which the signal frequency no longer exactly matches any of the discrete samples, so that it *leaks* across neighbouring spectrum bins (Lobo & Montero 1998), and also assess the statistical properties of the filter output (Montero 1998). Finally, in Section 6 we apply the method again to real data with an external control signal added, and show that it works satisfactorily. The paper closes with a summary of conclusions and future prospects.

## 2 LINEAR DATA FILTERING

We begin with a review of some fundamental concepts of linear data processing, fixing also the basic notation which we will be using throughout this article.

In the general case, let $\boldsymbol{u}(n)$ $(n = 0, \ldots, N-1)$ be the discrete set of samples which constitute our experimental data. A linear filter consists of a discrete set of numbers $\boldsymbol{g}(n; \mu_i)$ depending on several parameters, $\mu_i$, which acts on the experimental data as

$$\boldsymbol{y}(\mu_i) = \sum_{n=0}^{N-1} \boldsymbol{g}(n; \mu_i)\boldsymbol{u}(n), \tag{1}$$

producing what we shall call *the filter output*. It is usually assumed that $\boldsymbol{u}(n)$ is the sum of two different contributions: on the one hand is the signal, $\boldsymbol{x}(n)$, the presence of which we want to assess, represented by a deterministic function, and on the other hand the noise $\boldsymbol{r}(n)$, a stochastic process,

$$\boldsymbol{u}(n) = \boldsymbol{x}(n) + \boldsymbol{r}(n). \tag{2}$$

For any choice of parameters, it is appropiate to ask for the filter response both to the signal, $\boldsymbol{y}_x$, and to the noise, $\boldsymbol{y}_r$, the latter also being a stochastic process. The ratio of the mean square values of these quantities is called in the literature the output *signal-to-noise ratio* (SN),

$$\rho \equiv \frac{\boldsymbol{y}_x^2}{<\boldsymbol{y}_r^2>}, \tag{3}$$

and it is a measurement of the performance of the filter $\boldsymbol{g}(n; \mu_i)$. The theory of the matched filter (Helstrom 1968; Papoulis 1984) precisely determines, up to a global constant, the functional form which this must have, for given signal and noise, in order to maximize $\rho$.

In our case we shall assume that the noise $\boldsymbol{r}(n)$ can be adequately modelled by a zero-mean Gaussian and stationary stochastic process, whereas the signal $\boldsymbol{x}(n)$ will be the response of the cryogenic resonant detector Explorer to a pure monochromatic GW (Pallottino & Pizzella 1984; Montero 1997),

$$\boldsymbol{x}(n) = A_0 \cos(2\pi f_0 nT + \varphi_0). \tag{4}$$

Here, $A_0$ is the product of the amplitude by a conversion factor which defines the detector sensitivity at the frequency of the gravitational radiation, $f_{\rm g}$. This differs from $f_0$ by a constant shift (Frasca et al. 1992),

$$f_{\rm ini} = f_{\rm g} - f_0 = 900.026\,778\,1 \text{ Hz}, \tag{5}$$

introduced by the data acquisition system of the antenna, with the purpose of sampling the full bandwidth of the antenna (27.5088 Hz). The

matched filter for such a signal is then functionally equal to the latter (Helstrom 1968),

$$\boldsymbol{g}(n) = \mathcal{B}\cos(2\pi f_0 nT + \varphi_0), \tag{6}$$

where $\mathcal{B}$ is an arbitrary constant.

## 3 NON-LEAKING SIGNALS EMBEDDED IN KNOWN SPECTRUM NOISE

### 3.1 The signal

As noted in the Introduction, we want to develop in this article a general method, the operation of which does not depend on the existence of prior information about the source. So, in principle, the value of the frequency $f_0$ of the signal we can detect must be within the interval

$$0 \leq f_0 < \frac{1}{2T}, \tag{7}$$

where $1/T$ is our Nyquist frequency. Obviously no search strategy can afford the endless analysis of all the frequencies in that window, so we shall be forced to select a finite set of frequencies to scan. Nevertheless, the very functional form of the filter shows us that we shall perform discrete Fourier transforms (DFTs) in its implementation, and this defines the set of frequencies which will be searched in actual practice,

$$fT = \frac{k}{N}, \qquad k \in \{0, \ldots, N-1\}. \tag{8}$$

Moreover, for practical reasons, all the DFTs will be numerically computed using the fast Fourier transform (FFT) algorithm, a very optimized procedure which computes all spectral components at once, the only restriction being that the number of samples must be an exact power of 2.

In this section, we shall assume that the signal is *well matched* by the spectral template. By this we mean that the frequency of the signal is in fact one of those in equation (8), so that the entire signal is in one single bin of the FFT, with no *leakage* to the neighbouring ones. More precisely, we shall be assuming that

$$f_0 T = \frac{k_0}{N}, \tag{9}$$

where $k_0$ is one of $1, \ldots, N/2 - 1$, though we do not know which. We shall disregard the study of any $k_0$ bigger than $N/2$ because it would be redundant, since they represent nothing but negative frequencies. The value $k_0 = 0$ is also disregarded since, among other considerations, it represents, not a wave, but a constant signal.

Summing up, the target of the present analysis will be to assess the presence of a signal

$$\boldsymbol{x}(n) = A_0 \cos(2\pi k_0 n/N + \varphi_0) \tag{10}$$

in the experimental data series $\boldsymbol{u}(n)$, using a matched filter

$$\boldsymbol{g}(n; k, \varphi) = \mathcal{B}\cos(2\pi kn/N + \varphi), \tag{11}$$

depending on the two unknown paramenters, $k$ and $\varphi$, which we shall eventually estimate. Besides the advantageous property of the absence of frequency leakage in the filter output of such signals, equation (10) shows that $\boldsymbol{x}(n)$ is a periodic function over the entire processed period, because $\boldsymbol{x}(N)$ is equal to $\boldsymbol{x}(0)$. In fact, this relationship holds for any sample,

$$\boldsymbol{x}(n + N) = \boldsymbol{x}(n), \tag{12}$$

and it will be a crucial aspect of the developments which we shall introduce below.

### 3.2 The filter performace and the role of $\mathcal{B}$

Let us compute the two quantities $\boldsymbol{y}_x^2$ and $< \boldsymbol{y}_r^2 >$, in order to evaluate the actual goodness of the filter. $\boldsymbol{y}_x^2$ is different from zero only if a signal is really present *and* the value of the parameter $k$ matches $k_0$, i.e. the filtered signal does not leak across different frequency bins,

$$\boldsymbol{y}_x^2(k, \varphi) = \left[\frac{A_0 \mathcal{B} N}{2}\right]^2 \cos^2(\varphi - \varphi_0)\, \delta_{kk_0}. \tag{13}$$

For $< \boldsymbol{y}_r^2(k, \varphi) >$ we have the more complex formula,

$$< \boldsymbol{y}_r^2(k, \varphi) > = \frac{\mathcal{B}^2 N}{2} \sum_{n=-(N-1)}^{N-1} R(n)\left(1 - \frac{|n|}{N}\right)\cos(2\pi kn/N) - \mathcal{B}^2 \frac{\cos(2\pi k/N - 2\varphi)}{\sin(2\pi k/N)} \sum_{n=0}^{N-1} R(n)\sin(2\pi kn/N), \tag{14}$$

where we have introduced the autocorrelation function of the noise,

$$R(n) \equiv \langle \boldsymbol{r}(n' + n)\boldsymbol{r}(n')\rangle. \tag{15}$$

If we assume[1] that $R(n) \sim 0$ for $n \sim N$, it is clear that the second term shall become negligible in front of the first, and hence

$$< \boldsymbol{y}_r^2(k, \varphi) > \approx \mathcal{B}^2 \frac{NS(k; N)}{2T}, \tag{16}$$

[1] This, in fact, is a constraint on the number of samples that we want to filter. We must set $N$ large enough to sufficiently exceed the correlation time.

with

$$S(k;N) \equiv T \sum_{n=-(N-1)}^{N-1} R(n)\left(1-\frac{|n|}{N}\right)\cos(2\pi kn/N). \tag{17}$$

It can easily be shown that the quantity we have just defined is the mean value of a periodogramme,

$$S(k;N) = \frac{T}{N}\left\langle \left|\sum_{n=0}^{N-1} \boldsymbol{r}(n)\mathrm{e}^{-i2\pi kn/N}\right|^2 \right\rangle, \tag{18}$$

a well-known way for estimating the power spectral density of the noise at that particular frequency, based on the Wiener–Khinchine theorem (Kay 1990).

Putting together expressions (13) and (16) we finally get for the SN,

$$\rho = \frac{A_0^2 NT}{2S(k;N)}\cos^2(\varphi-\varphi_0)\,\delta_{kk_0} \equiv \rho_0\cos^2(\varphi-\varphi_0)\,\delta_{kk_0}, \tag{19}$$

with

$$\rho_0 \equiv \frac{A_0^2 NT}{2S(k;N)}, \tag{20}$$

the maximum value for $\rho$ we may achieve with the present filter.

The SN is obviously independent of the constant $\mathcal{B}$, so we can freely set it as we like in order to provide $\boldsymbol{y}$ with some advantageous property. Our particular choice is

$$\mathcal{B}(k) \equiv \sqrt{\frac{2T}{NS(k;N)}}, \tag{21}$$

the factor that makes $<\boldsymbol{y}_r^2(k,\varphi)>$ equal to one. The statistical properties of the noise and the linearity of the filter guarantee that $\boldsymbol{y}$ is still Gaussian. Then its probability distribution will be completely settled once we know its mean $<\boldsymbol{y}>$, and its variance, $\sigma_y^2$, which, in our case coincides with $<\boldsymbol{y}_r^2>$,

$$\sigma_y^2 \equiv <\boldsymbol{y}^2> - <\boldsymbol{y}>^2 = <\boldsymbol{y}_r^2> = 1. \tag{22}$$

So, on the one hand, we have forced $\sigma_y$ to take the same value regardless of the particular scanned frequency, and on the other hand, the mean of $\boldsymbol{y}(k,\varphi)$,

$$<\boldsymbol{y}(k,\varphi)> = \sqrt{\rho_0}\cos(\varphi-\varphi_0)\,\delta_{kk_0}, \tag{23}$$

shall be zero when either no signal is present in the data or, if there is a signal,[2] for any value of $k$ other than $k_0$. In this way, we have designed a bank of filters, the outputs of which corresponding to pure noise are statistically equivalent, and consequently *can be directly compared*.

### 3.3 Data splitting and averages

So far, we have implicitly assumed that $N$ represents the total amount of stored information we have access to or, in other words, that we can analyse the whole data series in a single filter pass. This is, in many senses, a rather optimistic assumption. First of all, since we want to process several months of experimental data, it should not be surprising that the available (or even existing) computing facilities could not afford such a calculation. Moreover, the output of any experimental device, like Explorer, will not be uniform in quality along all the data acquisition time, and the stationarity of the noise is not preserved over too-long periods of time. It could be worse to mix *bad* data (those, for instance, with a high level of noise) with *good* data in a single analysis, than simply to veto the stretch that we find unacceptable. However, the gaps that we may introduce in rejecting samples of the experimental set are not the unique discontinuities that we shall find in the time series, because in such a long-term operation of a detector it is not unlikely that the system will suffer sporadic stops. Also, the properties of the physical signal could be not so stable as to be satisfactorily fit by our models for extended periods of time.

We shall thus consider that each series of length $N$ is just one among a set of, say, $M$ consecutive[3] blocks (Astone et al. 1997a,b). The reasons for the choice of the particular values of $N$ and $M$ do not necessarily have to coincide, in general. In particular, it is possible that there exist several of those sequences of $N \times M$ data, eventually disconnected, which must be then processed separately.

So, we shall attach a new label $\alpha$ to each quantity in order to be able to specify which of the $M$ blocks of $N$ data we refer to:

$$\boldsymbol{y}_\alpha(k,\varphi) = \sqrt{\frac{2T}{NS(k;N)}}\sum_{n=0}^{N-1}\boldsymbol{u}_\alpha(n)\cos(2\pi kn/N+\varphi) \qquad (\alpha = 0,\ldots,M-1), \tag{24}$$

[2] In fact, we will put together both cases, since we shall take the criterion of typifying through $\rho_0 = 0$ any frequency that contains no signal, whether or not there is a signal at some other value of $k$.

[3] The relaxation of this condition, allowing for the existence of missing whole blocks, introduces minor changes in all the following discussion. So the derived expressions can be easily adapted to this case.

$\alpha$ is actually a shorthand which simplifies the notation,

$$\boldsymbol{u}_\alpha(n) \equiv \boldsymbol{u}(n + \alpha N). \tag{25}$$

It is obvious, however, that computing $\boldsymbol{y}_\alpha(k, \varphi)$ for each $\alpha$ will not change the individual values of $\rho$, as shown by equation (19). Our final goal should then be to combine them in a suitable way which allows us to make the final SN as high as possible. The definition (25) is, in this sense, very revealing, because when combined with (10) and (12) it shows the most important feature of a non-leaking monochromatic wave: the signal $\boldsymbol{x}_\alpha$ in fact does *not* depend on $\alpha$,

$$\boldsymbol{x}_\alpha(n) = A_0 \cos(2\pi k_0 n/N + \varphi_0) = \boldsymbol{x}(n). \tag{26}$$

We consequently see that, if the signal is present, each of these blocks contains an identical replica of the same stretch of sinusoid. This motivates us to define a new random variable $z(k, \varphi)$,

$$z(k, \varphi) \equiv \frac{1}{M} \sum_{\alpha=0}^{M-1} \boldsymbol{y}_\alpha(k, \varphi), \tag{27}$$

the mean value of which does not differ from that in equation (23), but the variance of which is reduced as a consequence of this averaging operation. Before we explicitly calculate this quantity and the value of the new associated SN, we are going to focus on the problem of choosing the right value of the phase parameter of the filter, $\varphi$.

The conceptually simplest method is to compute $z(k, \varphi)$ for many different values of that parameter, then select the best $\bar{\varphi}$, i.e. that which gives a larger output after the filtering procedure.

Nevertheless, we do not need to go into such a computationally long process, for the optimum value $\bar{\varphi}$ can be analytically determined as follows. According to its definition, we may write down $z(k, \varphi)$ as

$$z(k, \varphi) = \frac{1}{M} \sqrt{\frac{2T}{NS(k; N)}} \sum_{\alpha=0}^{M-1} [\mathrm{R}\{\tilde{\boldsymbol{u}}_\alpha(k)\} \cos\varphi + \mathrm{I}\{\tilde{\boldsymbol{u}}_\alpha(k)\} \sin\varphi], \tag{28}$$

where

$$\tilde{\boldsymbol{u}}_\alpha(k) \equiv \sum_{n=0}^{N-1} \boldsymbol{u}_\alpha(n)\, \mathrm{e}^{-2\pi \mathrm{i} kn/N} \tag{29}$$

is the DFT of $\boldsymbol{u}_\alpha(n)$. We define the $\bar{\varphi}$, imposing a local-maximum condition on $z(k, \varphi)$ of

$$\left.\frac{\partial z(k, \varphi)}{\partial \varphi}\right|_{\varphi=\bar{\varphi}} = 0. \tag{30}$$

We thus find[4]

$$\tan(\bar{\varphi}) = \frac{\sum_{\alpha=0}^{M-1} \mathrm{I}\{\tilde{\boldsymbol{u}}_\alpha(k)\}}{\sum_{\alpha=0}^{M-1} \mathrm{R}\{\tilde{\boldsymbol{u}}_\alpha(k)\}}. \tag{31}$$

Here it is useful to introduce the two random variables, $\mathbb{R}(k)$ and $\mathbb{I}(k)$,

$$\mathbb{R}(k) \equiv \frac{1}{M} \sqrt{\frac{2T}{NS(k; N)}} \sum_{\alpha=0}^{M-1} \mathrm{R}\{\tilde{\boldsymbol{u}}_\alpha(k)\}, \tag{32}$$

$$\mathbb{I}(k) \equiv \frac{1}{M} \sqrt{\frac{2T}{NS(k; N)}} \sum_{\alpha=0}^{M-1} \mathrm{I}\{\tilde{\boldsymbol{u}}_\alpha(k)\}, \tag{33}$$

because they can easily be combined to yield $z(k, \bar{\varphi})$,

$$\bar{z}(k) \equiv z(k, \bar{\varphi}) = \sqrt{[\mathbb{R}(k)]^2 + [\mathbb{I}(k)]^2}, \tag{34}$$

where we have defined $\bar{z}(k)$ as the best output of the filter at a given frequency, extending the notation used with $\varphi$.

The statistical properties of the actual filter output $\bar{z}(k)$ will differ strongly from those of $z(k, \varphi)$, since the new random variable is the fruit of a *non-linear* filtering process. For instance, its mean is no longer equal to zero, even if there is no signal in the experimental data, as we shall see. This is the reason why we did not undertake a very detailed study of $z(k, \varphi)$ in the first place.

### 3.4 Probability distribution of $\bar{z}(k)$

We shall build the probability density $p(\bar{z})$ starting from $p(\mathbb{R})$ and $p(\mathbb{I})$. Those two auxiliary random variables are Gaussian by construction, and

[4] We can ensure that $\bar{\varphi}$ defined in this way leads to a maximum and not a minimum of function $z(k, \varphi)$ because its second derivative, $\partial^2_\varphi z(k, \varphi = \bar{\varphi}) = -\bar{z}(k)$, is never positive.

are statistically independent, as they are the real and imaginary parts of the Fourier transform of a stationary stochastic process. Then we only have to know the respective means and variances in order to complete the information that will fully settle their probability densities. The mean values of $\mathbb{R}$ and $\mathbb{I}$ can be readily found from their definitions and equation (23):

$$< \mathbb{R}(k_0) > = \sqrt{\rho_0}\cos\varphi_0, \tag{35}$$

$$< \mathbb{I}(k_0) > = \sqrt{\rho_0}\sin\varphi_0, \tag{36}$$

while for the variances it can be found that both quantities are approximately equal, and

$$\sigma^2_{\mathbb{R}} \approx \sigma^2_{\mathbb{I}} \approx \frac{1}{M}, \tag{37}$$

because the correlation time is much shorter than the duration of the individual series, which is in essence equivalent to stating that, in spite of their consecutiveness, they are mutually uncorrelated.

The hypotheses we have made lead us to the following expressions for the probability density of $\mathbb{R}$ and $\mathbb{I}$:

$$p(\mathbb{R}) = \sqrt{\frac{M}{2\pi}}\ \exp\left[-\frac{M}{2}(\mathbb{R} - \sqrt{\rho_0}\cos\varphi_0)^2\right] \quad \text{and} \tag{38}$$

$$p(\mathbb{I}) = \sqrt{\frac{M}{2\pi}}\ \exp\left[-\frac{M}{2}(\mathbb{I} - \sqrt{\rho_0}\sin\varphi_0)^2\right]. \tag{39}$$

Hence $p(\bar{z})$ is given, after an integral is evaluated, by

$$p(\bar{z}) = M\bar{z}\exp\left[-\frac{M}{2}(\bar{z}^2 + \rho_0)\right] I_0(M\bar{z}\sqrt{\rho_0}), \tag{40}$$

where we have used one of the integral representations of the modified Bessel function of order zero (Olver 1970), $I_0(y)$,

$$I_0(y) \equiv \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-y\cos x}\mathrm{d}x. \tag{41}$$

In the absence of signal, i.e., when $\rho_0 = 0$, equation (40) reduces to

$$p(\bar{z}) = M\bar{z}\mathrm{e}^{-Mz^2/2}, \tag{42}$$

an expression which explicitly displays the statistical equivalence of all the frequencies which contain no signal, the property we want to achieve when we set the value of the constant $\mathcal{B}$. Moreover, in this case, $\bar{z}$ is Rayleigh distributed or, in other words, $\bar{z}^2$ follows a $\chi^2$ distribution with two degrees of freedom (Papoulis 1990).

### 3.5 Mean and variance of $p(\bar{z})$ – a new SN

We are now interested in the mean and variance of $\bar{z}$. These correspond to the first two moments of the probability distribution $p(\bar{z})$. It appears that a closed analytic expression can be found for the moments of any order, so we consider it here for completeness.

The $m$-th moment is defined by

$$< \bar{z}^m > \equiv \int_0^\infty M\bar{z}^{m+1}\exp\left[-\frac{M}{2}(\bar{z}^2 + \rho_0)\right] I_0(M\bar{z}\sqrt{\rho_0})\ \mathrm{d}z, \tag{43}$$

a calculation that becomes straightforward if one uses the relationship (Gradshteyn & Ryzhik 1980)

$$L_y(-z) = \frac{1}{\Gamma(y+1)}\int_0^\infty x^y \mathrm{e}^{-(x+z)} I_0(2\sqrt{xz})\mathrm{d}x, \tag{44}$$

where $L_y(z)$ is the Laguerre function of order $y$, assuming the normalization condition

$$L_y(0) = 1, \tag{45}$$

and $\Gamma(y)$ is Euler's *gamma* function. The expectation value of the $m$-th power of $\bar{z}$ is then

$$< \bar{z}^m > = \left(\frac{2}{M}\right)^{m/2}\Gamma\left(\frac{m}{2}+1\right)L_{m/2}\left(-\frac{M}{2}\rho_0\right). \tag{46}$$

The most relevant moments for our purpose are, as has been said, the first and the second. The mean,

$$< \bar{z} > = \sqrt{\frac{\pi}{2M}}L_{1/2}\left(-\frac{M}{2}\rho_0\right), \tag{47}$$

as we announced, will be different from zero even when $\rho_0$ vanishes, as a result of the property (45) of Laguerre functions,

$$< \bar{z} > |_{\rho_0=0} = \sqrt{\frac{\pi}{2M}}. \tag{48}$$

Nevertheless, the asymptotic behaviour of the Laguerre function,

$$L_y(-x) \overset{x \gg 1}{\longrightarrow} \frac{1}{\Gamma(y+1)} x^y \tag{49}$$

also shows that $< \bar{z} >$, when a signal is present, approaches the maximum mean value which according to (23) the random variables $\boldsymbol{y}_\alpha(k, \varphi)$ can possibly reach,

$$< \bar{z} > \approx \sqrt{\rho_0}. \tag{50}$$

It is worth noting that equation (49) ensures that the last expression holds not only when $\rho_0 \gg 1$ but when $\bar{\rho} \gg 1$, where we have introduced

$$\bar{\rho} = \frac{M}{2} \rho_0, \tag{51}$$

a quantity that plays the role of the new SN. The same conclusion can be obtained after the study of the explicit expressions for $\bar{z}_x^2$

$$\bar{z}_x^2 = \rho_0 \tag{52}$$

and $< \bar{z}_r^2 >$,

$$< \bar{z}_r^2 > = \frac{2}{M}. \tag{53}$$

It is relevant to point out that, in this case, SN linearly increases with the total number of filtered samples, $N \times M$:

$$\bar{\rho} = \frac{A_0^2 NMT}{4S(k;N)}, \tag{54}$$

whereas the more classical procedure of averaging the square of the moduli of the DFTs leads to

$$\bar{\rho}_0 = \frac{\sqrt{M}}{2} \rho_0 = \frac{A_0^2 N \sqrt{M} T}{4S(k;N)}, \tag{55}$$

which means that with our strategy for signals the frequency of which is one of the FFT samples, we enhance the value of $\bar{\rho}$ by a factor $\sqrt{M}$.

## 4 NON-LEAKING SIGNALS EMBEDDED IN UNKNOWN SPECTRUM NOISE

### 4.1 Replacing $S(k;N)$

It is almost redundant to say that the operative method we have just developed requires knowledge of the power spectral density of the noise. The aim of this section is the effective substitution of $S(k;N)$ in the definition of the constant $\mathcal{B}$ by a suitable estimate of this quantity obtained from the same data series.

Let us begin with a rearrangement of expression (18),

$$S(k;N) = \frac{T}{N} \langle [(\mathrm{R}\{\tilde{\boldsymbol{u}}_\alpha(k)\} - \mathrm{R}\{\tilde{\boldsymbol{x}}(k)\})^2 + (\mathrm{I}\{\tilde{\boldsymbol{u}}_\alpha(k)\} - \mathrm{I}\{\tilde{\boldsymbol{x}}(k)\})^2] \rangle. \tag{56}$$

There is one procedure in this formula that is certainly beyond our control: we cannot perform the statistical average. Our particular choice will be the substitution of that operation by a sum over the the entire rank of values of $\alpha$, because $S(k;N)$, in spite of the formal aspect of (56), is *independent of the block label*. The same applies upon replacement of $\tilde{\boldsymbol{x}}(k)$ (obviously also an unknown quantity), since $\tilde{\boldsymbol{x}}(k) = < \tilde{\boldsymbol{u}}(k)_\alpha >$,

$$\tilde{\boldsymbol{x}}(k) \rightarrow \frac{1}{M} \sum_{\alpha=0}^{M-1} \tilde{\boldsymbol{u}}_\alpha(k). \tag{57}$$

Summing up, the random variable we shall use in order to estimate $S(k;N)$ is $\mathbb{S}(k;N)$,

$$\mathbb{S}(k;N) \equiv \frac{T}{N} \frac{1}{M-1} \sum_{\alpha=0}^{M-1} \left[ \left( \mathrm{R}\{\tilde{\boldsymbol{u}}_\alpha(k)\} - \frac{1}{M} \sum_{\alpha'=0}^{M-1} \mathrm{R}\{\tilde{\boldsymbol{u}}_{\alpha'}(k)\} \right)^2 + \left( \mathrm{I}\{\tilde{\boldsymbol{u}}_\alpha(k)\} - \frac{1}{M} \sum_{\alpha'=0}^{M-1} \mathrm{I}\{\tilde{\boldsymbol{u}}_{\alpha'}(k)\} \right)^2 \right] \tag{58}$$

where we have divided by $M-1$ and not by $M$ because this way we get an unbiased estimator, i.e.,

$$< \mathbb{S}(k;N) > = S(k;N). \tag{59}$$

Now we can replace the unknown spectral density by $\mathbb{S}(k;N)$ in any preceding expression, thus obtaining a new filter output $\mathbb{Z}(k)$ which, unlike $\bar{z}(k)$, we are able to compute directly from the raw experimental data $\boldsymbol{u}_\alpha(n)$. With a procedure analogous to the one already explained, we obtain $p(\mathbb{Z})$ and all its related quantities, including the corresponding SN. Instead of starting from scratch, we shall calculate the probability density of $\mathbb{Z}$ in two steps, using previous results.

Let us introduce the auxiliary random variable $\mathbb{W}(k)$,

$$\mathbb{W}(k) \equiv \frac{\mathbb{S}(k;N)}{S(k;N)}, \tag{60}$$

which allows us to define $\mathbb{Z}(k)$ in a simple way,

$$\mathbb{Z} \equiv \frac{\bar{z}}{\sqrt{\mathbb{W}}}, \tag{61}$$

thanks to the fact that all the terms containing $S(k;N)$ in both random variables *mutually* cancel out. Since $\mathbb{W}$ and $\bar{z}$ are statistically independent, and $p(\bar{z})$ was given in the last section, we have thus reduced the problem to obtaining $p(\mathbb{W})$ and performing a final integration.

The probability density of $\mathbb{W}$ can be found in most reference books on probability (Papoulis 1990), because it is the arithmetic mean of the squares of $2M-2$ zero-mean independent Gaussian variables with unit variances. So $(2M-2)\,\mathbb{W}$ follows a $\chi^2$ distribution with precisely $2M-2$ degrees of freedom,

$$p(\mathbb{W}) = \frac{(M-1)^{M-1}}{\Gamma(M-1)}\mathbb{W}^{M-2}\mathrm{e}^{-(M-1)\mathbb{W}}. \tag{62}$$

### 4.2 The actual filter output $\mathbb{Z}$ and its distribution

The ratio in (61) which we have used for defining $\mathbb{Z}$ is a familiar one in elementary statistics, and is known to follow a Student's $t$-distribution if $\bar{z}$ is a zero-mean Gaussian variable. Nevertheless, the expression for $p(\bar{z})$ is far from a normal density function, as we have shown in (40), which compels us to perform an explicit calculation, leading to the result

$$p(\mathbb{Z}) = \frac{M\mathbb{Z}}{\left[1+\frac{M}{2M-2}\mathbb{Z}^2\right]^M}\exp\left(-\frac{M\rho_0}{2+\frac{M}{M-1}\mathbb{Z}^2}\right)L_{M-1}\left(-\frac{M^2\mathbb{Z}^2\rho_0}{4M-4+2M\mathbb{Z}^2}\right), \tag{63}$$

where once again we have made use of the formula (44). When no signal is present in the data at one particular frequency, expression (63) reduces to

$$p(\mathbb{Z}) = \frac{M\mathbb{Z}}{\left[1+\frac{M}{2M-2}\mathbb{Z}^2\right]^M}. \tag{64}$$

As a matter of fact, $\mathbb{Z}^2$ in this case follows a Fisher's $F$-distribution with 2 and $2M-2$ degrees of freedom, because it is the ratio of two $\chi^2$ random variables with those degrees of freedom, respectively.

In order to compute the moments of the density function of $\mathbb{Z}$, the simplest approach is not to use the final expression of $p(\mathbb{Z})$ but an intermediate formula,

$$p(\mathbb{Z}) = M\mathbb{Z}\frac{(M-1)^{M-1}}{\Gamma(M-1)}\mathrm{e}^{-(M/2)\rho_0}\int_0^\infty \mathbb{W}^{M-1}\mathrm{e}^{-\mathbb{W}[M-1+(M/2)\mathbb{Z}^2]}I_0(M\mathbb{Z}\sqrt{\rho_0\mathbb{W}})\mathrm{d}\mathbb{W}, \tag{65}$$

that will avoid the problem of the integration of Laguerre functions with negative arguments. We thus find

$$<\mathbb{Z}^m> = \left[(M-1)^{m/2+1}\frac{\Gamma(M-1-m/2)}{\Gamma(M)}\right]\left(\frac{2}{M}\right)^{m/2}\Gamma\left(\frac{m}{2}+1\right)L_{m/2}\left(-\frac{M}{2}\rho_0\right), \tag{66}$$

where we have chosen a layout that emphasizes the resemblance with the result corresponding to $<\bar{z}^m>$. The term inside the square brackets approaches unity when $M \gg m$, and we then recover the formula (46). It is especially interesting to note that, in particular, the newly defined SN remains unchanged. Let us split the second-order moment of $p(\mathbb{Z})$,

$$<\mathbb{Z}^2> = \frac{M-1}{M-2}\frac{2}{M}\left(1+\frac{M}{2}\rho_0\right), \tag{67}$$

into two terms, namely $<\mathbb{Z}_r^2>$ and $<\mathbb{Z}_x^2>$,

$$<\mathbb{Z}^2> = <\mathbb{Z}_r^2> + <\mathbb{Z}_x^2>. \tag{68}$$

When no signal is present, $\rho_0 = 0$, the value of $<\mathbb{Z}^2>$ is merely a result of the response of the noise to the filtering procedure,

$$<\mathbb{Z}_r^2> = \frac{M-1}{M-2}\frac{2}{M}, \tag{69}$$

so we will accordingly assign to the signal the rest of the outcome,

$$<\mathbb{Z}_x^2> = \frac{M-1}{M-2}\rho_0, \tag{70}$$

and therefore,

$$\bar{\rho} = \frac{<\mathbb{Z}_x^2>}{<\mathbb{Z}_r^2>} = \frac{M}{2}\rho_0. \tag{71}$$

### 4.3 A practical example

This section is devoted to showing the result of such a procedure when applied to a small stretch of data taken by the Explorer on 1991 August 3,

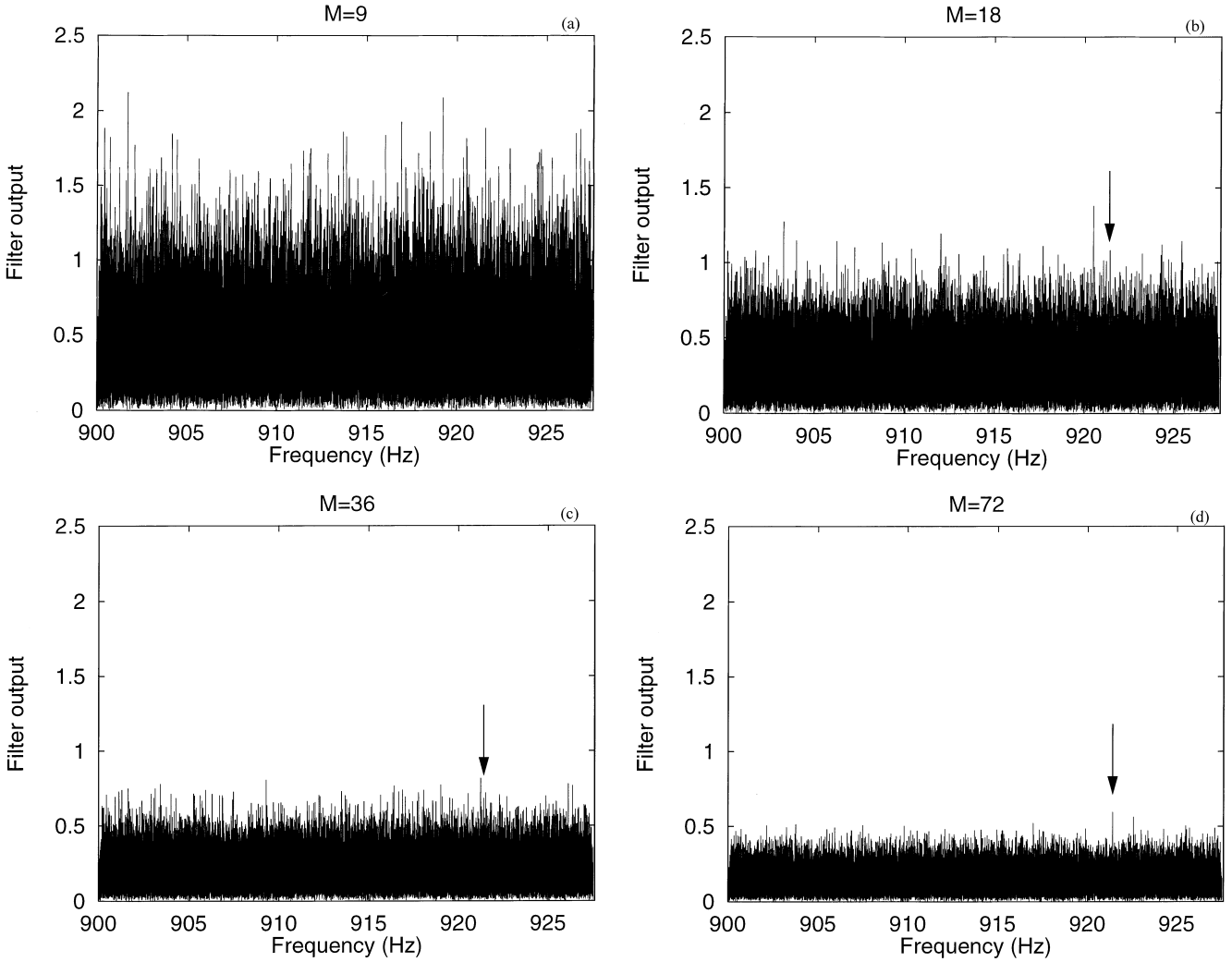


**Figure 1.** Output of the analysis procedure when the filtered data stretches extend over: (a) six hours, (b) twelve hours, (c) a day and (d) two days. Every time series begins at the same instant of 1991 August 3. The arrows point to the signal in each of the last three cases. Even though they are progressively more prominent, only in (d) is the signal the highest peak. In (c), for example, it is hidden between two taller noise fluctuations.

and on successive days. The starting date was randomly selected, since the final focus of the present analysis is not on extracting conclusions from the presence of GWs, but on the practical performance of the method itself. We have thus externally introduced a sinusoidal signal with the required absence of frequency leakage in order to check the ability of the method for revealing it. The signal, corresponding to a GW with an amplitude of $h_0 = 10^{-23}$, was placed at about 921.4 Hz, near the detector's *plus* resonance (Astone et al. 1993). For this particular date, the level of noise in the detector was such that the SN for this signal was $\rho_0 \sim 1/3$ for a number of filtered samples of $N = 131072$, a little less than forty minutes. The signal was therefore completely buried in the noise. The specific value $N = 131072$ may seem arbitrary in this context, but it has a physical reason: it ensures that no Doppler shift can be observed in the individual blocks of $N$ samples (Astone et al. 1997b). Once we set $N$ we can pin down the precise frequency bin which contains the control signal, $k_0 = 50918$.

Let us see what happens when we process six hours of data ($M = 9$). Looking at Fig. 1(a), it is by all means impossible to decide whether the signal is really present or not: the SN has only risen to a value near unity from the original 1/3 with few blocks processed. By increasing $M$, i.e. processing longer stretches of data, the SN grows and the signal becomes progressively more distinct, as we see in Figs 1(b)–(d), corresponding to half a day, one day and two days of filtered data, respectively. We must stress at this point that the theoretical prediction of equation (51), that energy SN grows linearly with the number $M$ of processed blocks, is very accurately observed in real practice, as we have numerically verified with the plotted data. The improvement by a factor of $\sqrt{M}$ relative to more standard procedures – see equation (55) – is thus firmly established not only in theory, but also in actual practice.

With the output of the filtering process for the values of $k$ other than $k_0$, we can compute[5] the distribution of $\mathbb{Z}$, because when no signal is present it does not depend on $k$. The case $M = 9$ is again very interesting because it offers us the possibility of comparing the experimental distribution with $p(\bar{z})$ and $p(\mathbb{Z})$, thus checking that $\mathbb{Z}$ really follows the second and not the first – see Fig. 2(a). For higher values of $M$, the two

[5] In fact, we do not need to remove $k_0$ (or any other presumed signal) necessarily before we compute the output distribution, because at most it is only one point in $2^{16}$.

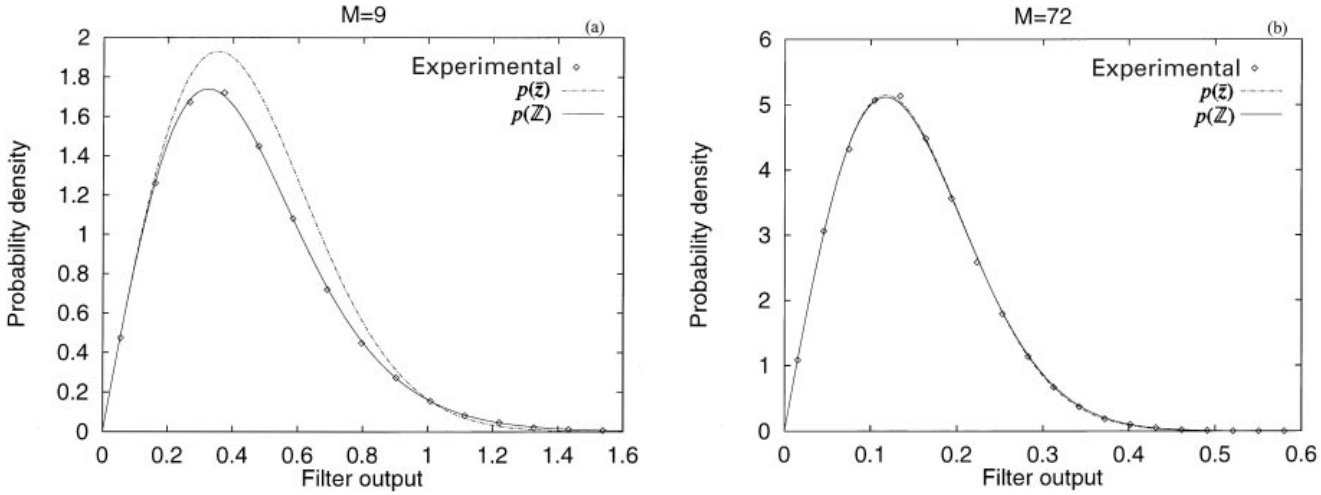


**Figure 2.** Comparison of the experimental output distribution of $\mathbb{Z}$ with $p(\bar{z})$ and $p(\mathbb{Z})$, both in absence of signal, when (a) $M = 9$ and (b) $M = 72$.

**Table 1.** Value of the threshold $\lambda_0$ corresponding to different choices of $\mathcal{Q}_0$ for each of the considered block number $M$.

| $\mathcal{Q}_0$ | $M = 9$ | $M = 18$ | $M = 36$ | $M = 72$ |
|---|---|---|---|---|
| $10^{-3}$ | 1.56 | 0.97 | 0.65 | 0.45 |
| $10^{-5}$ | 2.39 | 1.35 | 0.87 | 0.59 |

probability densities converge and become almost indistinguishable from one another (Fig. 2b). In both instances, however, the coincidence between theoretical and experimental distribution is remarkable.

Moreover, the explicit form of $p(\mathbb{Z})$ is very useful, not only in order to compare it with the experimental one, but also in order to fix a threshold $\lambda_0$, which will help us in the task of deciding whether a given crossing has statistical significance not. We will calculate the *error of the first kind*, or false-alarm probability, $\mathcal{Q}_0$, as a function of $\lambda_0$,

$$\mathcal{Q}_0 \equiv \int_{\lambda_0}^{\infty} p(\mathbb{Z})\mathrm{d}\mathbb{Z} = \frac{1}{\left[1 + \frac{M}{2M-2}\lambda_0^2\right]^{M-1}}. \tag{72}$$

and then we shall set the upper bound depending on the number of *false alarms* (i.e. mistakes) we can afford, using the well-known Neyman–Pearson criterion (Helstrom 1968):

$$\lambda_0 = \sqrt{\frac{2M-2}{M}\left(\frac{1}{\mathcal{Q}_0^{1/(M-1)}} - 1\right)}. \tag{73}$$

Table 1 shows how $\lambda_0$ varies both with respect to the value of $\mathcal{Q}_0$ and to the number of processed data blocks, $M$. The height of the signal, in units of the plots in Fig. 1, is 0.473, 0.684 and 0.595 for $M = 18$, 36 and 72, respectively. It is therefore above threshold if $M = 72$ with false alarm probability $\mathcal{Q}_0 = 10^{-5}$, and also if $M = 36$ and $M = 72$ with false alarm probability $\mathcal{Q}_0 = 10^{-3}$. Even so it cannot be clearly distinguished from other random fluctuations, as we see in Fig. 1. This is because it is very weak, of course. We shall come back to a discussion of the significance of these thresholds below.

## 5 LEAKING SIGNALS EMBEDDED IN KNOWN SPECTRUM NOISE

### 5.1 A leaking signal

We are going to start this section by considering the effects that the presence of a general frequency signal in the data may produce in the results we have given in the preceding sections. In particular, we shall study the new statistical properties of the random variables $\boldsymbol{y}_\alpha(k, \varphi)$, that determine the characteristics of $\mathbb{R}$ and $\mathbb{I}$, and consequently of $\bar{z}$.

So, in the following we shall relax the condition (9),

$$f_0 T = \frac{k_0 + \epsilon_0}{N}, \tag{74}$$

by introducing the real parameter $\epsilon_0$,

$$-\frac{1}{2} \leq \epsilon_0 < \frac{1}{2}, \tag{75}$$

the effective consequence of which is that $\boldsymbol{x}_\alpha(n)$ is no longer independent of $\alpha$, which appears in the form of an accumulative phase shift whenever $\epsilon_0$ is different from zero, i.e.

$$\boldsymbol{x}_\alpha(n) = A_0 \cos(2\pi[k_0 + \epsilon_0]n/N + 2\pi\alpha\epsilon_0 + \varphi_0). \tag{76}$$

While it is true that the frequency remainder also lowers the maximum filter output, as a result of the spectral leakage of the signal,

$$< \boldsymbol{y}_\alpha(k_0, \varphi) >= \sqrt{\rho_0}\frac{\sin(\pi\epsilon_0)}{N\sin(\pi\epsilon_0/N)}\cos\left(\pi\epsilon_0\left[1 - \frac{1}{N}\right] + \varphi_0 - \varphi + 2\pi\alpha\epsilon_0\right), \tag{77}$$

it is nevertheless the block dependence that damages the filtering procedure, because it is responsible for the sinc-like behaviour of the mean values of $\mathbb{R}(k_0)$ and $\mathbb{I}(k_0)$, when considered as functions of $M$:

$$< \mathbb{R}(k_0) > = \sqrt{\rho_0}\frac{\sin(\pi\epsilon_0 M)}{NM\sin(\pi\epsilon_0/N)}\cos\left(\pi\epsilon_0\left[M - \frac{1}{N}\right] + \varphi_0\right), \tag{78}$$

$$< \mathbb{I}(k_0) > = \sqrt{\rho_0}\frac{\sin(\pi\epsilon_0 M)}{NM\sin(\pi\epsilon_0/N)}\sin\left(\pi\epsilon_0\left[M - \frac{1}{N}\right] + \varphi_0\right). \tag{79}$$

This fact means that the search strategy is not as robust as we would like, in the sense that for a given $\epsilon_0$ other than zero, there always exists a value $M_0$ ($\sim 1/\epsilon_0$) for $M$, above which the SN decreases noticeably. The frequency band where the analysis method works efficiently thus decreases with the number of averaged blocks, a very undesirable feature. We now investigate how this problem can be addressed.

### 5.2 Phase-varying filter

As already stated, since the origin of the problem is a carried-over phase, we shall solve it by introducing a new block-dependent filter with one more parameter, $\epsilon$,

$$\boldsymbol{g}_\alpha(n; k, \varphi, \epsilon) = \sqrt{\frac{2T}{NS(k;N)}}\cos(2\pi kn/N + \varphi + 2\pi\alpha\epsilon), \tag{80}$$

with the intention of compensating for the phase shift. Starting from this new $\boldsymbol{g}_\alpha(n; k, \varphi, \epsilon)$, we can calculate the value of each $\boldsymbol{y}_\alpha(k, \varphi, \epsilon)$,

$$\boldsymbol{y}_\alpha(k, \varphi, \epsilon) = \sqrt{\frac{2T}{NS(k;N)}}|\tilde{\boldsymbol{u}}_\alpha(k)|\cos[\Phi_\alpha(k) - \varphi - 2\pi\alpha\epsilon], \tag{81}$$

where the following notation has been used:

$$\tilde{\boldsymbol{u}}_\alpha(k) = |\tilde{\boldsymbol{u}}_\alpha(k)|\mathrm{e}^{\mathrm{i}\Phi_\alpha(k)}. \tag{82}$$

Through a definition formally identical to (27), we shall establish $z(k, \varphi, \epsilon)$. Once again it is possible to obtain $\bar{\varphi}$ using a local-maximum condition, like that in (30),

$$\left.\frac{\partial z(k, \varphi, \epsilon)}{\partial \varphi}\right|_{\varphi=\bar{\varphi}} = 0, \Rightarrow \tan(\bar{\varphi}) = \frac{\sum_{\alpha=0}^{M-1}|\tilde{\boldsymbol{u}}_\alpha(k)|\sin[\Phi_\alpha(k) - 2\pi\alpha\epsilon]}{\sum_{\alpha=0}^{M-1}|\tilde{\boldsymbol{u}}_\alpha(k)|\cos[\Phi_\alpha(k) - 2\pi\alpha\epsilon]}. \tag{83}$$

The value of $\bar{\varphi}$ leads now to the following expression for $z(k, \bar{\varphi}, \epsilon)$,

$$z(k, \bar{\varphi}, \epsilon) = \frac{1}{M}\sqrt{\frac{2T}{NS(k;N)}}|\tilde{\tilde{\boldsymbol{u}}}(k;\epsilon)|, \tag{84}$$

a relationship that involves a new quantity, $\tilde{\tilde{\boldsymbol{u}}}(k;\epsilon)$, which formally is also a DFT,

$$\tilde{\tilde{\boldsymbol{u}}}(k;\epsilon) \equiv \sum_{\alpha=0}^{M-1}\tilde{\boldsymbol{u}}_\alpha(k)\mathrm{e}^{-2\pi\mathrm{i}\alpha\epsilon}. \tag{85}$$

The template for $\epsilon$, just as in the case of the frequency grid, will be dictated by the convenience of the use of the FFT algorithm in the computation of $\tilde{\tilde{\boldsymbol{u}}}(k;\epsilon)$. We shall therefore estimate $\epsilon_0$ within the following discrete rank of values[6] of $\epsilon$:

$$\epsilon = \frac{q}{M'} \qquad q \in \{0, \ldots, M' - 1\}, \tag{86}$$

where, in principle, $M'$, by the way an exact power of 2, must not necessarily coincide with $M$. $M'$ has to be greater than $M$ if we do not want to waste available information, but on the other hand, it seems that larger and larger values of $M'$ should produce an endless increment of precision in the estimation of $\epsilon_0$. Nevertheless, as we will show in the next section, $M'$ should be kept as low as possible.

Equation (85), together with the span condition (86), bears a considerable formal resemblance with the so-called *zoom transform* (Yip 1976; Hung 1981; de Wild, Nieuwkerk & van Sinttruyen 1987), in which the *twiddle* factor is now identically equal to one. It must however be

[6] As usual, any value for $q$ equal to or larger than $M'/2$ is related to a negative $\epsilon$.

stressed that, in the present context, ours is an *interpolation* formula rather than a frequency resolution algorithm. Our formalism can thus be naturally extended with the purpose of refining the spectral resolution, simply taking $M' = M$, i.e., data sets of $N \times M'$ points, where all possible values of $q$ must be considered, for a selected choice of frequency bins, $k$.

### 5.3 Statistical properties of $\bar{z}$

The mechanism for finding $\bar{q}$ is then very simple. We must compute all the $z(k, \bar{\varphi}, q/M')$ and select that $q$ which gives the largest output, just defining $\bar{z}(k) = z(k, \bar{\varphi}, \bar{q}/M')$, i.e.,

$$\bar{z}(k) = \max_{q \in \{0,\ldots,M'-1\}} \{z(k, \varphi, q/M')\}. \tag{87}$$

When no signal is present at some frequency, it can be proved that again,

$$p(z) = Mz\mathrm{e}^{-(M/2)z^2}, \tag{88}$$

no matter what the choices are for $k$ and $q$. So, in that case, the probability density of $\bar{z}(k)$, since it is the maximum of $M'$ equally distributed random variables (Papoulis 1990), is given by

$$p(\bar{z}) = MM'\bar{z}\mathrm{e}^{-(M/2)\bar{z}^2}\left[1 - \mathrm{e}^{-(M/2)\bar{z}^2}\right]^{M'-1}. \tag{89}$$

We have not been able to compute $<\bar{z}>$ other than in the form of an alternating finite series, which is almost useless for obtaining generic conclusions about it. Instead of the mean value of $\bar{z}$ we shall compute its most probable value. The function $p(\bar{z})$ reaches its maximum (when $M' \gg \sqrt{\mathrm{e}}$) for

$$\bar{z} \approx \sqrt{\frac{2}{M}\ln M'}, \tag{90}$$

a quantity that decreases as $M$ increases, and when $M'$ decreases. This shows the convenience of setting $M'$ as the first exact power of 2 greater than $M$. With the second-order moment of the distribution we have in principle a similar problem, although in this case the alternating series can be transformed into a non-alternating one,

$$<\bar{z}^2> = \frac{2}{M}\sum_{l=1}^{M'}\binom{M'}{l}\frac{(-1)^{l+1}}{l} = \frac{2}{M}\sum_{l=1}^{M'}\frac{1}{l}. \tag{91}$$

For large values of $M'$ we can approximate the result using the definition of Euler's $\gamma$ constant,

$$<\bar{z}^2> \approx \frac{2}{M}\left(\ln M' + \gamma\right). \tag{92}$$

Since the output of the procedure in the absence of noise is

$$\bar{z}_x = \sqrt{\rho_0}\frac{\sin(\pi\epsilon_0)}{N\sin(\pi\epsilon_0/N)}, \tag{93}$$

the new SN shall be (if $M' \gg 1$)

$$\bar{\rho} = \rho_0\left[\frac{\sin(\pi\epsilon_0)}{N\sin(\pi\epsilon_0/N)}\right]^2\frac{M}{2\left(\ln M' + \gamma\right)}. \tag{94}$$

The standard procedure of averaging the square of the moduli of the DFTs gives, for a general leaking signal, the following SN:

$$\bar{\rho}_0 = \rho_0\left[\frac{\sin(\pi\epsilon_0)}{N\sin(\pi\epsilon_0/N)}\right]^2\frac{\sqrt{M}}{2}, \tag{95}$$

which represents an improvement with the present method of

$$\frac{\bar{\rho}}{\bar{\rho}_0} = \frac{\sqrt{M}}{\ln M' + \gamma}, \tag{96}$$

a ratio larger than one when (92) holds, which increases with the number of blocks.

## 6 LEAKING SIGNALS EMBEDDED IN UNKNOWN SPECTRUM NOISE

### 6.1 A spectral estimator

The task of replacing $S(k; M)$ in the filter definition is much more complex than in the non-leaking case. The natural starting point is the value of $\tilde{\tilde{\boldsymbol{u}}}(k; q/M')$ for $q$ other than $\bar{q}$, but this presents two main problems. First of all, the fact that we are only able to choose $\bar{q}$ out of a discrete set leaves open the unpleasant possibility that the value of the signal frequency lies just in the middle of a bin. This means that we shall ignore the precise way in which the signal energy will be distributed among the different $q$'s, and thus the relative magnitude of $\tilde{\tilde{\boldsymbol{u}}}(k; \bar{q}/M')$, when compared with the rest of outputs. The scenario can be even worse, because we have no guarantees that $\bar{q}/M'$ corresponds to the best approach

to the actual $\epsilon_0$: the noise can mislead us into an inaccurate value of the signal frequency, leaving thus the *true* $\tilde{\tilde{\boldsymbol{u}}}(k;\epsilon_0)$ among the discarded ones.

Instead of constructing an estimator for $S(k;M)$ in the hope that none of the previously stated possibilities really takes place, which could lead us again to a filtering procedure too sensitive to the signal's peculiar properties, we are going to choose a *democratic* estimate $\mathbb{S}(k;N)$: perhaps it will not be as accurate as it might be, but it will not show appreciable differences in its performance depending on the actual frequency of the GW.

We define $\mathbb{S}(k;N)$ through an expression that closely resembles that in equation (58),

$$\mathbb{S}(k;N) \equiv \frac{T}{NM}\frac{1}{M'-2}\sum_{q\neq\bar{q}}\left[\left(\mathrm{R}\{\tilde{\tilde{\boldsymbol{u}}}(k;q/M')\}-\frac{1}{M'-1}\sum_{q'\neq\bar{q}}\mathrm{R}\{\tilde{\tilde{\boldsymbol{u}}}(k;q'/M')\}\right)^2+\left(\mathrm{I}\{\tilde{\tilde{\boldsymbol{u}}}(k;q/M')\}-\frac{1}{M'-1}\sum_{q'\neq\bar{q}}\mathrm{I}\{\tilde{\tilde{\boldsymbol{u}}}(k;q'/M')\}\right)^2\right], \tag{97}$$

where the $\tilde{\boldsymbol{u}}_\alpha(k)$ have been substituted by $\tilde{\tilde{\boldsymbol{u}}}(k;q/M')$, and the sums do not contain the term where the signal is supposed to be located, $\tilde{\tilde{\boldsymbol{u}}}(k;\bar{q}/M')$.

Once more we replace $S(k;N)$ by $\mathbb{S}(k;N)$ in the definition of $\bar{z}$, in order to get a new random variable $\mathbb{Z}$ which we can compute using only experimental data. This quantity inherits two characteristic traits from the way we estimate the spectral density of the noise.

On the one hand, if the signal is large enough, the filter output may have a saturation limit, which will depend upon the particular values of some parameters, such as $M$ or $M'$. This means that $\mathbb{Z}$ will not surpass a certain threshold, even if the amplitude of the signal increases indefinitely. The reason for this behaviour can be found in the fact that when the signal is much more intense than the noise, what becomes really difficult to assess is not the presence of the former but the properties of the latter. So in these cases $\mathbb{S}(k;N)$ will overestimate $S(k;N)$. It must, however, be stressed that the whole effect results in a change in the value of $\rho_0$, and thus does not actually set an upper bound in the SN $\bar{\rho}$. On the other hand, when no signal is present at a particular frequency, $\mathbb{S}(k;N)$ will underestimate $S(k;N)$, since we do not use $\bar{q}$ when computing it, and $\tilde{\tilde{\boldsymbol{u}}}(k;\bar{q}/M')$ has the biggest modulus among all the $\tilde{\tilde{\boldsymbol{u}}}(k;q/M')$.

To compute the overall probability density of $\mathbb{Z}(k)$ is very difficult, both in a single step (even if no signal is present) or in two steps, like in Section 4.1, since the auxiliary random variable $\mathbb{W}(k)$ defined as in (60) is no longer independent of $\bar{z}(k)$. We shall thus obtain it by a different method. We introduce a *tentative* $p(\mathbb{Z})$ inspired by the limiting probability density $p(\bar{z})$, because when $M$ is large enough both must coincide:

$$p(\mathbb{Z}) = MM'w\mathbb{Z}\mathrm{e}^{-(M/2)w\mathbb{Z}^2}\left[1-\mathrm{e}^{-(M/2)w\mathbb{Z}^2}\right]^{M'-1}. \tag{98}$$

The new parameter $w$, which condenses all the differences between $p(\bar{z})$ and $p(\mathbb{Z})$, measures in some suitable sense the *bias* of the spectral estimator $\mathbb{S}(k;N)$, i.e.,

$$w \sim \frac{<\mathbb{S}(k;N)>}{S(k;N)}. \tag{99}$$

This point of view is somewhat vain since we are not in a position to compute $<\mathbb{S}(k;N)>$ theoretically, as just stated. The value of $w$, however, can be estimated from the filter output itself, using for instance the relationship (91), replacing $\bar{z}$ by $\sqrt{w}\mathbb{Z}$, and the statistical mean by an average over the filter output,

$$w = \frac{N\sum_{l=1}^{M'} l^{-1}}{M\sum_{k=0}^{N/2-1}\mathbb{Z}^2(k)}. \tag{100}$$

The procedure we shall use to compute (100) also reminds us that $\mathbb{Z}(k)$ can be obtained from the experimental time series, irrespective of its probability density. We set the value of the $w$ quantity *a posteriori*, once the filtering process has finished. In fact, the functional form of $p(\mathbb{Z})$ has its very origin in the comparative study of the actual distribution $\mathbb{Z}(k)$ and the probability density of reference, $p(\bar{z})$, for different values for the free parameters. We pursue these ideas in the next section.

### 6.2 A single filtering process

Let us begin analysing a stream of about one day of the Explorer data with the layout used in Section 4.3, i.e. $N = 131072$ and $M = 36$, $M' = 64$. In fact, we will choose exactly the same time series, starting on 1991 August 3, in order to be able to compare the non-leaking and the leaking methods. We get the results shown in Fig. 3(a), where we have also externally introduced a signal analogous to that of Section 4.3 above, but its amplitude is now slightly bigger, $h_0 = 3\times10^{-23}$, i.e. the SN is about 1, because the new method is slightly less sensitive than the non-leaking one, as we have shown. And, of course, the signal spreads across different frequency bins: $k_0 = 50918$ and $\epsilon_0 = 0.1$.

A plot of the distribution of $\mathbb{Z}(k)$ is displayed in Fig. 3(b), where it is contrasted with $p(\bar{z})$ in (89), the *theoretical* probability density, and with $p(\mathbb{Z})$ once $w$ was computed following the prescription shown in (100), the *corrected* one. The agreement of the corrected probability density with the experimental probability density is again remarkable.

It is convenient to point out that when the number of processed blocks increases, the value of $w$ rapidly approaches one, thus becoming an irrelevant parameter. As a matter of fact Fig. 4, in which we present the same procedure with the same data but extend the processed time to two days, shows that $p(\bar{z})$ is then a sufficiently accurate expression for the probability density of the actual filter output.

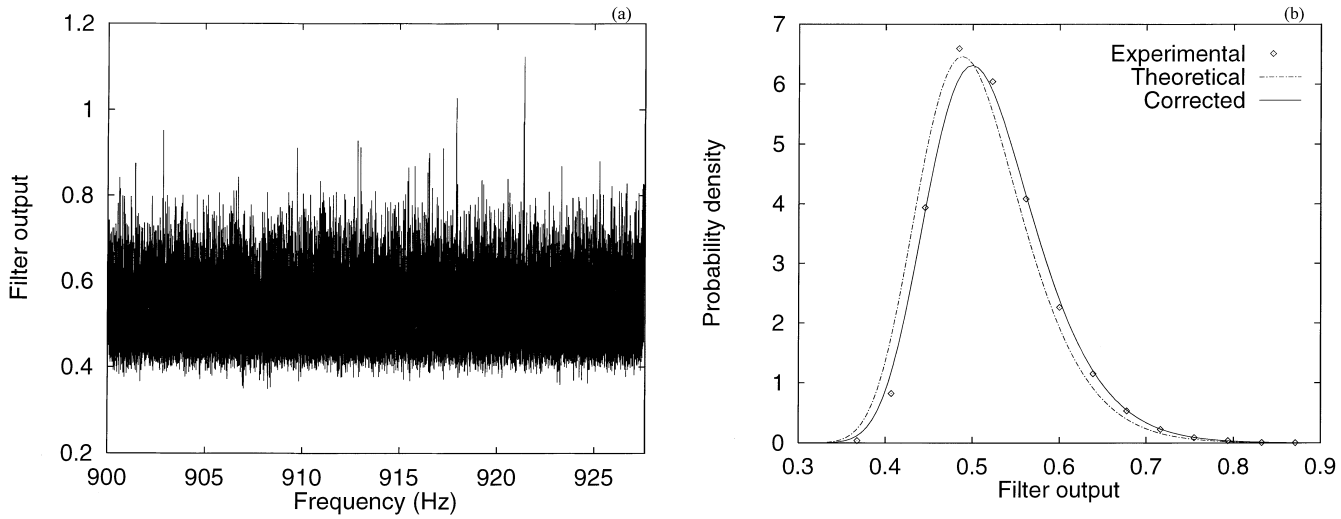


**Figure 3.** Here we find (a) the output of the filtering procedure, $\mathbb{Z}(k)$, for $N = 131072$, $M = 36$ and $M' = 64$, what represents about one day of data, and (b) its experimental distribution compared with the theoretically expected, $p(\bar{z})$, and corrected one, $p(\mathbb{Z})$.

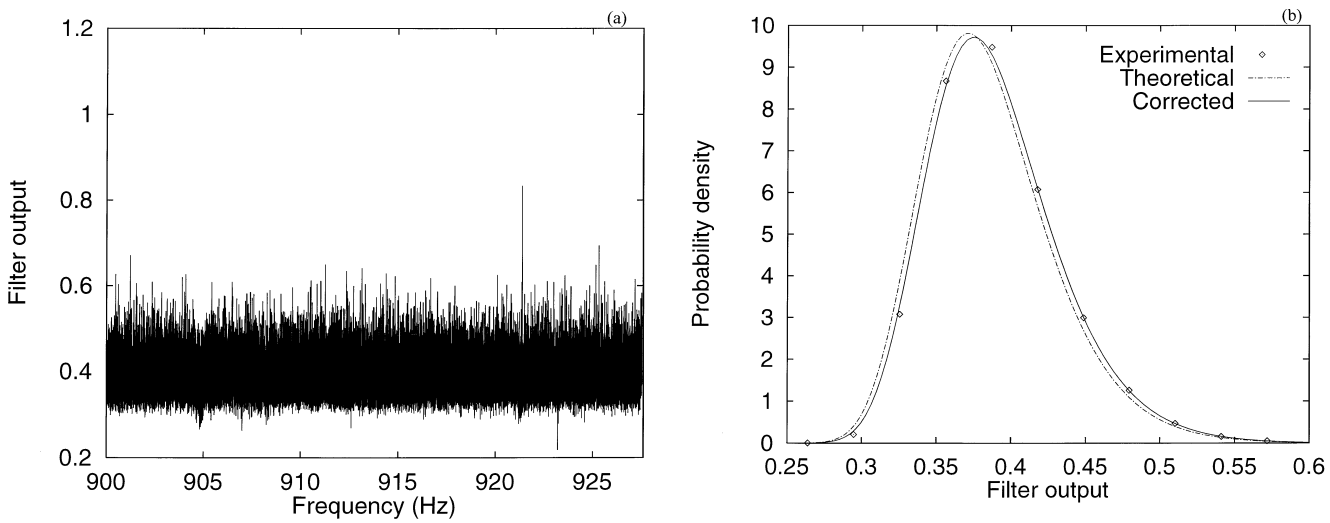


**Figure 4.** Here we find (a) the output of the filtering procedure, $\mathbb{Z}(k)$, for $N = 131072$, $M = 72$ and $M' = 128$, what represents about two days of data, and (b) its experimental distribution compared with $p(\bar{z})$ and $p(\mathbb{Z})$: in this case it is hard to distinguish from one another.

In the present case we can also compute the error of the first kind in terms of the threshold $\lambda_0$:

$$\mathcal{Q}_0 = 1 - \left[1 - \mathrm{e}^{-(M/2)w\lambda_0^2}\right]^{M'}, \tag{101}$$

and to invert this relationship,

$$\lambda_0 = \sqrt{-\frac{2}{Mw}\ln\left(1 - \left[1 - \mathcal{Q}_0\right]^{1/M'}\right)}. \tag{102}$$

By way of example, the threshold levels in units of the graphs in Figs 3 and 4 are 0.96 and 0.68, respectively, for a false alarm probability $\mathcal{Q}_0 = 10^{-5}$. In either case the signal is clearly above these thresholds, as it has respective heights of 1.12 and 0.83. From here we can rather accurately determine $\epsilon_0$, too: we find $\bar{q} = 6$ when $M' = 64$ (i.e., $\epsilon_{0,\mathrm{estimated}} = 0.094$), and $\bar{q} = 13$ when $M' = 128$ (i.e., $\epsilon_{0,\mathrm{estimated}} = 0.101$) for a real value of $\epsilon_0 = 0.1$.

## 7 OUTLOOK

The long-term operation of cryogenic GW detectors opens the possibility of looking for long-duration GW signals in the data generated by them, since signal-to-noise ratios are enhanced by the availability of long integration times. Monochromatic, as well as stochastic, signals belong in this category, though the latter require data from two or more independent antennas. In this paper we have addressed the problem of the design of suitable algorithms to allow us to single out possible monochromatic signals, coming from any direction in the sky, in the background of the noisy data produced by a cylindrical bar.

This is not such a simple problem, for a variety of reasons: computers are not arbitrarily powerful, detector duty cycles are not 100 per cent, data quality is not homogeneous, Doppler shifts distort monochromaticity, etc. Although some of the characteristics of the data are peculiar to the detector system, there are a number of procedures which should be quite generally usable. This paper is concerned with the problem of setting up filtering algorithms which enable the selection of candidate signals on the basis of the threshold-crossing criteria of the filtered data. To this end we have considered banks of filters which have a sinusoidal form but which enable signal phase estimation. We have determined the probability density functions at filter output and thence consistently established the probability of crossing a given level. We have also designed suitable estimators of the noise spectral density, which take into consideration the possibility that the signal might be in one or more of the FFT frequency bins for the complete duration of the analysed data.

These procedures have been checked by means of simulations on top of real data generated by the Explorer detector of the Rome group in 1991, and they work quite well: the theoretical predictions on the improvement in SN, as longer stretches of data are processed, remarkably coincide with the improvements observed in the real data analysis. They thus appear promising, and we would expect them to be useful in analysing other resonant detector data, whether the detectors are cylinders or the projected large spherical antennas, or the large-scale interferometers currently under construction, with suitable adaptive modifications.

As we have seen with the simulations in this paper, the Explorer detector comfortably sees amplitudes of $10^{-23}$, and therefore one can expect it to be sensitive to signals a few times $10^{-24}$ – see also Astone et al. (1997a). These are still rather high values for the expected amplitudes of signals of astrophysical origin, so that threshold-crossing criteria must be established with this in mind: if we set a very low false alarm tolerance, then the chances of missing the real event will grow, whereas if we set the tolerance rather low, the real signal will be treated as random noise.

Real signals, however, have characteristic Doppler patterns, the correct assessment of which should provide a more sound selection of candidate signals. In a forthcoming article we plan to apply the methodology here developed to a systematic analysis of the above-mentioned Explorer detector data over a long period of time, and to extend it to also include Doppler-shift effects in the data. The algorithms of Section 5 will be a powerful tool in this respect, if implemented under the perspective of the zoom transform (Yip 1976), for enhancing the frequency resolution within the spectral neighbourhood of pre-selected candidate lines.

## ACKNOWLEDGMENTS

The authors are grateful to the members of *ROG* group at Rome for giving them access to the data of the antenna Explorer, as well as the detector's transfer function details and procedures related to specific issues of the data analysis. We also thank Guido Pizzella for his critical reading of the manuscript and valuable suggestions, and for his continued hospitality. Bernard Schutz's comments, pointing us to the zoom transform algorithm, have been very helpful and are greatly appreciated as well. MM wishes to thank the DGR of the *Generalitat de Catalunya* for financial support, and JAL the Spanish Ministry Education for a grant, PB96-0384. We also acknowledge support from the *Institut d'Estudis Catalans*.

## REFERENCES

Astone P., 1998, in Proc. 2nd Edoardo Amaldi Conference, Geneva, July 1997, in press
Astone P., Bonifazi P., Pallottino G. V., 1990, Rev. Sci. Instrum., 61, 3899
Astone P. et al., 1991, Europhys. Lett., 16, 231
Astone P. et al., 1993, Phys. Rev. D, 47, 2
Astone P., Frasca S., Pallottino G. V., Pizzella G., 1997a, in Bassan M., Ferrari V., Francaviglia M., Fucito F., Modena I., eds, Proc. XII Italian Conf. on General Relativity and Gravitation Physics. World Scientific, Singapore
Astone P., Frasca S., Pallottino G. V., Pizzella G., 1997b, in Ciufolini I., Fidecaro F., eds, Proc. VIRGO Int. Conf. on Gravitational Waves: sources and detectors. World Scientific, Singapore
Bonazzola S., Gourgoulhon E., 1996, A&A, 312, 675
de Wild R., Nieuwkerk L. R., van Sinttruyen J. S., 1987, IEE Proc., 134, 659
Frasca S., La Posta C., 1991, Nuovo Cimento C, 14, 235
Frasca S. et al., 1992, in Nato S., Nakamura T., eds, Proc. MG6 Marcel Grossmann meeting, Gravitational Waves. World Scientific, Singapore
Gibbons G. W., Hawking S. W., 1971, Phys. Rev. D, 4, 2191
Gradshteyn I. S., Ryzhik I. M., 1980, Tables of Integrals, Series and Products. Academic Press, San Diego, CA
Hamilton W. O., 1994, in Coccia E., Pizzella G., Ronga F., eds, Proc. 1st Edoardo Amaldi Conf, Gravitational Wave Experiments. World Scientific, Singapore
Helstrom C. W., 1968, Statistical theory of signal detection. Pergamon Press, Oxford
Hung E. K. L., 1981, IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-29, 163
Kay S. M., 1990, Modern Spectral Estimation: theory and application. Prentice Hall, New Jersey
Lobo J. A., Montero M., 1997, in Miralles J. A., Morales J. A., Sáez D., eds, Proc. ERE-96 Conf., Some Topics on General Relativity and Gravitational Radiation. Editions Frontières, Paris
Lobo J. A., Montero M., 1998, in Bona C., Carot J., Mas L., Stela J., eds, Analytic and Numerical Approaches to Relativity: Sources of Gravitational Radiation. Univ. Illes Balears
Mauceli E., 1997, PhD thesis, Lousiana State Univ.
Montero M., 1997, in Buitrago J., Mediavilla E., Oscoz A., eds, Proc. ERE-95 Conf., General Relativity, Relativistic Astrophysics and Cosmology. World Scientific, Singapore
Montero M., 1998, PhD thesis, Univ. Barcelona
Niebauer T. M., Rüdiger A., Schilling R., Schnupp L., Winkler W., Danzmann K., 1993, Phys. Rev. D, 47, 3106

Olver F. W. J., 1970, in Abramowitz M., Stegun I. A., eds, Handbook of Mathematical Functions. Dover Press, New York
Pallottino G. V., Pizzella G., 1984, Nuovo Cimento C, 7, 155
Papoulis A., 1984, Signal analysis. McGraw-Hill, Singapore
Papoulis A., 1990, Probability and Statistics. Prentice Hall, New Jersey
Roland J., Frossati G., Teyssier R., 1994, A&A, 290, 364
Thorne K. S., 1987, in Hawking S. W., Israel W., eds, Three hundred years of gravitation. Cambridge Univ. Press, Cambridge
Weber J., 1969, Phys. Rev. Lett., 22, 1320
Yip P. C. Y., 1976, IEEE Trans. Comput., C-25, 287

This paper has been typeset from a TEX/LATEX file prepared by the author.