Class. Quantum Grav. **21** (2004) S821–S826 PII: S0264-9381(04)68301-8

First steps towards characterizing the hierarchical algorithm for curves and ridges pipeline

I S Heng[1], R Balasubramanian[2], B S Sathyaprakash[2] and B F Schutz[3]

[1] Max-Planck-Institute for Gravitational Physics (Albert-Einstein-Institute) and University of Hannover, Callinstr. 38, D-30167 Hannover, Germany
[2] Department of Physics and Astronomy, Cardiff University, PO Box 913, Cardiff, CF2 3VB, UK
[3] Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Golm, Am Mühlenberg 1, 14476 Golm, Germany

Received 1 September 2003
Published 10 February 2004
Online at stacks.iop.org/CQG/21/S821 (DOI: 10.1088/0264-9381/21/5/065)

Abstract
The hierarchical algorithm for curves and ridges is a variation of the burst gravitational wave search algorithm known as TFClusters. In this paper, we examine the detection efficiency of the hierarchical algorithm for curves and ridges by injecting sine-Gaussian burst waveforms with four different central frequencies into data acquired by the GEO600 gravitational wave detector during the S1 run. The fluctuation of the output signal-to-noise ratios was observed to be ∼5% for frequencies above 1.5 kHz and at least 15% for frequencies below 1.5 kHz. The uncertainty in the estimation of the arrival time is found to be less than 0.05 s.

PACS numbers: 07.05.Kf, 04.80.Nn

1. Introduction

The GEO600 gravitational wave detector is a laser interferometric gravitational wave detector with 600 m arm-lengths located near Hannover, Germany [1]. It acquired scientific data between 23 August and 9 September 2002 (henceforth referred to as the *S1 run*). One of the many analyses performed on the S1 run data is a search for burst gravitational waves. There are many algorithms used to identify burst gravitational wave candidates in data acquired by gravitational wave detectors. One such method used to identify burst gravitational waves in GEO600 data is the hierarchical algorithm for curves and ridges (HACR) [2]. HACR is an implementation of a time–frequency burst search algorithm, often referred to as *TFClusters* [3], within the GEO data analysis environment GEO++ [4].

To distinguish between genuine burst gravitational wave events and spurious environmental excitations, a search for coincident events is performed on candidate event lists from two or more widely spaced gravitational wave detectors. Since different detectors have different responses to incoming burst gravitational waves, signals must be injected (via

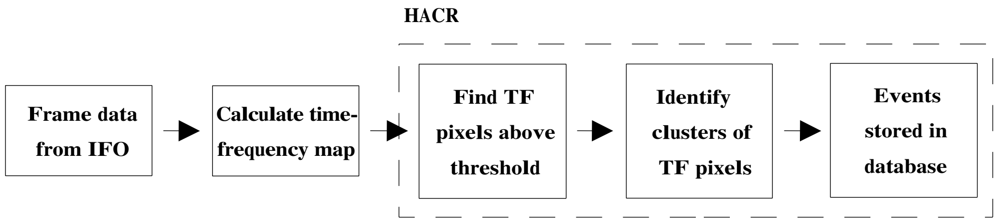

Figure 1. A block diagram of the HACR analysis pipeline.

both hardware and software) into the data and passed through the analysis pipeline to determine not only the detection efficiency of the single detector but also that of the network of detectors. Examining the arrival times of the injected signals that have been passed through the analysis pipeline allows one to determine the uncertainty in the estimate of the signal arrival time. This quantity is then used to determine the width of the time window within which events are considered to be coincident.

In this paper, we present the initial results from the study of software-injected sine Gaussians into GEO600 data acquired during the S1 run and passed through the HACR pipeline. We will begin by presenting a brief description of HACR before discussing the observations and concluding with a short summary and discussion of the presented work.

2. Injections into GEO S1 data

2.1. Description of HACR

As previously mentioned, HACR is based on TFClusters. Like TFClusters, HACR was developed to identify short burst of gravitational waves (lasting a few milliseconds to a few seconds) in broadband detectors such as an interferometric gravitational wave detector. A block diagram of the HACR analysis pipeline is illustrated in figure 1.

The algorithm starts by generating a spectrogram of the data segment being analysed. Then 10% of the samples at the extremes of the power distribution in each frequency bin of the spectrogram are removed before the mean and root-mean-squared (RMS) are calculated. This 10% is removed to avoid the extremes of the distribution, which could contain high signal-to-noise ratio (SNR) events, from biasing the estimate of the mean and RMS. The mean is then subtracted from all time samples in that frequency bin which are subsequently divided by the corresponding RMS value to obtain the SNR, in power, of each time–frequency 'pixel'.

A first SNR threshold is then applied to the spectrogram such that only time–frequency pixels with SNR above this threshold are considered. Any pixel above this threshold that has at least a neighbouring pixel above threshold is considered a *time–frequency cluster*. A threshold is then applied to the size of time–frequency clusters. Finally, only clusters with peak SNR values above a second SNR threshold are considered and the parameters of the remaining time–frequency clusters are stored in a database. Each cluster that survives the thresholding process is defined by its peak SNR value which we will refer to as the *HACR SNR*. The time assigned to each cluster is the calculated time at which the peak of the cluster occurs. If one considers the ideal case where the input data to HACR are stationary and Gaussian, then the output of HACR is a list of candidate burst gravitational wave events.

One of the differences between HACR and TFClusters is the way the spectrogram is constructed. For analysing data from the S1 run, the spectrogram for TFClusters was constructed without the use of a window nor were the segments overlapped. However, for

Table 1. HACR parameters used to study software injections.

Parameter name	Parameter value
Segment length	512 samples
Overlap	448 samples
Lower frequency cut-off	100 Hz
Upper frequency	3000 Hz
Lower (first) SNR threshold	20
Upper (second) SNR threshold	30
Fraction of outliers to exclude	0.1
Minimum cluster size	5 pixels

HACR, a Hanning window with 87.5% overlap between segments is used when constructing the spectrogram. Another difference is the way the first and second thresholds are defined in each algorithm. The first and second thresholds are defined in terms of the SNR of the time–frequency pixels in HACR, while TFClusters defines the first threshold in terms of the probability of the power of a particular time–frequency pixel and the second threshold on the size of the cluster. Also, for S1, TFClusters used a *generalized clusters* algorithm where two clusters that were within a certain distance in the time–frequency map were combined to form a single cluster. HACR does not perform such a combination. Finally, the SNR definitions used by the two algorithms are different. For TFClusters, the SNR of a trigger is defined by the total power of all the pixels in the cluster. For HACR, the SNR is the ratio of the peak power of the cluster with the mean power of the frequency bin in which the peak occurs in.

In the work that is reported in this paper, the parameters used by HACR to obtain the results are listed in table 1. The HACR parameters were tuned using data from a 2 h portion of GEO S1 data specially set aside for tuning gravitational wave signal search or veto identification algorithms (commonly referred to as *GEO S1 playground data*). The aim of the tuning process was to maximize the distance of the faintest burst gravitational wave source detectable by the GEO600 detector. However, this tuning was performed in an *ad hoc* manner and the quality of the results was determined by eye.

2.2. Injection and results

Sine Gaussians with four central frequencies were injected post-acquisition via software into data from the GEO S1 run and passed through the HACR pipeline. A plot showing the locations of the central frequencies of the sine Gaussians is shown in figure 2. Also, the power spectral density of the 1000 s of calibrated strain data from GEO600 (sampled at 16 384 Hz) into which the sine-Gaussian waveforms were injected is plotted in figure 2. One should note that during the S1 run, the GEO600 noise spectrum was highly non-stationary below 1.5 kHz. We expect this non-stationarity to increase the uncertainty in estimates of the amplitude and time of arrival of the observed signal.

The sine Gaussians, each lasting about 100 ms, were injected every 18 s into the 1000 s long stretch of data. The central frequencies were chosen to span the higher frequency region of the GEO noise curve, where most of the burst events will be identified. In addition, there was a random factor added to the central frequency of each signal so that the signal peak did not fall into a single frequency bin. To take into account the fact that higher frequency signals have more cycles for a given period of time than lower frequency signals, the quality

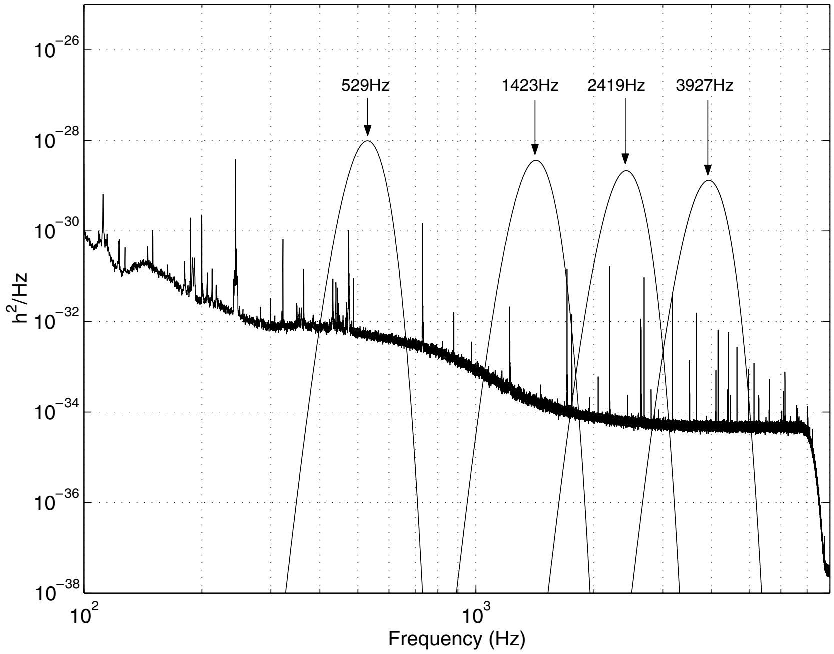

Figure 2. The power spectral densities of GEO600 S1 data from GPS second 715 608 013 to 715 609 013. The square of the Fourier transforms of the sine-Gaussian signals is also plotted in the same graph.

factor of all the sine-Gaussian waveforms (which governs how quickly the signal rings up and down) was fixed to an arbitrary choice of 4π (time constant, $\tau = 2/f_0$). The corresponding HACR SNR for each series of injected sine Gaussians fluctuates due to interactions between the injected signal and the detector noise. The injected signal responses were identified in the HACR output by choosing the strongest event within a ± 0.1 s window about the signal injection time. This step was necessary due to the presence of background events identified by HACR close to the injection time.

The HACR SNR of the identified injected signals is plotted in figure 3. The error bars correspond to 1 standard deviation in the distribution of the observed SNR. As expected, the slope of the lines on a log–log plot is approximately equal to 2. Also, the overall HACR SNR observed for the injected 1423 Hz sine Gaussians is approximately ten times larger than that for the 529 Hz sine Gaussians. This is expected if we compare the relative heights of the peaks of the Fourier transforms of two central frequencies in figure 2. For sine Gaussians with central frequencies 2419 Hz and 3927 Hz, the standard deviation is about 5% of the mean SNR. The standard deviation is proportionally greater for sine Gaussians with central frequencies of 529 Hz ($\sim$55%) and 1423 Hz ($\sim$15%). This is most likely due to the presence of non-stationarities in the noise at those frequencies. Most of the injected signals had a very high HACR SNR and were easily identified. However, for injected sine Gaussians with central frequencies of 2419 Hz, the observed HACR SNR is less than or equal to that of background events in a ± 0.1 s window for signal strain amplitudes smaller than 5×10^{-17}. The presence of these background events introduced an uncertainty in the HACR SNR estimation of the injected signal since. If we do not know the exact properties of the injected signal (as is the case when a burst gravitational wave impinges on the detector), we would not be able to identify which event was caused by the injected signal. The same uncertainty is observed for 529 Hz sine

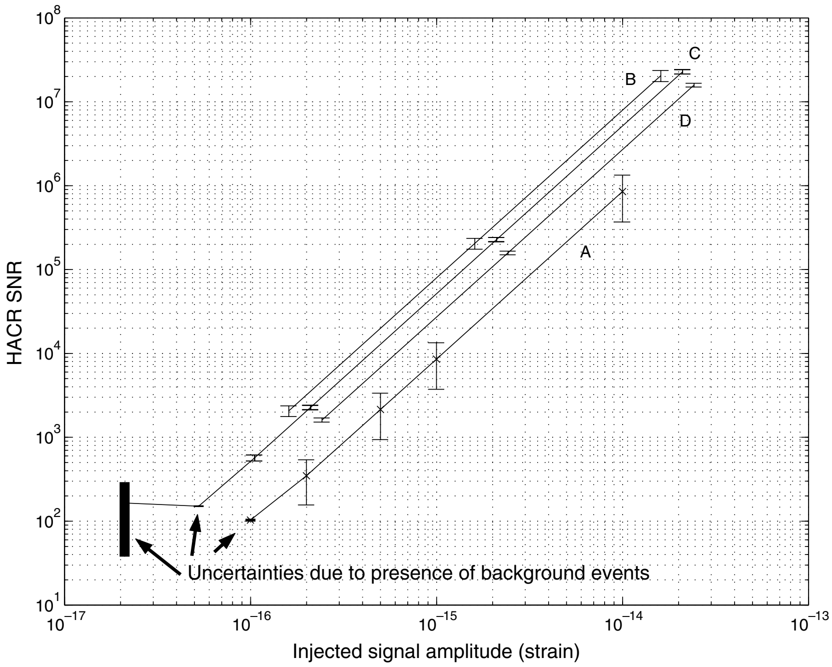

Figure 3. The observed HACR SNR plotted against the amplitude of the injected sine Gaussians with central frequencies (A) 529 Hz, (B) 1423 Hz, (C) 2419 Hz and (D) 3927 Hz. At lower SNRs, the presence of background events within the 0.1 s window creates an added uncertainty in the estimation of arrival times as shown by the rectangles.

Gaussians with strain amplitudes less than 10^{-16}. The region of uncertainty is represented by the rectangles in figure 3. The upper boundary of the rectangle marks the mean HACR SNR if one included the background noise events in the calculation. The lower boundary of the rectangle marks the mean HACR SNR if one only considers observed signals with a central frequency within 100 Hz of that of the injected signal from the calculation.

The uncertainty in the arrival time estimate can also be determined through signal injections. The standard deviation of the difference between the arrival time as estimated by HACR and the time when the sine Gaussian was actually injected is plotted in figure 4. The overall uncertainty increased as the central frequency of the injected sine Gaussian decreased. As with the observed HACR SNR, the presence of background events introduced an uncertainty in the arrival time estimation. From figure 4, it is reasonable for one to choose smaller coincidence windows for larger signals. However, the multi-interferometer coincidence analysis performed to date uses only a fixed window. Thus, we choose a conservative coincidence window by taking the upper value for the 2419 Hz injected sine Gaussians. This limits the minimum size of the coincidence window for strain amplitudes of 2×10^{-17} or greater to 0.05 s. This same minimum bound on the coincidence window should be applied to 529 Hz sine Gaussians with strain amplitudes greater than or equal to 10^{-16}.

3. Summary and discussion

The analysis of sine Gaussian injections into GEO600 S1 run data confirms the presence of non-stationarity at frequencies below 1.5 kHz. There were greater fluctuations in the observed

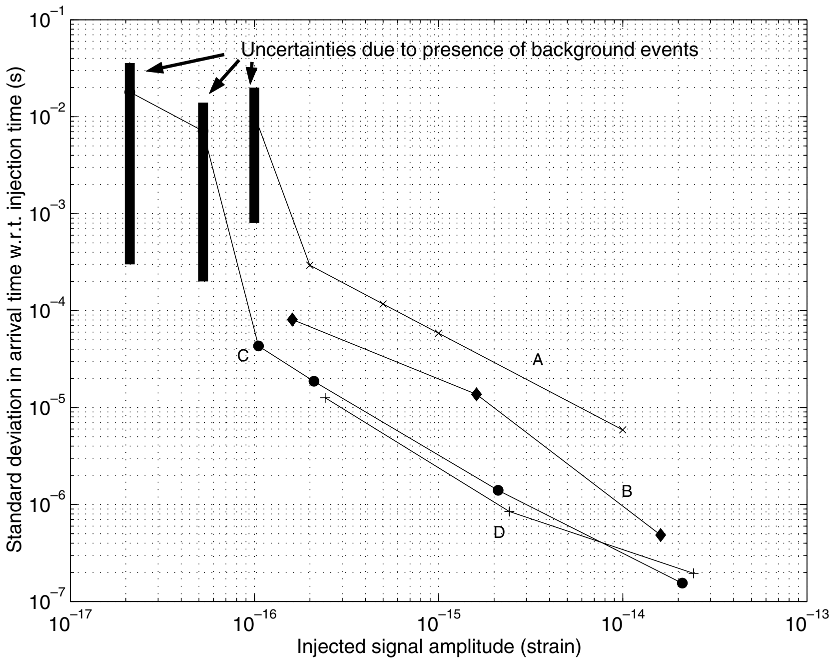

Figure 4. The uncertainty in the estimate of the arrival time with respect to the signal injection time for sine Gaussians with central frequencies (A) 529 Hz, (B) 1423 Hz, (C) 2419 Hz and (D) 3927 Hz. At lower SNRs, the presence of background events within the 0.1 s window creates an added uncertainty in the estimation of arrival times as shown by the rectangles.

HACR SNR for the 529 Hz and 1429 Hz sine Gaussians than for the 2419 Hz and 3927 Hz sine Gaussians. A conservative minimum bound for the width of the coincidence window has been set at 0.05 s.

These calculated numbers are the first part of the investigation into the response of GEO600 to burst gravitational wave signals. The next step would be to perform a Monte Carlo simulation of a particular source distribution (e.g. isotropically across the sky, galactic centre) from which one will determine the detection efficiency of GEO600 to the injected signals. It will also be important to consider signal injections in two or more detectors to determine the detection efficiency of a network of gravitational wave detectors. These efficiencies are an indication of how sensitive a detector is to incoming burst gravitational waves in different frequency bands. They are necessary for determining the rate at which burst gravitational waves are impinging on the Earth.

References

[1] Willke B *et al* 2002 *Class. Quantum Grav.* **19** 1377–87
[2] http://www.astro.cf.ac.uk/pub/R.Balasubramanian/geo++/doc++/classGEOPP_1_1HACRMon.html
[3] Sylvestre J 2002 *Phys. Rev.* D **66** 102004
[4] http://www.astro.cf.ac.uk/pub/R.Balasubramanian/geo++/doc++/