Application of a zero-latency whitening filter to compact binary coalescence gravitational-wave searches

Leo Tsukada,1,2,* Kipp Cannon,1,† Chad Hanna,3 Drew Keppel,4 Duncan Meacher,3 and Cody Messick3

1Research Center for the Early Universe (RESCEU), Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan
2Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan
3Department of Physics and Astronomy and Astrophysics The Pennsylvania State University, University Park, Pennsylvania 16802, USA
4California Institute of Technology, Pasadena, California 91125, USA

(Received 22 August 2017; published 14 May 2018)

Joint electromagnetic and gravitational-wave (GW) observation is a major goal of both the GW astronomy and electromagnetic astronomy communities for the coming decade. One way to accomplish this goal is to direct follow-up of GW candidates. Prompt electromagnetic emission may fade quickly, therefore it is desirable to have GW detection happen as quickly as possible. A leading source of latency in GW detection is the whitening of the data. We examine the performance of a zero-latency whitening filter in a detection pipeline for compact binary coalescence (CBC) GW signals. We find that the filter reproduces signal-to-noise ratio (SNR) sufficiently consistent with the results of the original high-latency and phase-preserving filter for both noise and artificial GW signals (called “injections”). Additionally, we demonstrate that these two whitening filters show excellent agreement in $\chi^2$ value, a discriminator for GW signals.

DOI: 10.1103/PhysRevD.97.103009

I. INTRODUCTION

Detecting gravitational wave signals from coalescing compact objects such as neutron stars and black holes is a primary goal for ground-based gravitational-wave detectors. The advanced LIGO [1] and advanced Virgo gravitational wave detectors [2] have successfully detected gravitational waves from several black hole binaries [3–6] following the initial binary black hole discovery of GW150914 [7]. Most recently, LIGO and Virgo detected the first ever neutron star collision, known as GW170817 [8]. Quite remarkably, GW170817 was detected in coincidence with a short gamma-ray burst (SGRB), which has long been expected to be one of the most promising electromagnetic (EM) counterpart candidates to binary neutron star mergers [9]. Furthermore, subsequent electromagnetic emission spanning from radio up to x-ray was detected within hours to weeks after the GW emission [10].

The association of GW170817 with electromagnetic transients across many wavelengths and the discoveries that followed were initiated by the rapid identification of the gravitational wave signal by the GstLAL detection pipeline [11]. Until the detection of GW170817, the question of what data processing latency is required, scientifically, for such a GW detection system to be most effective at enabling multimessenger astrophysics has been open. Theoretical work on GRB models has proposed a vast range of the time delay between GW emission and the onset of the following SGRB, from $< 10$ s [12] to $10^3$ s to $10^4$ s [13]. Now it is known the time lag is approximately 2 s [8], which motivates achieving alert latencies below 2 s.

Within the low-latency analysis paradigm, the GW detection problem consists of three main stages: data calibration and distribution, candidate identification, and alert distribution [11,14]. The GstLAL project (publicly available in [15]) began in 2009 in order to produce near zero-latency gravitational wave candidate identification [14]. The GstLAL candidate identification achieved latencies of approximately one minute in advanced LIGO’s first observing run. For example, the binary black hole known as GW151226 was detected within 70 s [3] by the GstLAL-based compact binary coalescence detection pipeline. Also, the GstLAL has inspired related works to pursue a low-latency search for CBC GW signals, such as J. Luan et al. [16], which proposes an algorithm to reduce computational cost for matched filtering by introducing chains of infinite impulse response filters.

This work addresses one of the open questions put forward in [14] for how to achieve theoretically zero latency candidate event identification, namely, data whitening, which was the largest source of candidate identification latency for the GstLAL pipeline in advanced LIGO’s first observing run. Data whitening is just one part GstLAL’s
modular framework for gravitational wave candidate event identification which also includes matched filtering an orthogonal decomposition of gravitational wave templates [17], and using a multidimensional likelihood ratio based ranking statistic to distinguish noise from signal [18]. The overarching pipeline is described in [11] and we will not elaborate further on the pipeline here.

In this work, we describe how the data whitening filter can be optimized in terms of latency. In particular, we implement a zero-latency algorithm with a linear finite-impulse-response (FIR) filter. We use the minimum-phase FIR filter approximation technique described by Damera-Venkata et al. in [19]. Other algorithms for whitening filter approximation are available, for example [20–24]. In principle, the zero-latency algorithm can be used by any GW data-analysis pipeline. For the study presented here we used the GstLAL [11] detection pipeline and note that this whitening algorithm was used during the detection of GW170817. In order to verify the suitability of the zero latency whitening algorithm, we compare relevant parameters of simulated gravitational wave candidates, namely SNR and $\chi^2$, to the original frequency-domain whitening algorithm. Since, as will be shown below, we find the Damera-Venkata et al., algorithm to meet the needs of low-latency searches for compact-object mergers, we have not investigated the performance of other techniques at this time.

This paper is organized as follows. In Sec. II, we give an overview of statistical method to analyze CBC GW signals and the comparison between the frequency-domain and zero-latency whitening algorithms. In Sec. III, we present the performance tests of the zero- latency whitening filter, including the comparison to the original whitening filter. Lastly, we conclude in Sec. IV.

II. METHOD

A. Matched filtering and $\chi^2$ test

One statistic used to estimate the detection significance of GW signals is the signal-to-noise ratio (SNR), $\rho$, computed using a matched filter [25]. For the calibrated strain data $s(t)$ and a template waveform $h(t)$, the output of the matched filter is

$$z = (s(t), h(t)) \equiv 4 \int_0^\infty \frac{\tilde{h}(f)\tilde{s}(f)}{S_n(f)} df,$$

where $S_n(f)$ is the one-sided power spectral density of the detector strain noise. $\rho$ is defined as the output of the matched filter in the case of a normalized GW signal

$$\rho \equiv \frac{z}{\sigma},$$

where

$$\sigma^2 \equiv \langle h(t), h(t) \rangle.$$

Additionally, strain data contain noise transients which do not obey a Gaussian distribution and may accidentally produce high $\rho$. Such non-Gaussian transient (referred to as “glitches”) are distinct from real GW signals in that they do not have the morphology of the template $h(t)$. Making use of this distinction, we employ another statistic, $\chi^2$, defined below, in order to distinguish the transients [11]. The time-dependent SNR of data is compared with that expected from a real signal using the autocorrelation function of the template at its time of peak amplitude, $R(t)$. A $\chi^2$ value is computed for each trigger using the time-dependent SNR $\rho(t)$, the peak SNR $\rho_p$ at the time stamp of $t_p$, the noise-weighted autocorrelation function of a template $R(t)$.

$$\chi^2 = \frac{1}{\mu} \int_{t_p-\delta t}^{t_p+\delta t} |\rho(t) - \rho_p R(t)|^2 dt$$

where

$$\rho(t) = \frac{4}{\sigma^2} \int_0^\infty \frac{\tilde{h}(f)\tilde{s}(f)}{S_n(f)} e^{2\pi if(t-t_p)} df$$

and

$$R(t) = \frac{4}{\sigma^2} \int_0^\infty \frac{\tilde{h}(f)^2}{S_n(f)} e^{2\pi if(t-t_p)} df$$

The factor $\mu$ is to normalize the $\chi$ value for a well-fit signal. The time window $\delta t$ is a tunable parameter.

Both SNR and $\chi^2$ values are used to derive a likelihood ratio necessary for ranking triggers [18].

B. Frequency-domain whitening filter

As can be seen in (1), a matched filter is interpreted simply as an inner product between $\tilde{s}(f)$ and $\tilde{h}(f)$ with the weight of $1/\sqrt{S_n(f)}$ for each. This weighting process is called “whitening,” named after the fact that the transformation ideally returns only white noise. Referring to Fig. 1, the power spectrum density (PSD) of 32 s chunks of input data is measured for the subsequent whitening (i). Since the discrete Fourier transform (DFT) processes a chunk of a time series as if it were a periodic infinite series, we apply a Hann window function (ii) to suppress discontinuities on the boundaries of each period. These discontinuities will lead to click noise in the whitened data and may produce fake GW signals. Furthermore, in order to prevent any remaining discontinuity on the boundaries, some additional samples around the 32 s time series have been filled with zeros after applying the Hann window. Then, the spectrum of the Hann-windowed block is computed (iii) and the block is whitened with the PSD (iv). After the inverse DFT (v), we further apply a Tukey window (vi). The purpose of this windowing is to suppress time-domain leakage which appears through the whitening and IDFT. The above procedure is repeatedly applied for
every 32 s block with a 50% overlap and the PSD is updated every 16 s. In the end, all whitened data chunks, each of which is separately processed, are added with a consecutive 16 s shift in order to output a continuous time series of the whitened data (vii). The algorithm’s main drawback is latency. Since a 32 s block is processed all at once every 16 s, the latency depends on the sample’s location in the block, and can be anywhere from 16 s to 32 s.

C. Time-domain whitening filter

1. FIR-filter-based algorithm

Here, we present an alternative FIR-filter-based algorithm to the frequency-domain whitening described above. For the given LIGO strain data, the square root of the inverse PSD is employed to construct the FIR of a linear-phase filter shown in Fig. 2. Therefore, it is possible to replace step (iv) in Fig. 1 with this FIR filter. It should be noted that due to its peak location, the FIR-filter-based algorithm still has a latency of 16 s. Figure 3 is amplitude and phase responses of the filter. The phase response shows no phase shift during this whitening process, which guarantees preserving the phase of the data.

2. Zero-latency algorithm

According to the discussion in the previous paragraph, the peak of the filter must be moved to the left for the latency reduction. It is not possible to change the filter’s latency without changing the whitening transformation. The result will be an approximation of the original filter. We adopt the technique of Damera-Venkata et al. [19] which derives a minimum-phase approximation of the desired filter by applying a discrete Hilbert transform to the logarithm of a given magnitude response. Using more samples for the given magnitude response, one can more accurately approximate the magnitude response of the computed minimum-phase filter. The result is shown in Fig. 4. In both Figs. 2 and 4, the time origin can be uniquely determined by the requirement that the time stamps at which artificial GWs (called “injections”) are recovered be preserved (described in Sec. III C and Figs. 12 and 13). Thus, Fig. 4 indicates that the FIR filter does not use any information of future input samples, which is the reason
why it is called a “zero-latency whitening filter”.\textsuperscript{1} As described in Sec. II B, this whitening filter is equally applied to both of the templates and the strain data. This can be expressed by writing (1) as

\begin{equation}
 z = 4 \int_0^\infty \left( \frac{\tilde{h}(f)}{\sqrt{S_n}} \right)^* \left( \frac{\tilde{s}(f)}{\sqrt{S_n}} \right) \text{d}f. \tag{7}
\end{equation}

The desired whitening transformation is a frequency-dependent scale factor ($\sqrt{S_n(f)}$ is real-valued, it does not alter phases). We approximate this transformation with a purely causal FIR filter. The difference between the actual transformation and the desired transformation can be described by introducing an error factor in the frequency domain representation. If $1/\sqrt{S_n(f)}$ is the transformation we wish to apply to $\tilde{h}(f)$ and $\tilde{s}(f)$, let $\delta \sqrt{S_n(f)}/\sqrt{S_n(f)}$ be the transformation performed by the causal FIR approximation, so

\begin{equation}
 z = 4 \int_0^\infty \left( \frac{\tilde{h}(f)}{\sqrt{S_n}} \right)^* \left( \frac{\tilde{s}(f)}{\sqrt{S_n}} \right) \delta \sqrt{S_n} \text{d}f \tag{8}
\end{equation}

\begin{equation}
 = 4 \int_0^\infty \left( \frac{\tilde{h}(f)}{\sqrt{S_n}} \right)^* \left( \frac{\tilde{s}(f)}{\sqrt{S_n}} \right) |\delta \sqrt{S_n}|^2 \text{d}f. \tag{9}
\end{equation}

\textsuperscript{1}We ignore the computation time as the time required to compute each output sample is significantly less than the sample period of the data stream.

We see the matched filter output is insensitive to $\arg(\delta \sqrt{S_n})$, the phase errors arising from the causal FIR approximation (shown in Fig. 5), it is sensitive only to the amplitude errors. Figures 6–8 show that the causal FIR filter approximation succeeds in whitening the data, producing zero

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{Fig3}
\caption{The amplitude phase response of the linear-phase whitening filter. The Nyquist frequency is 2048 Hz.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{Fig4}
\caption{The impulse response of the zero-latency whitening filter. The impulse response is zero for all negative times, indicating that this filter is purely causal.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{Fig5}
\caption{The amplitude and phase response of the minimum-phase whitening filter. The Nyquist frequency is 2048 Hz.}
\end{figure}
mean, unit variance, stationary (nearly) white Gaussian noise, so the amplitude errors are not significant.

3. Smooth PSD tracking

In order to replace the whole algorithm in Fig. 1, we have implemented an alternative method of the windowing process along with the whitening transformation. Specifically, we have allowed the PSD transition to occur continuously. Here, we have created a function which returns a linear combination of the newest and next newest filters during their transition as described by (10). The coefficient of the newer filter smoothly shifts from zero to one, sample by sample, according to a sinusoidal function. The zero-latency algorithm applies this function recursively any time a new whitening filter becomes available.

\[
    s'(t) = \begin{cases} 
    s_{\text{old}}(t) & (t < t_{\text{up}}) \\
    s_{\text{old}}(t) \cos \frac{\pi (t - t_{\text{up}})}{2\Delta t_{tr}} + s_{\text{new}}(t) \sin \frac{\pi (t - t_{\text{up}})}{2\Delta t_{tr}} & (t_{\text{up}} \leq t < t_{\text{up}} + \Delta t_{tr}) \\
    s_{\text{new}}(t) & (t_{\text{up}} + \Delta t_{tr} \leq t) 
    \end{cases}
\]

where \( s'(t) \) is the resulting filter, \( s_{\text{old,new}} \) is the FIR of an older and newer filter respectively, \( t \) is the current time stamp, \( t_{\text{up}} \) is the time stamp when the PSD is updated and \( \Delta t_{tr} \) is the duration of the filter transition. Particularly, we set \( \Delta t_{tr} \) as 0.125 s in this work so that the transition timescale lies outside the frequency band of interest, which starts at 10 Hz. Therefore, the detail of the transition does not affect the detectability of GW signals. Note that this method is not unique in the application of a whitening filter but can be used for other time-dependent filtering.
III. TESTS

Unlike the frequency-domain algorithm described in Sec. II B, the zero-latency whitening filter does not conserve the phase of input data, which potentially harms the GW detectability. Therefore, it is necessary to demonstrate how significantly this change affects the resulting SNR and $\chi^2$.

Here, we implement each of the zero-latency and frequency-domain whitening filters in the CBC gstLAL pipeline, which is compiled in LSC Algorithm Library (LAL) [26]. The pipeline scans LIGO strain data for any GW signal candidates (called “triggers”) and computes SNR and $\chi^2$ for each trigger. In this work, we employ strain data from H1 with a duration of 45,056 s (from 08:25:23 to 20:56:19 UTC on 2005/11/27) during the fifth science run, called S5 [27]. A template bank is used spanning: component masses $3 M_\odot \leq m_1, m_2 \leq 6 M_\odot$; total mass $6 M_\odot \leq M_{\text{total}} \leq 12 M_\odot$; a minimal match of 97%; sampling frequency of 2048 Hz; non-spinning waveform to second post-Newtonian (PN) order. Along with statistical tests described below, we conduct two kinds of tests, namely, a noise-based and an injection-based test. In the noise-based test, the pipeline computes SNR and $\chi^2$ from the strain data with no GW signal. Therefore, all triggers in this test arise from detector noise which accidentally produce higher SNR than the threshold. On the other hand, the injection-based test requires artificial GW signals, and so we add injections to the same strain data. An injection is generated once every 31.4 s in the data, so the number of injections amounts to 1435. These two tests examine the agreement between the two whitening filters for noise and signals.

A. Statistical tests

Figures 6 and 7 show the auto-correlation and amplitude histogram created from an output stream of the zero-latency whitening filter. For comparison, expected curves are shown as dashed lines in the both figures, each of which indicates a delta-function with some variance and a Gaussian distribution respectively. Both plots show good agreement between the output and pure white noise. In particular, there is no apparent peak of the auto-correlation at around 100 Hz, at which the amplitude response shows its peak (See Fig. 5). Also, the averaged power spectral density is shown in Fig. 8. The spectrum is flatten throughout the shown frequency domain. Therefore, we conclude that the zero-latency algorithm sufficiently functions as a whitening filter.

B. Noise-based test

In Fig. 9, we plot SNR and $\chi^2$ computed for each noise trigger by the zero-latency whitening filter versus the frequency-domain one. Here, we associate each counterpart by spotting a pair of triggers within the end-time window of $10^{-1}$ s and with identical component masses ($m_1$ and $m_2$). The two whiteners produced triggers with an SNR of 5–60 and a $\chi^2$ of $10^{-1}$–$10^3$. Figure 9 presents good agreement of both SNR and $\chi^2$ between the two whitening filters.

**FIG. 9.** Scatter plots of SNR and $\chi^2$ computed by the frequency-domain and zero-latency FIR whiteners for noise triggers. The value of SNR and $\chi^2$ range 5–60 and $0.2 \cdot 10^3$ respectively. We have associated a pair of triggers within the end-time window of $10^{-1}$ s and with identical component masses, $m_1$ and $m_2$. 

103009-6
**C. Injection-based test**

Figure 10 shows SNR and $\chi^2$ computed with the zero-latency whitening filter versus the frequency-domain one in the presence of injections. Also, coalescence phase for every injection is shown in Fig. 11. The coalescence phase is a phase of injection waveform at the coalescence time and determined by the ratio between cosine and sine components of a chosen template. In this test, we first simulate waveforms, based on a collection of parameters chosen randomly from a given probability distribution: $m_1$ and $m_2$ from a Gaussian distribution with the mean of $4.5 M_\odot$ and the standard deviation of $0.5 M_\odot$; cosine of the inclination chosen from a uniform distribution; nonspinning waveform to second post-Newtonian (PN) order with the cutoff frequency of 30 Hz. Next, the pipeline searches for and extracts injection by spotting the one with the highest SNR among all located within 1 s of its true end time.

**FIG. 10.** Scatter plots of SNR and $\chi^2$ computed by the frequency-domain and zero-latency FIR whiteners for injection triggers. First, 1289 injections with the SNR of $5 \times 10^4$ were generated and truncated so that the maximum SNR would be 100. This is because triggers with higher SNR do not help to examine the critical discernibility of GW signals. Consequently, the values of SNR and $\chi^2$ range 5–100 and 0.3–400 respectively. The smaller plot located on the left panel shows the data with SNR < 10.

**FIG. 11.** The scatter plots of coalescence phase computed by the frequency-domain and zero-latency FIR whiteners for injection triggers. These triggers, whose SNR fall into the range from 5 to 100.

**FIG. 12.** Histogram of the discrepancy between the true and estimated end time for each injection in case of the frequency-domain whitening filter. Each bin has the number of injection triggers whose end-time difference lies in the bin’s range.
time. After the trigger extraction, SNR, $\chi^2$ and coalescence phase of every injection trigger is recorded. In the end, only those with an SNR less than 100 are left to fit into the scatter plot, Fig. 10. The above procedure is conducted for both whitening filters and we identify each counterpart by the end time of each injection. We have also conducted a consistency test for an injection’s end time in the case of the two whitening filters. Figures 12 and 13 show histograms of the discrepancy between the true and estimated end time of each of the filters. In the both figures, the central peak has a tail width of $\sim 100$ ms, which is consistent with the typical tail width of the autocorrelation function of injected waveforms, suggesting that the pipeline properly generates and extract the injections from the triggers. As a result of the injection-based test, we find the SNR and $\chi^2$ computed with the zero-latency whitening filter to agree with those of the frequency-domain one. Although some triggers indicate that the new whitening filter slightly underestimates an SNR compared to the original one (See Fig. 10), it will not harm the GW detectability since this case lies in the higher SNR regime.

IV. CONCLUSION

We have applied and implemented an algorithm that optimizes latency for the whitening filter in the CBC data-analysis pipeline. Through statistical tests between the frequency-domain and zero-latency whitening filters, we have found that the two statistical values, SNR and $\chi^2$, are in sufficient agreement for both noise and injection triggers. As a result, we have achieved a 16 s latency reduction in the whitening process. It should be noted that this work has yielded the first confirmation that a zero-latency whitening filter can be employed in the data-analysis pipeline for CBC GW searches.

ACKNOWLEDGMENTS

The authors are grateful to the LIGO lab for an offer of the S5 strain data. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation. This research was supported by the National Science Foundation through PHY-1454389 and ACI-1642391. Funding for this project was provided by the Charles E. Kaufman Foundation of The Pittsburgh Foundation.

[1] B. P. Abbott et al. (LIGO Scientific Collaboration), Classical Quantum Gravity 32, 074001 (2015).
[2] F. Acernese et al. (Virgo Collaboration), Classical Quantum Gravity 32, 024001 (2015).
[3] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 116, 241103 (2016).
[4] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 118, 221101 (2017).
[5] B. P. Abbott et al., Astrophys. J. Lett. 851, L35 (2017).
[6] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 119, 141101 (2017).
[7] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 116, 061102 (2016).
[8] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 119, 161101 (2017).
[9] B. D. Metzger and E. Berger, Astrophys. J. 746, 48 (2012).
[10] B. Abbott et al., Astrophys. J. Lett. 848, L12 (2017).
[11] C. Messick et al., Phys. Rev. D 95, 042001 (2017).
[12] X. Li, Y.-M. Hu, Y.-Z. Fan, and D.-M. Wei, Astrophys. J. 827, 75 (2016).
[13] L. Rezzolla and P. Kumar, Astrophys. J. 802, 95 (2015).
[14] K. Cannon, R. Cariou, A. Chapman, M. Crispin-Ortuzar, N. Fotopoulos, M. Frei, C. Hanna, E. Kana, D. Keppel, L. Liao, S. Privitera, A. Searle, L. Singer, and A. Weinstein, Astrophys. J. 748, 136 (2012).
[15] LSC, https://wiki.ligo.org/DASWG/GstLAL.
[16] S. Hooper et al., Phys. Rev. D 86, 024012 (2012).
[17] K. Cannon, Phys. Rev. D 85, 081504(R) (2012).
[18] K. Cannon, C. Hanna, and J. Peoples, arXiv:1504.04632.
[19] N. Damera-Venkata, B. L. Evans, and S. R. McCaslin, IEEE Trans. Signal Process. 48, 1491 (2000).
[20] S. C. Pei and H. S. Lin, IEEE Trans. Circuits Syst. II: Express Briefs 53, 1113 (2006).
[21] S. C. Pei, H. S. Lin, and P. H. Wang, IEEE Trans. Circuits Syst. I: Regular Papers 57, 925 (2010).
[22] S. Kidambi and A. Antoniou, IEEE Trans. Circuits Syst. II: Express Briefs 64, 472 (2017).
[23] B. C. Garai, P. Das, and A. K. Mishra, Signal Processing 91, 1812 (2011).
[24] M. Abo-Zahhad and Q. Al-Zoubi, Digit. Signal Process. 16, 211 (2006).
[25] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Phys. Rev. D 85, 122006 (2012).
[26] LSC, https://wiki.ligo.org/DASWG/LALSuite.
[27] B. P. Abbott et al. (LIGO Scientific Collaboration), Rep. Prog. Phys. 72, 076901 (2009).