Amaldi 9 and NRDA 2011 IOP Publishing
Journal of Physics: Conference Series **363** (2012) 012027 doi:10.1088/1742-6596/363/1/012027

Progress on the Low-Latency Inspiral Gravitational Wave Detection algorithm known as SPIIR

Shaun Hooper[1,2], Linqing Wen[1,2], Chad Hanna[3], Kipp Cannon[4], Drew Keppel[5], David Blair[1], Shin-Kee Chung[1,2], Leo Singer[6], Yanbei Chen[6]

[1]Australian International Gravitational Research Centre, School of Physics, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia
[2]ICRAR-Fairway M468, School of Physics, The University of Western Australia, Crawley, WA 6009, Australia
[3]Perimeter Institute of Theoretical Physics
[4]Canadian Institute of Theoretical Astrophysics
[5]Albert-Einstein-Institut
[6]California Institute of Technology

E-mail: `shaun.hooper@uwa.edu.au`

Abstract. Low-latency event triggers to signify the presence of gravitational waves from coalescing binaries will be required to make prompt electromagnetic follow-up observations of electromagnetic counterparts. We present the recent progress made on implementing the time-domain low-latency detection algorithm known as summed parallel infinite impulse response (SPIIR) filtering into a real gravitational wave search pipeline.

1. Introduction

Neutron star binary mergers are currently the leading model of the progenitors of short hard gamma-ray bursts (short GRBs) [1, 2]. A very short delay (from 0.1 seconds to hundreds of seconds [3, 4]) is thought to happen between the final GW emission and the onset of the GRB. Presently the electromagnetic emission of the GRB event is not well understood. A prompt emission in X-ray and optical wavelengths followed by a delayed afterglow of cascading wavelengths is expected to be related to the initial GRB. For long GRBs (those where the burst duration is >2s) it is known that there is a prompt optical component that may occur tens to hundreds of seconds after the initial burst [5]. It is expected that for short GRBs there is also a prompt optical component (and possibly afterglow) happening on short time scales. Low-latency detection of the associated GW will be required in order to make follow-up observations of the electromagnetic counter-parts.

The current strategy to search for the existence of inspiral waveforms in the detector data is based on matched filtering [6]. This method, based on Wiener optimal filtering, is a correlation of an expected inspiral waveform template and the detector data, weighted by the inverse noise spectral density of the detector [7]. In order to save computational costs, this correlation is performed in the frequency domain, via a Fourier transform of a finite segment of detector data. Although computationally efficient, this naturally incurs a latency due to the fact that data must be collected before the Fourier transform can be applied. A method known as *Multi-Band Template Analysis* (MBTA) is currently under development to reduce the latency by splitting the matched filtering over two frequency bands [8]. However this is still a frequency domain method. Another low-latency method known as *Low-Latency On-line Inspiral Data* analysis (LLOID) can also give low-latency triggers in the time domain by first down-sampling the incoming data into multiple streams and then applying time domain finite impulse response (FIR) filters [9]. The computational cost of this pipeline is reduced by decreasing the number of templates via singular value decomposition [10]. In both cases, the computational cost scales with the length of the gravitational waveform. Advanced LIGO detectors will have improved bandwidth, especially at the low-frequency end. This means that the gravitational waveform length will increase, and consequently the computational cost will also increase.

The authors have previously [11, 12] introduced a new method to detect inspiral signals in the time domain using infinite impulse response (IIR) filters. This method works by approximating an inspiral waveform by a summation of time shifted exponentially increasing sinusoids. Each sinusoid can be searched for by applying a single pole IIR filter. In this manner the IIR filter acts as a narrow bandpass filter responding to a peak characteristic frequency. In order to search for the template a bank of single-pole IIR filters is applied in parallel. When each appropriately delayed IIR filter output is added coherently, the total output approximates the matched filter output of the exact waveform. We call this the *summed parallel infinite impulse response* (SPIIR) method.

In this proceeding we will discuss the current status and recent progress on the implementation of the low-latency method known as SPIIR.

2. The SPIIR Method

The *summed parallel infinite impulse response* (SPIIR) method has been previously been outlined in [11] and its mathematical concepts detailed in [12]. In this section, we will briefly go over the core parts of the SPIIR method.

The SPIIR method is designed as a low-latency time domain replacement to the optimal search strategy, matched filtering, commonly used in inspiral searches [6]. The authors have previously shown that is it possible to approximate the inspiral gravitational waveform components $h_{c,s}(t)$ by a summation of damped sinusoids u_l [11, 12],

$$h_{c,s}(t) = A(t)e^{i\phi(t)} \simeq U(t) = \sum_l u_l(t) = \sum_l b_{0,l} e^{(\gamma_l + i\omega_l)(t-t_l)} \Theta(t_l - t). \tag{1}$$

It is then possible to search for each sinusoid using the discrete single pole IIR filter

$$y_{k,l} = a_{1,l} y_{k-1} + b_{0,l} x_{k-d_l}. \tag{2}$$

Where the index k denotes discretely sampled times of t, namely $t_k = k\Delta t$. Figure 1 shows a schematic of the data processing pipeline for a one IIR filter. This is exactly equivalent to a cross-correlation of the damped sinusoids u_l where

$$a_{1,l} = e^{-(\gamma_l + i\omega_l)} \tag{3}$$

and d is a delay term equivalent to t_l. Here γ_l represents a damping factor and ω_l the peak frequency of the l-th IIR filter. The IIR filter (2) is a narrow band pass filter, centered on the frequency ω_l. As is shown in [11, 12], the summation of a *bank* of M IIR filters run in parallel, each with a different b_0, d, γ, ω, can approximate the optimal matched filter output,

$$z_k \simeq 2\Delta t \sum_{l=1}^{M} y_{k,l}. \tag{4}$$

for each template. The matched filter output z divided by a normalization constant gives the signal to noise ratio (SNR).

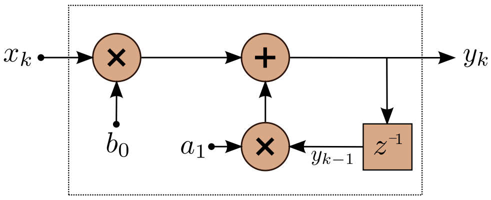

Figure 1. A signal processing schematic showing the flow of data through a single-pole IIR filter. The input, x_k is multiplied by a complex constant b_0, then added to the previous output that has been multiplied by another complex constant a_1, resulting in the current output y_k. It should be noted that this filter, in principle, should be have been run forever.

3. Progress

3.1. Advanced LIGO template bank

In order to implement the SPIIR method for the advanced detector era [13, 14], the first necessary step was to produce IIR banks for each template in the entire template bank. We choose to focus on the $1 - 3M_\odot$ mass range. Using the predicted advanced LIGO noise curve [15] with a lower cut off of 15 Hz we obtain 21890 mass-pairs in the given mass range with a minimal match of 97%. For each mass-pair, we produce a restricted second post Newtonian (2PN) waveform and find a bank of M IIR filters using the method outlined in [11]. We calculated the inner product between the exact 2PN waveform h and the IIR bank impulse response U to see if the impulse response approximates the waveform. We call the value of this inner product the overlap. Please see section IIIA of [11] for greater detail as to how the overlap is calculated. The overlap in all 21890 cases was above 99.2% (see figure 2). This step was necessary to assure that the SPIIR method would be reasonably capable of recovering detections all templates in the mass range $1 - 3M_\odot$.

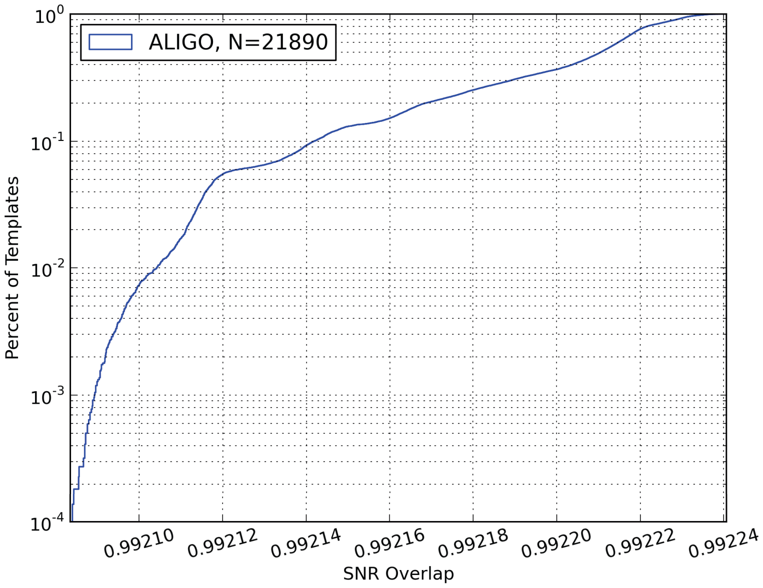

Figure 2. A cumulative distribution plot of the template SNR overlaps. The SNR overlaps is defined as the normalized inner product of the exact 2PN waveform, and the impulse response of the IIR bank.

3.2. Implementation with `gstlal`

Recent work by the authors has focused on the implementation of the SPIIR method via integration with an existing gravitational wave detection framework, `gstlal` [9]. `gstlal` is the combination of the LIGO Algorithm Library (LAL) and the open-source real-time multimedia handling software GStreamer [16]. Since processing gravitational wave data is very similar to processing audio data, GStreamer's stock signal processing elements, such as re-samplers and filters can be used as a powerful framework with which to create new gravitational wave search pipelines.

This software library is an ideal platform to integrate the SPIIR method, as there currently exists a pipeline that focuses on low-latency detection of inspiral signals; the LLOID pipeline [17]. The LLOID pipeline consists of many stock GStreamer ”elements” along with gravitational wave specific elements connected together to take raw detector data, whiten the data, and then matched filtering the data via the SVD method [10]. Once the SNR time series is produced, the pipeline then passes time segment “buffers” to an element that generates standard trigger files.

Because of the relative maturity of `gstlal`, the authors have focused on replicating the LLOID pipeline, but with the core processing element replaced with an “IIR bank” element. In order to implement the SPIIR method into the larger structure of `gstlal`, a core “element” code was written, which is now a part of the `gstlal` project. The element we wrote is capable of taking the single incoming channel of data and applying a set of N IIR banks simultaneously (one for each template) thereby producing N channel outputs. Each template IIR bank is a vector of M a_1's, b_0's and d's.

3.3. Multi-rate filtering

In order to reduce computational costs, the authors have worked to implement a very similar multi-rate scheme that exists in the LLOID pipeline. The motivation for this is as follows; IIR filters with characteristic frequencies ω much lower than the native sample rate need only be run at a sample rate of twice (or slightly more) ω. In order to achieve this, the coefficients a_1 and b_0, as well as d must be redefined for the new sample rate. This comes at the cost of first down-sampling the whitened detector data, as well as up-sampling the sub IIR bank output to the original sample rate. However preliminary theoretical calculations show that the additional computational costs of re-sampling are negligible compared the savings made from running the IIR filters at a reduced rate.

For each template's IIR bank, the M IIR filters are divided into different sample rate bins. For each sample rate bin, there will be a different *sub* IIR bank that must be filtered. Figure 3 shows the basic outline of the multi-rate pipeline.

The main goal of the SPIIR method is to have a time domain method that uses less computational resources than the tradiational time-domain matched filter (based on FIR filtering) and the frequency domain matched filter for low latencies. In [12], the authors outline the theoretical computational cost of the standard SPIIR scheme and multi-rate SPIIR scheme, and compare the cost with both the time domain and frequency domain matched filter. Figure 4 shows that if an extremely low latency is required, the SPIIR method can be faster than both time domain and frequency domain matched filters.

4. Future Work

As the advanced detectors won't be online until at least 2014, the authors are currently planning to test the `gstlal` SPIIR method in the next engineering run, scheduled for January 2012. The engineering run is proposed to test important software infrastructure that will be used in the advanced detector era. As a preliminary preparation for this experiment, the pipeline will be tested on using detector data from previous science runs. The triggers generated from this preliminary experiment will be compared with other methods, such as the LLOID pipeline.

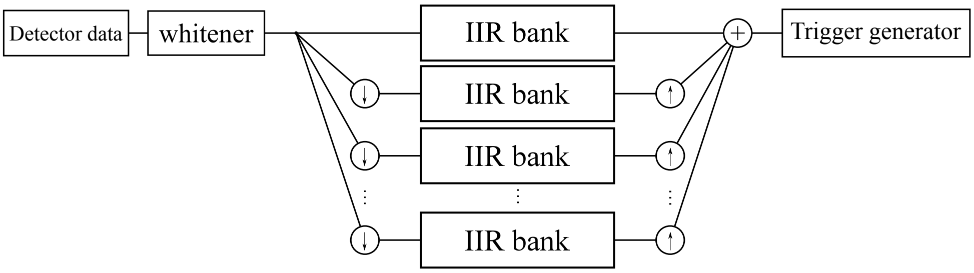

Figure 3. Basic data flow chart of the `gstlal` SPIIR method. After detector data is acquired, it is then whitened. Next the whitened data is sent to several down-samplers. Each down-sampler reduces the sample rate to an appropriate rate, as well as applying a low-pass filter. The data is then passed to IIR bank processing elements, where a bank of IIR filters runs on the data. The output of the IIR banks is the up-sampled to the highest sample rate, and added to the output of the other up-samplers. The SNR stream is then passed to a trigger generator, which compiles a database of events.

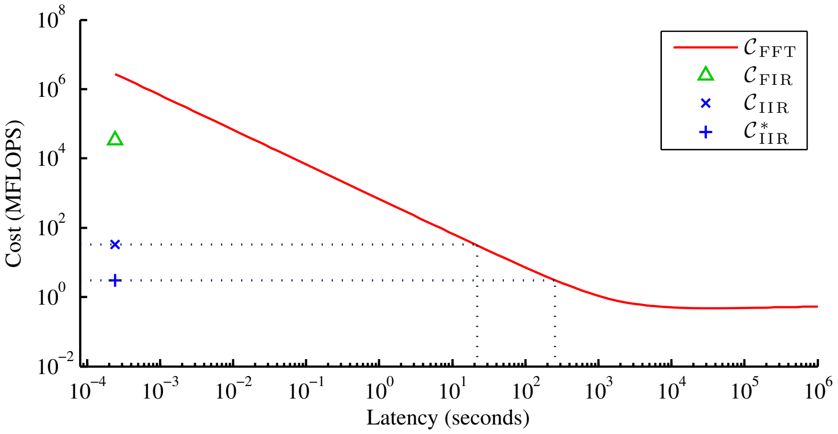

Figure 4. Theoretical computational cost of the frequency-domain (FFT) matched filter (solid line), time-domain (FIR) matched filter (triangle), IIR method (cross) and multi-rate IIR method (plus) to filter a representative $1.4 + 1.4\,\mathrm{M}_{\odot}$ template in advanced LIGO. The frequency domain filter requires segmenting, and is therefore a function of latency. Time domain methods have inherently sample time latencies.

As a side project to the `gstlal` implementation, the authors are also working on a GPU based SPIIR development (see [18] for details). The parallelization of the SPIIR method executed on GPU's is expected to result in significant computational speed improvements.

A further reduction in computational costs are expected to be made by the introduction of the template interpolation scheme introduced in [12]. This method works by reducing the complete, *fine*, template bank into a reduced *coarse* bank. It can be shown that different frequency bands of the coarse bank templates can be manipulated to produce all templates in the fine bank. This frequency dividing, much like the MBTA method, is ideally executed using the sub IIR banks. Although [12] demonstrates this procedure for Newtonian waveforms, we plan to extend this idea to post-Newtonian waveforms.

References

[1] Fox D B, Frail D A, Price P A, Kulkarni S R, Berger E, Piran T, Soderberg A M, Cenko S B, Cameron P B,

Gal-Yam A, Kasliwal M M, Moon D, Harrison F A, Nakar E, Schmidt B P, Penprase B, Chevalier R A, Kumar P, Roth K, Watson D, Lee B L, Shectman S, Phillips M M, Roth M, McCarthy P J, Rauch M, Cowie L, Peterson B A, Rich J, Kawai N, Aoki K, Kosugi G, Totani T, Park H, MacFadyen A and Hurley K C 2005 *Nature* **437** 845–850 (*Preprint* `arXiv:astro-ph/0510110`)

[2] Nakar E 2007 *Phys. Rep.* **442** 166–236 (*Preprint* `arXiv:astro-ph/0701748`)

[3] van Putten M H P M 2009 (*Preprint* `0905.3367`)

[4] Zhang B and Mészáros P 2004 *International Journal of Modern Physics A* **19** 2385–2472 (*Preprint* `arXiv:astro-ph/0311321`)

[5] Klotz A, Boër M, Atteia J L and Gendre B 2009 AJ **137** 4100–4108 (*Preprint* `0902.0898`)

[6] Abbott B P, Abbott R, Adhikari R, Ajith P, Allen B, Allen G, Amin R S, Anderson S B, Anderson W G, Arain M A and et al 2009 Phys. Rev. D **79** 122001 (*Preprint* `0901.0302`)

[7] Wainstein L A and Zubakov V D 1962 *Extraction of Signals from Noise* (Prentice-Hall)

[8] Buskulic D, for the LIGO Scientific Collaboration and the Virgo Collaboration 2010 *Classical and Quantum Gravity* **27** 194013–+

[9] `https://www.lsc-group.phys.uwm.edu/daswg/projects/gstlal.html`

[10] Cannon K, Chapman A, Hanna C, Keppel D, Searle A C and Weinstein A J 2010 *Phys. Rev. D* **82** 044025

[11] Hooper S *et al.* 2011 (*Preprint* `1108.3186`)

[12] Luan J, Hooper S, Wen L and Chen Y 2011 (*Preprint* `1108.3174`)

[13] Harry G M, for the LIGO Scientific Collaboration 2010 *Classical and Quantum Gravity* **27** 084006

[14] Acernese F *et al.* 2009 `https://tds.ego-gw.it/itf/tds/file.php?callFile=VIR-0027A-09.pdf`

[15] `https://www.lsc-group.phys.uwm.edu/daswg/projects/lal/nightly/docs/html/LALAdvLIGOPsd_8c_source.html`

[16] `http://gstreamer.freedesktop.org/`

[17] Cannon K *et al.* 2011 (*Preprint* `1107.2665`)

[18] Liu Y 2011 *J. Comput. Phys.* (in preparation)