

Summed parallel infinite impulse response filters for low-latency detection of chirping gravitational waves

Shaun Hooper,^{1,2,*} Shin Kee Chung,^{1,2} Jing Luan,³ David Blair,¹ Yanbei Chen,³ and Linqing Wen^{1,2,†}

¹*Australian International Gravitational Research Centre, School of Physics, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia*

²*ICRAR-Fairway M468, School of Physics, The University of Western Australia, Crawley, Western Australia 6009, Australia*

³*Theoretical Astrophysics 350-17, California Institute of Technology, Pasadena, California 91125, USA*

(Received 16 August 2011; published 5 July 2012)

With the upgrade of current gravitational wave detectors, the first detection of gravitational wave signals is expected to occur in the next decade. Low-latency gravitational wave triggers will be necessary to make fast follow-up electromagnetic observations of events related to their source, e.g., prompt optical emission associated with short gamma-ray bursts. In this paper we present a new time-domain low-latency algorithm for identifying the presence of gravitational waves produced by compact binary coalescence events in noisy detector data. Our method calculates the signal to noise ratio from the summation of a bank of parallel infinite impulse response filters. We show that our summed parallel infinite impulse response method can retrieve the signal to noise ratio to greater than 99% of that produced from the optimal matched filter.

DOI: [10.1103/PhysRevD.86.024012](https://doi.org/10.1103/PhysRevD.86.024012)

PACS numbers: 04.25.Nx, 04.30.Db, 04.80.Cc, 04.80.Nn

I. INTRODUCTION

The interferometric gravitational wave (GW) detectors LIGO [1], and Virgo [2] have reached a sensitivity at which the detection of GWs is possible. The LIGO detectors are currently undergoing a major upgrade to Advanced LIGO, for which the sensitivity will be improved tenfold relative to Initial LIGO [3]. Hence in the era of advanced detectors GWs produced from inspiralling compact binaries will be detectable within a volume of space one thousand times larger than that of Initial LIGO, out to $\sim 200\text{--}300$ Mpc [4].

The emission of GWs produced by compact binary coalescence (CBC) can be modeled with a high degree accuracy [5]. When two compact bodies, such as neutron stars or black holes are in orbit, Einstein's equations predict the generation of GWs. As the bodies spiral towards each other a GW is created that increases in frequency over time until the bodies merge, following what is known as the inspiral waveform. Ground based detectors have frequency passbands that allow them to be sensitive to the final stages of such events up to total system masses of several hundred M_{\odot} .

Neutron star binary mergers are widely thought to be the progenitors of short hard gamma-ray bursts (short GRBs) [6,7]. The delay between the final GW emission and the onset of the GRB is estimated to be as short as 0.1 seconds or as long as tens to hundreds of seconds [8,9]. The electromagnetic emission of the GRB event is not well understood. Related to the initial GRB there is thought to be a prompt emission in X-ray and optical wavelengths followed by a delayed afterglow of cascading wavelengths.

Prompt optical emission may occur tens to hundreds of seconds after the initial burst. The low-latency detection of the GW associated with a neutron star merger could lead to the localization of a GRB source event on the sky, enabling fast moving telescopes to observe the prompt optical emission. Data collected from a multitude of sources—GWs, gamma-rays, X-rays and optical counterparts of the GRB—will lead to maximum insight into these highly energetic events.

The standard strategy for searching for the existence of inspiral waveforms in the detector data is based on matched filtering [5] (and references therein). 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 [10]. 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. In previous LIGO searches, the detector data is split up into “science blocks,” which are further divided into “data segments” chosen to be at least twice the length of the longest waveform in the template bank [11]. Each data segment is chosen to overlap the previous one by 50%. Each segment therefore must be matched filtered in a time that is half the length of the segment for a real-time analysis; that is, the filter output rate is equal to the data input rate. In this case, the matched filter process has a minimum latency (from signal arrival to signal detection) that is proportional to the longest template (see [12] for more details). Advanced LIGO will have an increased bandwidth over Initial LIGO, with the lower bound dropping from 40 to 10 Hz [4]. GW signals from CBC events spend much more time at these lower frequencies. Hence waveforms used for matched filtering in Advanced LIGO

*shaun.hooper@uwa.edu.au

†linqing.wen@uwa.edu.au

will be much longer (thousands of seconds). This in turn means the segment length will be increased, further increasing the latency. The latency of this method to produce GW triggers is longer than the time to onset of prompt optical emission after coalescence (10 to 100 s of seconds). After this amount of time, the early electromagnetic counterpart of a GRB event will be significantly faded, and may be missed by telescopes altogether.

A low-latency GW detection method is required to trigger follow-up electromagnetic observations of the prompt optical emission. So far two frequency domain methods have been developed to solve this issue. The VIRGO group has produced a low-latency pipeline based on multiband template analysis (MBTA) [13], and LIGO is also working on a new method, low-latency on-line inspiral data (LLOID) analysis method. In MBTA the matched filtering technique is split over two frequency bands, and the output is coherently added, reducing latency. A latency of less than 3 minutes until the availability of a trigger using this method has been achieved [13]. Low-latency in the LLOID method is achieved by first down-sampling the incoming data into multiple streams and then applying frequency domain finite impulse response (FIR) filters [14]. The computational cost of this pipeline is reduced by decreasing the number of templates via singular value decomposition [15].

We introduce a new method to detect CBC signals in the time domain using infinite impulse response (IIR) filters. Approximating an inspiral waveform by a summation of time shifted exponentially increasing sinusoids enables us to construct a bank of parallel single-pole IIR filters. Each IIR filter acts as a narrow bandpass filter. When each appropriately delayed IIR filter is added the coherent output approximates the matched filter output of the exact waveforms. We call this the summed parallel infinite impulse response (SPIIR) method. Figure 1 visually demon-

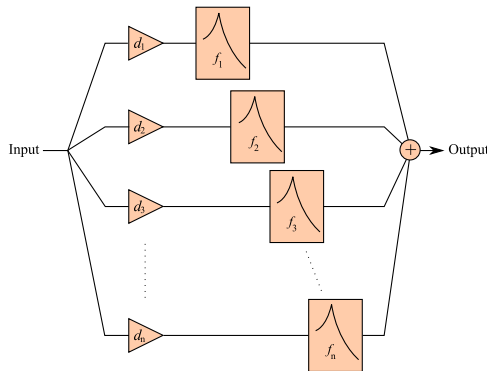


FIG. 1 (color online). A schematic overview of the SPIIR method. The input is split into different channels, time delayed by an amount d_i , then passed through a narrow bandpass IIR filter, each with a different central frequency f_i . Finally the output of each individual IIR filter is summed, giving the output of the SPIIR method.

strates the idea of using a bank of IIR filters as narrow bandpass filters. For a full explanation of the mathematical principles, see [12].

In this follow-up paper, we numerically address the issues essential to the practical use of this method for the upcoming advanced detectors. We calculate the filter coefficients and demonstrate via numerical simulations how well our method approximates the optimal matched filter as a function of the number of filters per bank using a range of parameters. We also show that the detection rate of the SPIIR method is very similar to that of the matched filter method. It has been shown theoretically that in order to get the same latency as the SPIIR method, the frequency domain matched filter method would require greater computational resources [12].

The structure of this paper is as follows: In Sec. II we will go through the formal introduction of the inspiral waveform and matched filtering, and how to get from the continuous frequency domain matched filter to the time-domain discrete matched filter. This will lead to a demonstration on how it is possible to approximate an inspiral signal by a sum of exponentially increasing sinusoids. The methodology is explained in Sec. III and will cover how we set up our simulation to test the efficiency of the SPIIR method as opposed to the frequency domain matched filter. Section IV will analyze the results of the simulation and Sec. V will discuss the implications of these results for advanced detectors.

II. METHODOLOGY

Gravitational wave interferometers output the strain induced by gravitational waves incident on the detector, as well as inherent noise. In unitless strain, the detector output will be

$$s(t) = \begin{cases} n(t) & \text{if signal is absent} \\ n(t) + h(t) & \text{if signal is present,} \end{cases} \quad (1)$$

where $n(t)$ is the noise inherent in the detector, assumed to be a stationary Gaussian process with mean zero. The sensitivity of the instrument can be characterized by the (one-sided) power spectral density $S_n(f)$ defined by

$$\langle \tilde{n}(f) \tilde{n}^*(f') \rangle = \frac{1}{2} S_n(f) \delta(f - f'), \quad (2)$$

where $\langle \cdots \rangle$ denotes the ensemble average over detector noise, and the tilde represents the forward Fourier transform,

$$\tilde{q}(f) = \int_{-\infty}^{\infty} q(t) e^{-2\pi i f t} dt. \quad (3)$$

A. The inspiral waveform

The gravitational wave strain incident at the interferometer is given by

$$h(t) = F_+(\theta, \varphi, \psi)h_+(t) + F_\times(\theta, \varphi, \psi)h_\times(t), \quad (4)$$

where the detector antenna response functions F_+ and F_\times are functions of (θ, φ) —the standard spherical polar coordinates measured with respect to the detector's frame, and ψ is the polarization angle. The detector antenna response functions, F_+ and F_\times , can be found in [16]. The $+$ and \times polarizations of the waveform are

$$h_+(t) = \left(\frac{1 + \cos^2 \iota}{2} \right) A(t) \cos(\phi(t) + \phi_c) \quad (5)$$

and

$$h_\times(t) = (\cos \iota) A(t) \sin(\phi(t) + \phi_c). \quad (6)$$

For nonspinning binaries with component masses m_1, m_2 in the range of $(1-3)M_\odot$ —which we will hereafter assume—the waveforms can be modeled to very high accuracy using the restricted post-Newtonian (PN) expansion [17–19] in the LIGO band (assumed to be 10–1500 Hz for Advanced LIGO). For restricted waveforms, only the leading order of the amplitude $A(t)$ is taken,

$$A(t) = -\frac{GM\eta}{Dc^2} \left(\frac{\eta}{5GM/c^3} (t_c - t) \right)^{-1/4}, \quad (7)$$

where $M = m_1 + m_2$ is the total mass, and $\eta = m_1 m_2 / M^2$ is the symmetric mass ratio. The phase of the gravitational waveform ϕ produced by a coalescing compact binary system evolves at twice the rate of the instantaneous orbital phase. The orbital phase can be approximated via the post-Newtonian expansion. To second post-Newtonian order the phase of the inspiral gravitational waveform is [19]

$$\phi(t) = -\frac{2}{\eta} \left[\Theta(t)^{5/8} + \left(\frac{3715}{8064} + \frac{55}{96} \eta \right) \Theta(t)^{3/8} - \frac{3\pi}{4} \Theta(t)^{1/4} + \left(\frac{9275495}{14450688} + \frac{284875}{258048} \eta + \frac{1855}{2048} \eta^2 \right) \Theta(t)^{1/8} \right], \quad (8)$$

where we have used the convenient dimensionless time parameter Θ ,

$$\Theta(t) = \left(\frac{c^3 \eta}{5GM} \right) (t_c - t). \quad (9)$$

In addition to the component masses m_1, m_2 , there are several unknown parameters: the time of coalescence t_c , the phase of the gravitational waveform at coalescence ϕ_c , distance from observer to source D , the inclination angle of the binary's orbital plane relative the line of sight ι , and the polarization angle ψ . One can use the linear combination trigonometric identity

$$a \cos \theta + b \sin \theta = \sqrt{a^2 + b^2} \cos(\theta - \alpha), \quad (10)$$

where $\tan \alpha = b/a$ to reexpress the strain (4) as

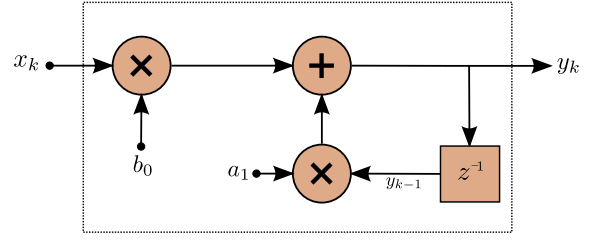


FIG. 2 (color online). A signal processing schematic showing the flow of data through a digital single-pole IIR filter. The input x_k is multiplied by a complex constant b_0 , then added to the previous output that has been multiplied by another complex constant a_1 , resulting in the current output y_k . It should be noted that this filter, in principle, should have been run forever.

$$h(t) = \frac{D}{D_{\text{eff}}} A(t) \cos(\phi(t) - \phi_0) \quad (11)$$

$$= \frac{(1 \text{ Mpc})}{D_{\text{eff}}} [h_c(t) \cos \phi_0 + h_s(t) \sin \phi_0], \quad (12)$$

where the factor D_{eff} obtained from the trigonometric identity (10) is

$$D_{\text{eff}} = \frac{D}{\sqrt{F_+^2 (1 + \cos^2 \iota)^2 / 4 + F_\times^2 (\cos \iota)^2}} \quad (13)$$

and ϕ_0 , an unknown phase term is

$$\phi_0 = \arctan \frac{F_\times (2 \cos \iota)}{F_+ (1 + \cos^2 \iota)} - \phi_c. \quad (14)$$

Here we define h_c and h_s as the “cosine” and “sine” components of the gravitational waveform. They are equivalent to the polarizations of a gravitational waveform that would be produced by an optimally orientated inspiralling binary. An optimal orientated binary system is one that has its orbital plane perpendicular to and has its orbit centered on the detector's z axis (i.e. where the unknown phase term would be $\phi_0 = 0$ and $\phi_0 = \pi/2$ respectively). Although not strictly necessary, the usual convention is to scale the cosine and sine components to an optimally orientated template at a distance of 1 Mpc, which we have done by setting D in Eq. (12). Hence, the cosine and sine components are defined as

$$h_c(t) = A_{1 \text{ Mpc}}(t) \cos \phi(t), \quad (15)$$

$$h_s(t) = A_{1 \text{ Mpc}}(t) \sin \phi(t), \quad (16)$$

where $A_{1 \text{ Mpc}}$ is simply (8) with $D = 1 \text{ Mpc}$.

B. Two-phase matched filter

The matched filter is the optimal linear filter for detecting known signals in noisy data [10]. In this paper, we will follow the derivation of the matched filter as it appears in Sec. III of [20], which itself is based on classical signal

analysis methods. We define the output of the matched filter as a correlation of the detector data s and the filter Q , weighted by the noise-spectral density $S_n(|f|)$

$$z = 2 \int_{-\infty}^{\infty} \frac{\tilde{s}(f) \tilde{Q}^*(f)}{S_n(|f|)} df. \quad (17)$$

Note that z may be a complex value depending on the choice of Q . In the case that the detector data contains Gaussian noise $n(t)$ only, the expectation value of the matched filter output z is zero. In such a case, the variance of the output of the matched filter z is

$$\sigma_Q^2 = \langle z \cdot z^* \rangle = 2 \int_{-\infty}^{\infty} \frac{|\tilde{Q}(f)|^2}{S_n(|f|)} df. \quad (18)$$

Let us assume the gravitational waveform is present in the detector data, ending at some time t_c not known *a priori*. There is also a constant phase term ϕ_0 that is not known ahead of time. A common way [5,17,20] to search for the unknown time of coalescence t_c and phase term ϕ_0 is to filter both components h_c and h_s (which correspond to orthogonal phases $\phi_0 = 0$ and $\phi_0 = \pi/2$) separately and then combine them to form a complex signal. This can be done by using the complex filter

$$\tilde{Q}^*(f) = [\tilde{h}_c^*(f) + i\tilde{h}_s^*(f)]e^{2\pi i f t_c}, \quad (19)$$

where we now explicitly use $\tilde{h}(f)$ to represent the Fourier transform of $h(t)$ when $t_c = 0$. This convention shall be used throughout the remainder of the paper. This way, the unknown time of coalescence t_c can now be searched over as an extrinsic parameter by changing the value of t_c in the complex matched filter

$$\begin{aligned} z(t_c) = 2 \int_{-\infty}^{\infty} \frac{\tilde{s}(f) \tilde{h}_c^*(f)}{S_n(|f|)} e^{2\pi i f t_c} df \\ + i2 \int_{-\infty}^{\infty} \frac{\tilde{s}(f) \tilde{h}_s^*(f)}{S_n(|f|)} e^{2\pi i f t_c} df. \end{aligned} \quad (20)$$

This is a cross correlation of the components h_c and h_s with detector data s , weighted by inverse noise-spectral density. The real part corresponds to the matched filter output for the (real) template h_c , and the imaginary part corresponds to the matched filter output for the (real) template h_s . In the stationary phase approximation [21] the components h_c and h_s are exactly orthogonal. It then follows that $\tilde{h}_c(f) = -i\tilde{h}_s(f)$ for $f > 0$. When this property is applied to Eq. (20), we have the form of the two-phase matched filter as

$$z(t_c) = 4 \int_0^{\infty} \frac{\tilde{s}(f) \tilde{h}_c^*(f)}{S_n(|f|)} e^{2\pi i f t_c} df, \quad (21)$$

commonly found in inspiral search papers [5,17]. In this paper we prefer to maintain the form of the two-phase filter in Eq. (20). Following convention (cf. [22]), the amplitude signal to noise ratio (SNR) ρ is defined as the absolute

value of the two-phase matched filter z divided by the standard deviation of its real part:

$$\rho(t) = \frac{|z(t)|}{\sigma}, \quad (22)$$

where

$$\sigma^2 = 2 \int_{-\infty}^{\infty} \frac{|\tilde{h}_c(f)|^2}{S_n(|f|)} df. \quad (23)$$

In the absence of a waveform, the SNR squared ρ^2 is chi-square distributed with two degrees of freedom (one for each of the components). Hence the probability of finding an SNR value greater than ρ_* in the absence of a waveform is [23]

$$P(\rho^2 > \rho_*^2) = e^{-\rho_*^2/2}. \quad (24)$$

C. Discrete time-domain filtering

The two-phase matched filter (20) is a cross correlation of each component $h_{c,s}(t)$ and the detector output $s(t)$, weighted by the inverse noise-spectral density $S_n(f)$. By defining the quantity x as the *over-whitened* strain data,

$$x(t) = \int_{-\infty}^{\infty} \frac{\tilde{s}(f)}{S_n(|f|)} e^{2\pi i f t} df, \quad (25)$$

we can use the cross correlation theorem to define the two-phase matched filter (20) in the time domain:

$$z(t) = 2 \int_{-\infty}^t x(t') h_c(t' - t) dt' + i2 \int_{-\infty}^t x(t') h_s(t' - t) dt' \quad (26)$$

$$= 2 \int_{-\infty}^t x(t') \hat{h}(t' - t) dt', \quad (27)$$

where we have redefined $h_c(t)$ and $h_s(t)$ to represent the cosine and sine components (15) and (16) when $t_c = 0$. For simplicity of later equations, we define $\hat{h}(t) = h_c(t) + i h_s(t) = A_{1 \text{ Mpc}}(t) e^{i\phi(t)}$ where the amplitude and phase terms have $t_c = 0$.

In practice the detector data $s(t)$ is sampled at intervals of Δt . The discretized form of the continuous time-domain matched filter (27) is

$$z_k = 2 \sum_{j=-\infty}^k x_j \hat{h}_{j-k} \Delta t, \quad (28)$$

where the index k denotes discretely sampled times of t , namely $t_k = k\Delta t$. In practice, the inspiral waveform template h_i is bounded (because the detector is only sensitive over a bandwidth) and the summation becomes finite, making this a FIR filter.

D. Infinite impulse response filter

Now let us introduce an alternative digital filter, the IIR filter. The difference equation of a general IIR filter is

$$y_k = \sum_{n=1}^N a_n y_{k-n} + \sum_{m=0}^M b_m x_{k-m}, \quad (29)$$

where y_k is the filter output at time step k ($t_k = k\Delta t$), x_k is the filter input, and a 's and b 's are complex coefficients.

Examples of IIR filters in common usage are Chebyshev, Butterworth and elliptic filters [24,25]. IIR filters use much less computational resources than an equivalent FIR filter. This is because they have “memory”—the previous outputs are fed back into the filter. However *digital* IIR filter design is a more complex process than FIR design. Obtaining the coefficients is usually done by first constructing an equivalent analog filter and applying well-known methods, such as the bi-linear transform or impulse invariance. Multiple IIR filters used together have different forms, such as direct form I & II, cascade (series) and parallel. In a series configuration, the overall transfer function is the multiplication of each IIR filter transfer function. In a parallel bank of IIR filters, where the output is summed together, the overall transfer function is the summation of the different transfer functions.

First, let us analyze the simplest single-pole IIR filter. The difference equation of this filter is

$$y_k = a_1 y_{k-1} + b_0 x_k. \quad (30)$$

(Figure 2 shows a schematic of the single pole filter data flow.) A solution to this first-order linear inhomogeneous difference equation is

$$y_k = \sum_{j=-\infty}^k x_j b_0 a_1^{k-j}. \quad (31)$$

By defining the complex coefficient a_1 in the form

$$a_1 = e^{-(\gamma+i\omega)\Delta t}, \quad (32)$$

and comparing (28) and (31), it is easy to see that the output of the simple filter (30) is the cross correlation of x_k and a complex sinusoid u_n with frequency ω and a magnitude that increases with an exponent factor γ for $n < 0$:

$$u_n = b_0 e^{(\gamma+i\omega)n\Delta t} \Theta(-n), \quad (33)$$

where $\Theta(-n)$ is the Heaviside function.

E. Approximation to an inspiral waveform

Since $\phi(t)$ is not linear in time, a complex sinusoid (33) cannot approximate the $h_{c,s}$ components of the inspiral waveform $\hat{h}(t) = A_{1\text{Mpc}}(t)e^{i\phi(t)}$. However we can easily linearize the components by a first-order Taylor expansion about the time t_l^* :

$$A_{1\text{Mpc}}(t)e^{i\phi(t)} \simeq A_{1\text{Mpc}}(t_l^*)e^{i\phi(t_l^*)+i\dot{\phi}(t_l^*)(t-t_l^*)}; \quad (34)$$

since the amplitude $A_{1\text{Mpc}}(t)$ does not increase at the same rate as $\phi(t)$, only a linear expansion of $\phi(t)$ is required. Multiplying by the window function $e^{\gamma_l(t-t_l)}\Theta(t_l-t)$ makes this approximation an exponentially increasing constant-frequency complex sinusoid with cutoff time t_l :

$$u_l(t) = A_{1\text{Mpc}}(t_l^*)e^{i(\phi(t_l^*)+\dot{\phi}(t_l^*)(t-t_l^*))}e^{(\gamma_l+i\dot{\phi}(t_l^*)(t-t_l))}\Theta(t_l-t). \quad (35)$$

The expansion point t_l^* is chosen to be near the cutoff time $t_l^* = t_l - \alpha T_l$, where α is a tunable parameter and the interval T_l is the duration in which the approximation is valid:

$$\left| \frac{1}{2} \ddot{\phi}(t_l) T_l^2 \right| = \epsilon < 1 \quad (36)$$

with ϵ a tunable parameter chosen to be too small. Equation (35) implies that the coefficient b_0 for the l th complex sinusoid is

$$b_{0,l} = A_{1\text{Mpc}}(t_l^*)e^{i(\phi(t_l^*)+\dot{\phi}(t_l^*)(t_l-t_l^*))} \quad (37)$$

and the frequency $\omega_l = \dot{\phi}(t_l^*)$.

In this paper, we chose the cutoff time t_l of the first sinusoid to correspond to the time at which the waveform has the highest frequency detectable by the LIGO detector band. The next sinusoid is chosen by moving to an earlier time, $t_{l+1} = t_l - T_l$. Since we want the l th sinusoid to be mostly present on the interval $t_l - T_l < t < t_l$, we choose the damping factor to be $\gamma_l = \beta/T_l$, where β is a tunable parameter. This procedure is repeated until the time t_l corresponds to a time in the waveform that has frequency below the LIGO detector band. Hence the number of sinusoids is dependent on the value of ϵ , the rate of frequency change $\ddot{\phi}(t)$, which is dependent on the masses of the system, and the detector bandwidth. For more information on this procedure, see [12].

We can now approximate the components $\hat{h}(t) = A_{1\text{Mpc}}(t)e^{i\phi(t)}$ by an addition of a series of damped sinusoids $u(t)$ with cutoff times t_l :

$$A_{1\text{Mpc}}(t)e^{i\phi(t)} \simeq U(t) = \sum_l u_l(t) = \sum_l b_{0,l} e^{(\gamma_l+i\omega_l)(t-t_l)} \Theta(t_l-t). \quad (38)$$

Figure 3 shows an illustration of how damped constant-frequency sinusoids can add to give an inspiral-like waveform.

F. Summed parallel IIR filtering

Each complex sinusoid $u_l(t)$ in Eq. (38) can be searched for in the data x using the single-pole IIR filter (30). Here the cutoff time is incorporated by running each filter on a delay, $d_l = t_l/\Delta t$. The output of the l th filter at time k is

$$y_{k,l} = a_{1,l} y_{k-1,l} + b_{0,l} x_{k-d_l}. \quad (39)$$

The linear summation of the output of *all* filters is the cross correlation of the data x and the approximate waveform $U(t)$ in (38):

$$z_k \simeq 2\Delta t \sum_l y_{k,l}. \quad (40)$$

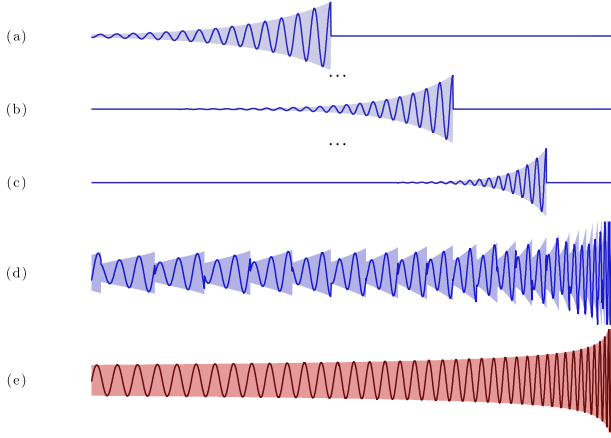


FIG. 3 (color online). An illustrative diagram demonstrating the ability to linearly sum exponentially increasing constant-frequency sinusoids to approximate an inspiral-like waveform. The top three panels (a–c) show three example sinusoids with different damping, frequency and cutoff time factors. Panel (d) shows the linear addition of all the sinusoids (at different scales). Panel (e) shows the exact inspiral-like waveform. Note that this figure is only for illustrative purposes.

Here z is equivalent to the value computed by the discrete time-domain two phase filter (28) when using a template $\hat{h}(t) = U(t)$. From Eq. (22), it follows that the absolute value of the summation (40) divided by σ_U is the SNR, which we term the output of the SPIIR. The normalization factor σ_U is defined as

$$\sigma_U^2 = 4 \int_0^\infty \frac{|\tilde{U}_c(f)|^2}{S_n(|f|)} df, \quad (41)$$

where $\tilde{U}_c(f)$ is the Fourier transform of the real part of $U(t)$, which approximates $h_c(t)$. The similarity of the SPIIR output and the matched filter output will depend on how well $U(t)$ approximates the given template.

III. IMPLEMENTATION FOR PERFORMANCE TESTING

A. IIR bank construction

To confirm the ability of the SPIIR method to recover a good SNR, it is first necessary to show that the approximate inspiral waveform (38) is a good “match” to the theoretical inspiral waveform (12). We define the *overlap* Δ as the inner product of the normalized approximate waveform U and the template h :

$$\Delta = \frac{1}{\sigma \cdot \sigma_U} \sqrt{\left(2 \int_{-\infty}^{\infty} \frac{\tilde{h}_c(f) \tilde{U}_c^*(f)}{S_n(|f|)} df\right)^2 + \left(2 \int_{-\infty}^{\infty} \frac{\tilde{h}_s(f) \tilde{U}_s^*(f)}{S_n(|f|)} df\right)^2}, \quad (42)$$

where $U_s(t)$ approximates $h_s(t)$. We initially approximate a canonical 2PN 1.4–1.4 M_\odot inspiral waveform band-limited to 10–1500 Hz using the value of the tunable parameters ϵ , α and β to be consistent with the high overlap results of [12]. With some minor variation of their values, we aim to recover the highest overlap possible. Once a good choice of α and β is found for the 2PN 1.4–1.4 M_\odot template, we use the same values for other templates, but vary the value ϵ (and consequently the number of IIR filters in each bank) to see the effect on overlap.

B. Detector data simulation

To test the detection efficiency of the SPIIR method compared to the frequency domain matched filter, we will filter two mock signals, one for which the input data are just LIGO-like noise, and the other with the same noise plus an inspiral waveform injection scaled to represent a source at a chosen effective distance D_{eff} .

For this test, we need to construct a finite segment of detector data to filter. Being infinite impulse response filters, in principle the filters should be run for an infinite length of the input data before the output has stabilized. In order to approximate this behavior, we need to run the IIR bank for a finite “warm-up” period before the output is consistent with that of an IIR filter that has been running

for an infinite amount of time. We choose to run each filter for 2 e -foldings of time before we accept the output as being identical to one which has run for an infinite amount of time. Additionally, since each IIR filter in the bank runs on a delay, the summed output of *all the IIR filters* will not be produced until after the longest delay time (d_{max}) has passed. The filter that has the longest delay (d_{max}) is also the one that has the longest decay rate γ_{max} . In total, the input data must be at least $d_{\text{max}} + 2\gamma_{\text{max}}^{-1}$ in length before any output is produced. Hence the length of the input data is

$$N_{\text{input}} = d_{\text{max}} + 2\gamma_{\text{max}}^{-1} + N_{\text{analysis}}, \quad (43)$$

where N_{analysis} is the length of analysis period, which we choose to be 4 seconds. Hence the 4 s SPIIR output will tell us whether there is an injection that ended somewhere within those 4 seconds. At a sample rate of 4096 Hz, the analysis period is $N_{\text{analysis}} = 16834$ data points long. In our simulation, we find $d_{\text{max}} = 4081683$ and $2\gamma_{\text{max}}^{-1} = 149432$, resulting in $N_{\text{input}} = 4247499$.

1. Noise generation

The LIGO-like noise data is produced by creating a normally distributed white noise time series of length N_{input} , then coloring it by the theoretical Advanced LIGO

noise spectrum $S_n(f)$ (A1). We then over-whiten this time series using Eq. (25) to produce the waveform-free noise input data x :

$$x_{\text{noise}}(t) = n^{\text{ow}}(t). \quad (44)$$

2. Waveform injection

We create our waveform injections by first producing an inspiral waveform band-limited between 10 and 1500 Hz. The injection is padded with zeros so that it has the length N_{input} . The end of the waveform is chosen so that it finishes somewhere after $d_m + 2\gamma_m^{-1}$ data points. The injection signal is then over-whitened using Eq. (25). The over-whitened injection can then be placed in the over-whitened noise signal:

$$x_{\text{noise+injection}}(t) = x_{\text{noise}}(t) + h^{\text{ow}}(t). \quad (45)$$

3. Matched filter comparison

As a comparison, we will also perform a frequency domain correlation matched filter. For this process, since the input data is already over-whitened, it only needs to be cross-correlated with the waveform. Section II B outlines how this is done. The cosine component $h_c(t)$ gets pre-padded with enough zeros to get to length N_{input} . This ensures that $\tilde{h}_c(f)$ has the same spectral resolution as $\tilde{s}(f)$. The matched filter (21) produces a time series of N_{input} length. However the first $N_{\text{input}} - N_{\text{analysis}}$ data points are erroneous wrap-around caused by the fast-Fourier transform. Only the interval $[N_{\text{input}} - N_{\text{analysis}} + 1, N_{\text{analysis}}]$ is used to determine if a waveform is present.

C. Detection efficiency

To test the detection efficiency of the SPIIR method compared to the traditional matched filter method we will construct several receiver operating characteristic (ROC) curves for 2PN 1.4-1.4 M_\odot waveforms injected for different effective distances D_{eff} . To create each ROC curve, we first find the false alarm rate. The false alarm rate is found by realizing an N_{input} length LIGO-like noise time series, filtering this input data, and analyzing the output of the 4 s analysis period (the SNR). We will count this realization as a false positive if at any point within the 4 seconds the SNR goes over a given SNR threshold. Several thresholds will be chosen, giving the false positive as a function of threshold. After $>10^6$ noise realizations, the false alarm rate is simply the ratio of total number of false positives to number of noise realizations. Likewise, to see if the IIR filter doesn't miss too many true positives, we inject a 2PN 1.4-1.4 M_\odot waveform using the prescribed method in III B 2 for a given D_{eff} into LIGO-like noise. After filtering, if at any point within the analysis period the SNR is above a given threshold, this realization is counted as a true positive. Again, after $>10^6$ noise realizations, we calculate the

detection rate as a ratio of the total number of true positives to number of realizations. The plot of false alarm rate versus detection rate gives the ROC curve.

IV. RESULTS

A. Inspiral waveform overlap

Starting with the canonical 1.4-1.4 M_\odot second order post-Newtonian binary waveform band limited to be between 10 and 1500 Hz we found, using the parameters $\epsilon = 0.04$, $\alpha = 0.99$, $\beta = 0.25$ in the procedure outlined in Sec. II E, that we can recover an overlap of 99% using 687 IIR filters.

We find that increasing the value of ϵ will in general increase the overlap, as the frequency space is more finely sampled. However there seems to be a limit, as the damping factor γ causes the adjacent IIR filters to run into each other.

With this choice of α and β we are able to recover a high overlap for different mass pairs as well. Figure 4 shows the overlap as a function of number of IIR filters for six different mass pairs.

B. Ability to recover SNR

Figure 5 shows the SNR produced from both the matched filter technique and the SPIIR method. The input time series is constructed following Sec. III B. The injection of a 2PN 1.4-1.4 M_\odot waveform scaled for an effective distance of 500 Mpc is added to LIGO-like noise. The x axis of the plot is centered about the end of the injection ($t = \tau_c$), which is directly in the middle of the analysis period. Around this time, the SNR peaks to 8, which is near the expected value of 7.9 for an injection at this distance. This plot shows that the SPIIR method is capable of recovering a very similar SNR to the matched filter at all times.

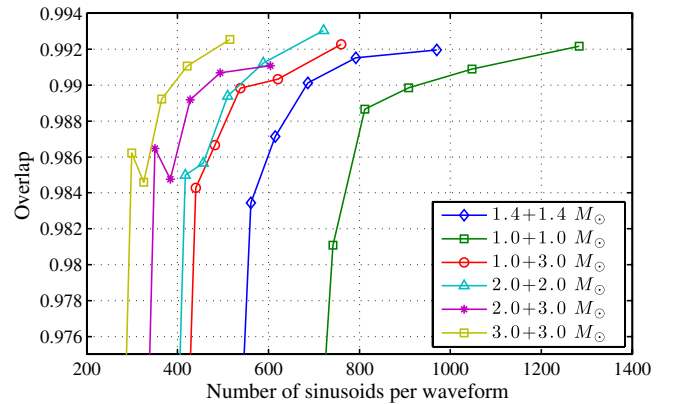


FIG. 4 (color online). The overlap between the exact inspiral waveform and the approximate inspiral waveform as a function of number of damped sinusoids. In general the greater the number of sinusoids per waveform, the greater the overlap. However the choice of γ/s greatly affects the overlap.

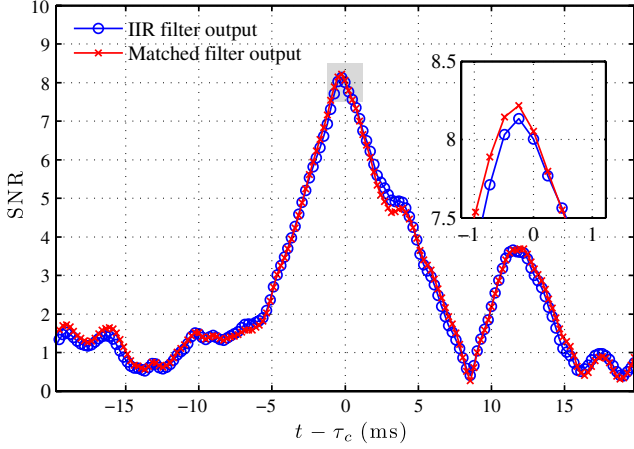


FIG. 5 (color online). The SNR output of both the SPIIR method and a traditional matched filter method. The plot is centered on $t - \tau_c$ where τ_c is the time at which the injection ends. From the two curves, it is clear that the SPIIR method can return a very similar SNR to that from the optimal filter.

C. Detection efficiency

We analyzed over 10^6 independent noise realizations, for which the waveform had been injected at D_{eff} of 500, 600, 700, 800 Mpc. We performed both IIR filtering and traditional matched filtering. Figure 6 shows that the SPIIR method recovers most of the same events as the traditional matched filter method. At false alarm rates of greater than 10^{-5} , the SPIIR method recovers greater than 99% of the injections recovered by the matched filter when searching for injections at an effective distance of 500 Mpc ($\text{SNR} \sim 8$). Even in the worst case, at a false alarm rate of 10^{-6} , the SPIIR method catches 4.5% of injections scaled at an extreme 800 Mpc ($\text{SNR} \sim 5$), whereas the matched filter catches 5% of injections at this scale.

V. SUMMARY AND DISCUSSION

We have demonstrated that, through the use of a parallel bank of single-pole IIR filters, it is possible to approximate the SNR derived from the matched filter with greater than 99% overlap. The main advantage of our SPIIR method is that it operates completely in the time domain, and in principle it has zero latency (not taking into account whitening or computational time). The SPIIR method recovers most of the injections the optimal matched filter recovers.

The use of a bank of simple IIR filters for each template as opposed to the matched filter method enables us get two extra processes for minimal additional cost. The first is that the individual IIR filter outputs can be arranged into groups, such that their total summed output is roughly independent and orthogonal to each other. This enables, with minimal extra overhead, the calculation of a χ^2 distributed statistic, giving a secondary method of verification. We will demonstrate this in an upcoming paper. The

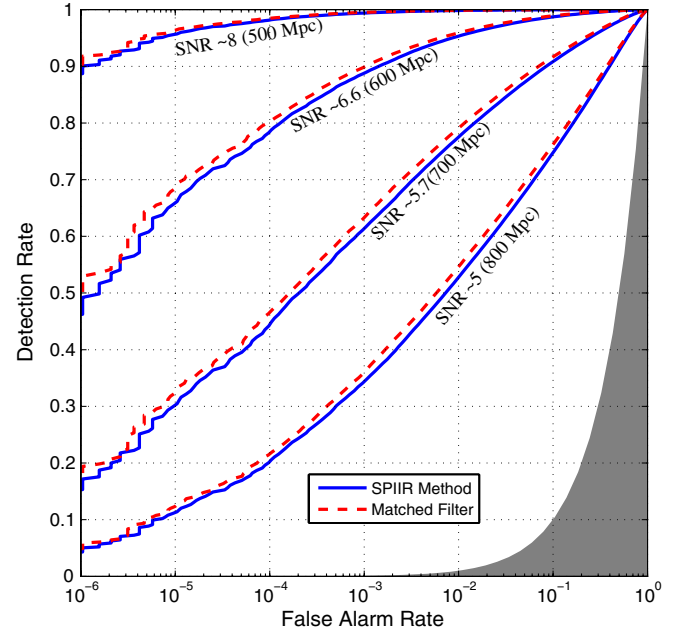


FIG. 6 (color online). The receiver operating characteristics (ROC) of both the IIR filter method and the traditional matched filter method. The x axis shows the false alarm rate, and the y axis the detection rate. A one-to-one relationship, which is the worst case scenario, is shown by the boundary of the shaded area. We show four different ROC curves, where each curve represents the detection rate as a function of false alarm rate for waveforms injected at effective distances of 500, 600, 700 and 800 Mpc ($\text{SNR} \sim 8, 6.6, 5.7$ and 5 respectively).

second natural advantage of using a parallel bank of single-pole IIR filters is that they can easily be executed in parallel using multithreaded processors, such as graphics processing units. Indeed, a side study has shown that this is possible [26]. This leads to the future possibility that a single personal computer may be able to process the detection of GWs from inspiralling compact binaries.

A further way to reduce the computation of the IIR calculation is to split the incoming data into differently down-sampled channels. The output of each IIR filter in the bank is the correlation of a fixed frequency sinusoid and the incoming data. For the sinusoids that have frequencies < 124 Hz, the incoming data need only be sampled at 256 Hz. The current pipeline of LLOID uses a similar multichannel down-sampling in their detection pipeline. Their pipeline consists of the integration of the open-source real-time multimedia handling software GStreamer and the LIGO Algorithm Library (LAL) [14]. This software library is an ideal platform to integrate the SPIIR method. The total computation can also be further reduced by sharing IIR filters (via interpolation) between different templates [12].

Although the design of the IIR filter so far only applies to chirping, post-Newtonian approximation inspirals, we have performed preliminary tests using more complicated

combinations of single-pole IIR filters to replicate the waveform of an inspiral with spin. If the amplitude/frequency beating of a spinning inspiral waveform can be simulated by the linear addition of two different nonspinning inspirals with different masses, then it can be approximated by a linear addition of damped sinusoids. In this case, the SPIIR method can produce the SNR for the beating waveform. There is also the possibility of using higher order IIR filters, although designing the coefficients can be very difficult.

We foresee that the use of IIR filters for time-domain filtering of Advanced LIGO will be ideal, as the waveforms will be much longer. The frequency domain matched filter will take more time to calculate GW triggers, essentially ruling out the possibility of triggering the detection of prompt optical emission related to neutron star mergers (GRBs). We have shown that the use of a parallel bank of IIR filters requires less computational cost, with minimal detection rate loss, and most importantly can be calculated in the time domain with near zero latency.

ACKNOWLEDGMENTS

We would like to thank Kipp Cannon, Drew Keppel and Chad Hanna for detailed discussion on the design and

implementation of low-latency detection algorithms. This work was done in part during the LIGO Visiting Student Researcher program, which was partially funded by the 2009 UWA Research Collaboration Award. This research was supported by the Australian Research Council. S. H. gratefully acknowledges the support of an Australian Postgraduate Award. L. W. acknowledges the support of the Australian Research Council Discovery Grants and Future Fellow program.

APPENDIX: NOISE-SPECTRAL DENSITY

The noise-spectral density (in units of strain/ $\sqrt{\text{Hz}}$) we use is based on an algebraic expression prediction of the Advanced LIGO noise curve given in the LAL suite reference manual [27] defined by

$$S_h(f) = S_0 \left\{ \left(\frac{f}{f_0} \right)^{-4.14} - 5 \left(\frac{f}{f_0} \right)^2 + 111 \left(\frac{1 - \left(\frac{f}{f_0} \right)^2 + 0.5 \left(\frac{f}{f_0} \right)^4}{1 + 0.5 \left(\frac{f}{f_0} \right)^2} \right) \right\}, \quad (\text{A1})$$

where $f_0 = 215$ Hz and $S_0 = 10^{49}$ strain/ $\sqrt{\text{Hz}}$.

-
- [1] <http://www.ligo.org>.
 - [2] <http://www.virgo.infn.it/>.
 - [3] J.R. Smith *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **26**, 114013 (2009).
 - [4] Advanced LIGO Team, LIGO Technical Report No. M060056-08-M, 2007 (unpublished).
 - [5] B. P. Abbott *et al.* (LIGO Scientific Collaboration), *Phys. Rev. D* **79**, 122001 (2009).
 - [6] D. B. Fox *et al.*, *Nature (London)* **437**, 845 (2005).
 - [7] E. Nakar, *Phys. Rep.* **442**, 166 (2007).
 - [8] M. H. P. M. van Putten, [arXiv:0905.3367](https://arxiv.org/abs/0905.3367).
 - [9] B. Zhang and P. Mészáros, *Int. J. Mod. Phys. A* **19**, 2385 (2004).
 - [10] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise* (Prentice-Hall, Englewood Cliffs, NJ, 1962).
 - [11] B. Abbott *et al.* (The LIGO Scientific Collaboration), *Phys. Rev. D* **77**, 062002 (2008).
 - [12] J. Luan, S. Hooper, L. Wen, and Y. Chen, *Phys. Rev. D* **85**, 102002 (2012).
 - [13] D. Buskulic *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Classical Quantum Gravity* **27**, 194013 (2010).
 - [14] <https://www.lsc-group.phys.uwm.edu/daswg/projects/gstlal.html>.
 - [15] K. Cannon, A. Chapman, C. Hanna, D. Keppel, A. C. Searle, and A. J. Weinstein, *Phys. Rev. D* **82**, 044025 (2010).
 - [16] W. G. Anderson, P. R. Brady, J. D. E. Creighton, and É. É. Flanagan, *Phys. Rev. D* **63**, 042003 (2001).
 - [17] B. Abbott *et al.* (LIGO Scientific Collaboration), *Phys. Rev. D* **69**, 122001 (2004).
 - [18] L. Blanchet, T. Damour, and B. R. Iyer, *Phys. Rev. D* **51**, 5360 (1995).
 - [19] L. Blanchet, B. R. Iyer, C. M. Will, and A. G. Wiseman, *Classical Quantum Gravity* **13**, 575 (1996).
 - [20] B. Allen, *Phys. Rev. D* **71**, 062001 (2005).
 - [21] S. Droz, D. J. Knapp, E. Poisson, and B. J. Owen, *Phys. Rev. D* **59**, 124016 (1999).
 - [22] B. Abbott *et al.* (LIGO Scientific Collaboration), *Phys. Rev. D* **72**, 082001 (2005).
 - [23] P. R. Brady and S. Fairhurst, *Classical Quantum Gravity* **25**, 105002 (2008).
 - [24] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
 - [25] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975).
 - [26] Y. Liu (unpublished).
 - [27] https://www.lsc-group.phys.uwm.edu/daswg/projects/lal/nightly/docs/html/LALAdvLIGOPsd_8c_source.html.