A novel method for detecting coalescing binaries in near realtime with Advanced LIGO and beyond

Submit to PRD or CQG?

Kipp Cannon[1], Adrian Chapman[2], Nickolas Fotopoulos[2], Chad Hanna[3], Drew Keppel[4,5], Antony C. Searle[2], Leo Singer[2], Alan J. Weinstein[2]

[1] Canadian Institute for Theoretical Astrophysics, Toronto, ON, Canada
[2] LIGO Laboratory - California Institute of Technology, Pasadena, CA, USA
[3] Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada
[4] Albert-Einstein-Institut, Max-Planck-Institut für Gravitationphysik, Hannover, Germany
[5] Leibniz Universität Hannover, Hannover, Germany

Abstract. Conventional matched filter bank methods for the detection of gravitational waves from the inspiral of compact binaries are computationally expensive, have hundreds of seconds of unavoidable intrinsic latency, and require arrays of largely redundant matched filters. Novel detection methods that are more computationally efficient and have lower latency will be required to realize the full potential of advanced of gravitational wave detectors that are currently under construction. In this paper, we describe a new detection method that exploits the properties of inspiral waveforms using multi-rate filtering, principal component analysis, and hierarchical detection. We provide receiver operating characteristics from a prototype search pipeline that is capable of low-latency or near-realtime detection with greatly reduced computational requirements in comparison with previously described methods.

PACS numbers: 95.55.Ym, 84.30.Vn, 95.75.Wx

1. Introduction

Low-mass inspiraling binaries are the most promising sources of gravitational radiation for Advanced LIGO. Observation demonstrates that they must exist in the local Universe in some abundance [1]. A detection would strongly constrain mass and spin [2, 3], and with sufficient signal strength, the neutron star equation of state [4, 5], general relativity in the strong-field regime [6], the graviton mass [7, 8], and the Lorentz Invariance principle [9]. In the final stages of inspiral, the neutron star is tidally disrupted, providing the matter required to produce bright electromagnetic counterparts [10]. Conveniently, the gravitational waveforms are understood and implemented to the level required by Advanced LIGO to detect [11], allowing the sensitive technique of matched filtering.

The parameter space of compact binary coalescence signals is large [12, 13] leading to large filter banks and significant computational cost. Following the series of detector upgrades that are now under way, LIGO and Virgo will attain their Advanced configurations, gaining a tenfold improvement in amplitude sensitivity over Initial configurations [14, 15] and a corresponding thousandfold increase in observable volume

in the local Universe. The computational requirements of conventional detection strategies will be greatly increased by this upgrade for two reasons. First, advanced detectors will have much improved low-frequency sensitivity, so chirp signals from coalescing binaries will remain in the detectors' bands for a significantly longer time. This will demand that matched filter based searches use longer templates, requiring more memory and more cycles to process. Second, background noise for advanced detectors will be more spectrally uniform, so it will be possible to resolve the intrinsic parameters of a source that determine the time-frequency structure of the gravitational wave signal with much greater accuracy. This will require that matched filter bank based searches employ more matched filters, again demanding more memory and more cycles.

Beyond raw throughput, the ability to detect signals in near realtime will become increasingly valuable as the gravitational-wave detection horizon pushes outward. Having an electromagnetic or neutrino counterpart to a gravitational-wave detection would not only increase the confidence in the detection but will also greatly increase the astrophysical information available from the event. Most models that predict simultaneous gravitational-wave and electromagnetic observations also predict that the peak amplitude of electromagetic radiation will occur soon after gravitational-wave emission [16]. Thus in order to maximize the chance of a successful electromagnetic followup the latency of gravitational-wave signal analysis must be made minimal.

dk: I don't think this is true for neutrino counterparts. Just as there is no need to point GW detectors, there is no need to point neutrino detectors so the association can easily be done offline.

More tantalizingly, these long waveforms enable us to accumulate a significant fraction of the signal-to-noise ratio before coalescence. With enough signal-to-noise ratio (and a lot of work!) it should be possible to reconfigure the detectors on-the-fly to more sensitively capture the later merger part of the signal, where neutron star equation of state information can be extracted. With similar caveats, it is certainly possible to send alerts to telescopes several seconds before prompt emission. There are no theoretical hurdles to negative latency, but careful attention must be paid to the algorithms and implementation of search pipelines to take advantage of a fortuitously loud signal.

dk: This is not true for all EM bands. Radio afterglows from short GRBs are not expected to be seen until 1 month after observation and then observable for several months.

In 2010, LIGO and Virgo completed S6/VSR3, a short period of joint data taking during which several all-sky detection pipelines operated in a low-latency configuration [CITATION NEEDED]. A few candidates of moderate significance were promptly sent for electromagnetic followup [CITATION NEEDED] to several telescopes including Swift, ROTSE, TAROT, and Zadko. They achieved latencies of BLAH, but required human vetting of each candidate before alerting the telescope partners. As confidence in the infrastructure increases, humans can be removed from the loop, demonstrating a best latency of BLAH.

More speculative and no reference. Thoughts? Move to discussion? Worth adding any plots?

This work will describe how to exploit degeneracy in the signal parameter space to answer more quickly and economically whether or not a gravitational wave is present. Specifically, we will exploit the time-frequency structure of chirp signals to downsample different parts of the waveform, use the singular value decomposition (SVD) factors to identify the redundancy of the bank and reduce the effective number of filters required to search the data, and use matched subspace filters built from SVD to identify when computation can be avoided. We note that others have applied the use of SVD to gravitational wave data analysis to analyze optimal gravitational-wave burst detection [18, 19] and coherent networks of detectors [20].

If we publish before the LUMIN papers, we have to redo this section. We can't make claims about those private communications with the telescopes. Else, cite the pre-S6 LUMIN paper [17].

The paper is organized as follows. First we provide an overview of the standard method for detecting compact binary coalescence signals and describe how it can be modified to accomodate low latency analysis. We then describe the pipeline we have

nvf: I use some rather different words here to summarize the techniques. I don't want to simply repeat words from later, but I don't want to jar a reader with changing jargon either. Feel free to edit.

constructed to implement these changes. To validate the approach we present results of simulations and finish with some concluding remarks.

2. Method

2.1. The standard approach: matched filtering

Searches for gravitational waves from compact binary coalescences typically employ matched filter banks [21]. Potential inspiral signals are continuously parameterized by time, amplitude, phase, and a set of intrinsic source parameters θ, which in this paper we shall take to consist of the two component masses of a binary, $\theta = (m_1, m_2)$. Let $h_+(\theta)$ and $h_\times(\theta)$ be, respectively, the '+' and '×' polarization gravitational wave signals that would arise from a fiducial face-on binary at some distance. Because, for inspiral signals, h_+ and $h_\times$ are nearly in quadrature, they are generally combined into a single complex-valued template $h = h_+ + ih_\times$.

The detection procedure for just one set of intrinsic source parameters θ starts by *whitening* the measurement data stream x. This involves finding a linear filter that renders the detector's noise IID and Gaussian. This filter is applied to the measurement, yielding the whitened data stream x^{W}. The same linear filter is applied to the template $h(\theta)$, yielding the whitened template $h^{\mathrm{W}}(\theta)$. The matched filter is the normalized cross-correlation of h^{W} and x^{W},

After introducing the ⋆ notation for cross correlation, must we define it for continuous and discrete time?

$$\rho(\theta) = \frac{h^{\mathrm{W}}(\theta) \star x^{\mathrm{W}}}{|h^{\mathrm{W}}(\theta)|}.$$

This is called the signal to noise ratio, or SNR. The detection statistic is the modulus of this, $|\rho(\theta)|$, which has a χ^2 distribution with 2 degrees of freedom in the absence of signal.

To construct a filter bank, matched filters are realized for discrete signal parameters θ_1, θ_2, $\cdots$, θ_N, such that any possible signal will have a maximum cross-correlation of at least 0.97 with at least one template. Such a template bank is said to have a 97% *minimum match*. This technique is designed so that an inspiral signal can be detected without any prior knowledge of its intrinsic parameters: at most 3% of the SNR is lost by a signal's parameters not exactly coinciding with a template's. A trigger is reported for the template parameters θ_i and time t for which $|\rho|$ is a maximum over some moving interval in θ and t.

Citation needed for template placement procedure?

2.1.1. Latency and overhead The matched filter bank can be implemented with finite impulse response (FIR) filters. FIR filters are perfectly suited for realtime detection, because they do not introduce any latency at all. However, FIR filters are very expensive: for N templates of M samples each, a FIR matched filter bank costs $\mathcal{O}(MN)$ operations per sample.

Much more commonly, the matched filter bank is implemented using FFT convolutions, costing only $\mathcal{O}(M \lg N)$ per sample but having a latency of at least M samples. For example, for a bank of 1 ks templates sampled at 4096 Hz, the FFT implementation requires about about 2.2×10^4 times fewer floating point operations per sample than the FIR implementation. However, the FFT implementation has a latency of at least 1 ks.

lg *is* $\log_2$. *Explain nomenclature? Switch to* $\log_2$*?*

This presents a dilemma: it seems that low latency detection using the FIR implementation is prohibitively expensive, whereas computationally cheap detection with the FFT method comes with minutes to hours of latency.

In the rest of the section, we will describe a detection strategy that makes use of some very general properties of inspiral template banks in order to evaluate a matched filter bank with far lower computational cost than the conventional FIR method and far less latency than the conventional FFT method.

2.2. Selectively reducing the sample rate of the data and template waveforms

Our first innovation is to split each template into disjoint intervals, or *time slices*. A matched filter is constructed for each time slice, whose outputs form an ensemble of partial SNR streams. By linearity, these partial SNR streams can be suitably time delayed and summed to reproduce exactly the SNR with respect to the original template. By exploiting some general properties of inspiral waveforms, we shall see that time slices permit a reduction in latency relative to the FFT matched filter, but with comparable computational overhead.

A similar idea has been demonstrated by the Virgo Collaboration's MBTA pipeline [22, 23, 24, 25], which operated in a low-latency mode during LIGO's sixth science run and Virgo's second science run in 2010.

MBTA should be cited in the introduction.

Each partial matched filter can be realized with either the FIR or the FFT method. If all of the time slices make use of the FFT method, the resulting filter will have lower latency than a conventional FFT matched filter, and also less computational overhead than an FIR matched filter. A further benefit is that all of the time slices may be processed in parallel for additional speedup.

If the template is known to be quasi-bandlimited, then the partial matched filter for each time slice may be processed at a reduced sample rate. Compact binary inspiral waveforms fit this description superbly: they are chirps whose frequency and amplitude rise according to power laws of time. They are not only quasi-bandlimited, but quasi-monochromatic. By exploiting this property, we can implement a matched filter for an inspiral waveform using a multirate filter bank that requires far fewer floating point operations.

For concreteness and simplicity, let us consider an inspiral waveform in the quadrupole approximation, for which the time-frequency relation is

$$f = \frac{1}{\pi \mathcal{M}} \left[\frac{5}{256} \frac{\mathcal{M}}{-t} \right]^{3/8} . \tag{1}$$

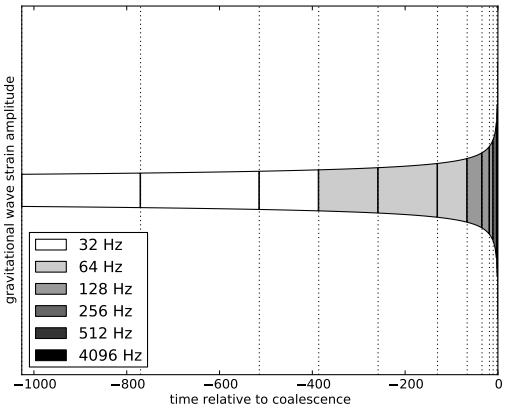
Here, $\mathcal{M}$ is the chirp mass of the binary in units of time (where $GM_\odot/c^3 \approx 5\mu$s) and t is the time relative to the coalescence of the binary [21, 26, 27, 28]. Usually the template is truncated at some prescribed time $t = t_0$, often chosen to correspond to the innermost stable circular orbit at which $f = f_{\mathrm{ISCO}} \approx 4400\, M_\odot/M$. An inspiral signal will enter the LIGO band at a low frequency cutoff, $f = f_{\mathrm{low}}$ corresponding to a time t_{low}. This template has a duration of $t_0 - t_{\mathrm{low}}$ and is critically sampled at a rate of $2f_{\mathrm{isco}}$.

Way too many citations here.

The time slices for this template consist of the K intervals $(t_K, t_{K-1}], \ldots, (t_2, t_1], (t_1, t_0]$ sampled at frequencies $f_{\mathrm{K}-1}, \ldots, f_1, f_0$ where $f_0 = 2f_{\mathrm{ISCO}}$, $t_0 = t_{\mathrm{ISCO}}$, $f_{K-1} \geqslant 2f_{\mathrm{low}}$, and $t_K \leqslant t_{\mathrm{low}}$.

If all of these time slices are implemented using the FFT, then the latency of this filter bank is $\max_j \left[-\left(2(t_j - t_{j-1}) - t_{j-1}\right)\right] = \max_j \left(3t_{j-1} - 2t_j\right)$.

Table 1: Example of critically sampled, power-of-2 time slices for a $1.4 - 1.4\,M_\odot$ template extending from $f_{\rm low} = 10\,{\rm Hz}$ to $f_{\rm ISCO} = 1571\,{\rm Hz}$ with a time frequency structure given by (1).



f_k (Hz)	$(t_{k+1},\, t_k]$ (s)	Samples
4096	$(-0.25,\, 0]$	1024
512	$(-2.25,\, -0.25]$	1024
256	$(-10.25,\, -2.25]$	2048
128	$(-18.25,\, -10.25]$	1024
128	$(-34.25,\, -18.25]$	2048
128	$(-66.25,\, -34.25]$	4096
64	$(-130.25,\, -66.25]$	4096
64	$(-258.25,\, -130.25]$	8192
64	$(-386.25,\, -258.25]$	8192
32	$(-514.25,\, -386.25]$	4096
32	$(-770.25,\, -514.25]$	8192
32	$(-1026.25,\, -770.25]$	8192

An example time slice design satisfying these constraints for a $1.4 - 1.4\,M_\odot$ is shown in table 1. For this example, the latency for this time slice design is just 125.75 s even if all of the time slices are implemented with the FFT method. This set of time slices will require 619 operations per sample per template, compared with 372 for pure FFT cross-correlation without time slices, or 8.4×10^6 for the FIR filter method without time slices.

2.3. Reducing the number of filters with the singular value decomposition

Our second innovation exploits the fact that the templates in inspiral template banks are, by design, highly correlated. This is due to the fact that points are chosen in parameter space to be templates such that any point from the region parameter space we cover has a match to the nearest template greater than our chosen minimum match. It is possible to greatly reduce the number of matched filters required to achieve a particular minimum match by designing an appropriate set of orthonormal *basis templates*. A purely numerical technique based on the application of the singular value decomposition (SVD) to inspiral waveforms is demonstrated by the authors in [29].

One may regard a bank of $N/2$ discretely-sampled, complex-valued templates of length M samples, $h^{\rm W}(\theta_i; t_j) = [\mathbf{H}]_{ij} = H_{ij}$, as a real matrix of shape $N \times M$, where we have unpacked each complex-valued template into two real-valued templates. The singular value decomposition is an exact factorization that exists for any matrix such that

$$H_{ij} = [\mathbf{V\Sigma U}]_{ij} = \sum_{k=1}^{N} v_{ik}\sigma_k u_{kj}, \tag{2}$$

where $\mathbf{V}$ and $\mathbf{U}$ are both unitary matrices and $\mathbf{\Sigma}$ is a diagonal matrix. For our purposes, we associate the rows of $\mathbf{U}$ with a minimal set of basis templates, which become the kernels of basis filters. These filters give rise to the orthogonal SNRs, $\rho_k^{\perp} = u_k \star x^{\rm W}$. The rows of $\mathbf{V\Sigma}$ become reconstruction coefficients that map linear

combinations of the orthogonal SNRs from the basis filters back onto SNRs for the original templates of interest: $\rho_i = \sum_k v_{ik}\sigma_k\rho_k^\perp$.

In many applications of the SVD, including ours, the matrix can be well approximated by truncating the summation in equation (2) at $N' \ll N$:

$$H'_{ij} = \sum_{k=1}^{N'} v_{ik}\sigma_k u_{kj} \tag{3}$$

The cumulative sum of squares of the *singular values*, σ_k, measures the Frobenius norm of the approximation, such that

$$\frac{\|\mathbf{H}'\|}{\|\mathbf{H}\|} = \left(\sum_{k=1}^{N'} |\sigma_k|^2\right)^{1/2} \left(\sum_{k=1}^{N} |\sigma_k|^2\right)^{-1/2}. \tag{4}$$

This is also called the SVD tolerance. In our application, it relates to how much SNR is lost by discarding $N - N'$ of the basis filters with the lowest singular values.

This result differs from other work that models gravitational-wave chirp signals in approximate ways [30, 31, 32] by starting with an exact representation of the desired template family and producing a rigorous approximation with a tunable accuracy.

2.4. Composite detection statistic and hierarchical detection

We should cite [33] here.

Although the SVD allows us to reduce the number of matched filters, this comes at the price of having to perform a matrix multiplication by the reconstruction matrix $\mathbf{V\Sigma}$ at every sample. In some circumstances, this matrix multiplication may be more expensive than applying the orthogonal matched filters.

Further speedup may be gained in a hierarchical detection scheme. In general, the orthogonal SNRs alone provide some indication of whether any template in the original template bank is likely to have a large SNR. Consider some composite detection statistic that is a scalar function of the orthogonal SNRs, $\Gamma(\rho_1^\perp, \rho_2^\perp, \cdots, \rho_L^\perp)$. Suppose that we can infer the distribution of Γ for a data stream with the signal present or with the signal absent. Then we may use a threshold crossing of Γ to trigger the conditional application of the expensive reconstruction matrix only during the times when a signal is likely to be present.

Following the ideas in [34], we employ is a weighted sum of squares, $\Gamma = \sum_{k=1}^{N'} w_k(\rho_k^\perp)^2$, with the particular weights

$$w_k = \frac{\sigma_k^2}{\sigma_k^2 + N/A^2}, \tag{5}$$

where A^2 is a desired SNR scale that is set by the analyst. As the authors show in [35], this choice of weights has a better receiver operating characteristic for a signal of SNR $= A$ than some other obvious choices, $w_k = \sigma_k^2$ or $w_k = 1$.

2.5. Comparison of computational costs and latency

Using all of these innovations, we are able to greatly reduce the compuational cost of searching for these signals. An example of this can be seen in table 2 where we compute the computational cost of filtering 256 templates from a 97% minimal match template bank whitened with an (FIXME: is this correct?) Advanced LIGO ASD. It is interesting to note that by themselves, both the SVD and time-slice innovations

increase the overall computational cost when combined with FFT filtering. It is only when combined with conditional reconstruction do these methods yield computational savings. Details of the computational cost of different operations can be found in Appendix A.

It would be good to illustrate the layout of this particular template bank: masses spanned, time slice layout ...

Table 2: Operation counts per sample for six different detection methods. The operation counts for LLOID assume a reconstruction duty cycle of 5%. Note that the FFT method with LLOID is almost 10 times faster than the conventional FFT method, despite having substantially lower latency.

operations/sample	latency (s.)	method
3 714 580 480	2.4×10^{-4}	conventional FIR method
49 937	1.8×10^{3}	conventional FFT method
1 233 400 832	2.4×10^{-4}	FIR method with SVD
60 740	1.8×10^{3}	FFT method with SVD
1 233 359 245	2.4×10^{-4}	FIR method with LLOID and no time slices
19 153	1.8×10^{3}	FFT method with LLOID and no time slices
1 587 712	2.4×10^{-4}	FIR method with time slices
173 926	2.5×10^{-1}	FFT method with time slices
120 673	2.4×10^{-4}	FIR method with time slices and SVD
74 498	2.5×10^{-1}	FFT method with time slices and SVD
51 285	2.4×10^{-4}	FIR method with LLOID
5 110	2.5×10^{-1}	FFT method with LLOID
2 146	2.5×10^{-1}	same, with cascade topology
1 612	2.5×10^{-1}	same, with cascade topology 2

3. Software Framework

The detection method described above is much more efficient in terms of floating point operations than the traditional matched filter bank method. However, time slices and conditional reconstruction greatly complicate queueing, synchronizing, and bookkeeping of intermediate signals. A low latency implementation capable of recruiting more than one CPU core would be difficult to achieve within the familiar serial programming framework because of the nontrivial time-delay relationships between samples. Due to these complications, we chose to prototype the search using an open source signal processing environment called GStreamer [36]. Primarily used for playing, authoring, or streaming media on Linux systems, GStreamer is an integral component of the popular Gnome desktop.

4. Results

We generated receiver operating characteristic (ROC) curves for a selection of choices for the composite detection statistic threshold and the SVD tolerance. For each ROC curve, or 'run', we analyzed mock data with injections to assess detection efficiencies and mock data without injections to evaluate false alarm rates.

4.1. Detector noise characteristics

We tested the new detection method with mock Advanced LIGO data having a power spectrum prescribed by the "zero detuning, high power" noise model in [37]. Colored Gaussian noise is generated by passing 5 independent realizations of white Gaussian noise sampled at 16384 Hz through a bank of 5 third-order IIR filters, then summing the filters' outputs. This carefully designed filter bank reproduces the noise model very faithfully, but since it is composed of IIR filters it can produce mock data in realtime very cheaply. See Appendix B for implementation details.

In this section, we refer to aLIGO run parameters, but the results are actually using initial LIGO noise and initial LIGO parameters. This will be fixed.

4.2. Injection population

For each run, we performed about 10^4 injections into our mock dataset. Injections were 2PN post-Newtonian inspiral waveforms with component masses independently and identically distributed in $[1, 3]$ $M_\odot$. Injections were tapered at a frequency of 10 Hz where they enter the Advanced LIGO detection band. This dictated that the longest injection waveform had a duration of 1 777 s. In order to guarantee that no two injection waveforms overlapped in the data, the coalescence times of injections were spaced apart by 2 000 ± 100 s. For the sake of economizing disk space, we simultaneously analyzed 16 statistically independent sets of about 10^3 injections apiece in the same 14 day stretch of mock strain data.

Check –taper-injection option with Drew.

Injections were distributed uniformly in log distance. We used the following heuristic to provide a reasonably balanced number of distant, marginally detectable signals, and nearby, easily detected signals:

$$d_{\min} = \frac{1}{2} \cdot \frac{\rho_{\text{candle}}}{8.0 \cdot \min d_{\text{eff}}}, \quad d_{\max} = \frac{1}{4} \cdot \frac{\rho_{\text{candle}}}{5.5 \cdot \max d_{\text{eff}}}.$$

Here, ρ_{candle} is a fiducially selected SNR, and d_{eff} is the distance at which a face-on-system would produce a matched filter SNR of ρ_{candle}.

Orientations of injection sources are drawn uniformly, such that the sky location is distributed uniformly in 4π and the cosine of the binary inclination angle ι is distributed uniformly in $[0, \pi]$.

4.3. Noninjection population

We completed one injection-free pass over the 14 day dataset.

*And **why**? This part is too terse.*

4.4. Triggering and data reduction

For each template with parameters θ_i, a trigger is recorded any time the absolute value of the SNR, $|\rho(\theta_i)|$, is greater than a threshold $\rho^* = 5.5$, and is also a maximum over the 10 ms before and after. A trigger is therefore an ordered triple (t, i, ρ), recording the time, template number, and complex SNR of an excursion in the outputs of the matched filter bank.

This is a proposed trigger criterion that is much simpler than what we currently use.

Offline, triggers for all templates collectively are clustered by time, such that within a sliding window of 4 seconds, all triggers except the one with the highest ρ is kept.

For the injection-free passes, all clustered triggers are regarded as false alarms. We accumulate the SNRs of all of the false alarms into the set $\{|\rho|\}_{\text{false alarm}}$.

For each injection, we check whether there was any clustered trigger that was within 2 seconds of the injection's time. If there was, then the injection is considered to be recovered at the SNR of that trigger. Otherwise, the injection was not recovered because the SNR threshold ρ^* was not met, so we consider the injection to be recovered only at an SNR of ∞. From all found injections, we accumulate the injected distance and the recovered SNR into the set $\{(d, |\rho|)\}_{\text{found}}$.

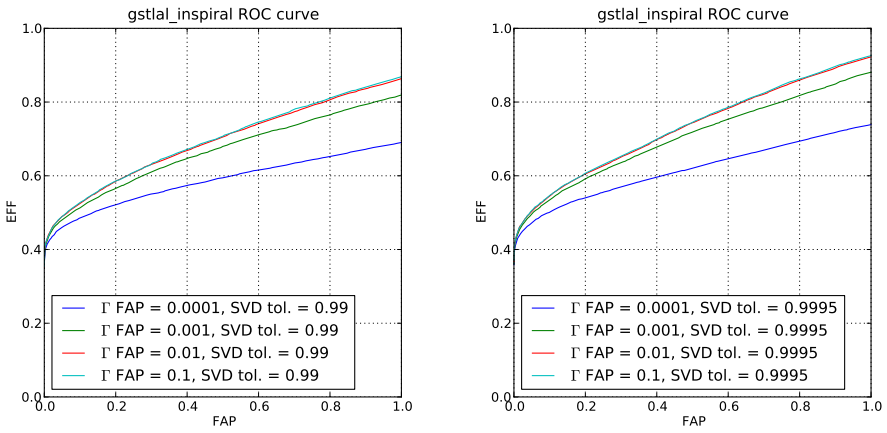


Figure 1: Receiver operating characteristic (ROC) curve of detection efficiency (EFF) versus false alarm probability (FAP).

5. Conclusions

Acknowledgments

LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement PHY-0107417. This paper has LIGO Document Number LIGO-P0900004-v2.

Appendix A. Floating point operation counts

Each addition and each multiplication is counted as a single floating point operation. We are not assuming that a multiply-accumulate is available as a single operation.

The filter bank can be implemented using finite impulse response (FIR) filters, which are just sliding window dot products. If there are M templates of length n, and the data stream contains N samples, then applying the filter bank requires $2MNn$ operations.

More commonly, the matched filters are implemented using the FFT convolution. This entails applying FFTs to blocks of D samples, with $2n \leq D$, each block overlapping the previous one by n samples. There are $N/(D-n)$ such blocks. Modern implementations of the Cooley-Tukey FFT, such as the ubiquitous `fftw`, require about $4N \lg N$ operations to evaluate a DFT of size N [38]. A D sample cross-correlation

This is more commonly known as "overlap-save". We should find someone else's operation count and cite it.

Table A1: Number of floating point operations per sample (multiplications and divisions) required for a selection of signal processing operations used in LLOID.

Process	ops/sample
FIR matched filter, M templates of length n	$2Mn$
FFT matched filter, M templates of length n, blocks of length D	$\frac{4(M+1)\lg D+2M}{1-n/D}$
n tap FIR resampling filter, sample rates $f_1 < f_2$	$2Mnf_1/f_2$
multiply $M \times L$ real matrix by $L \times 1$ real vector	$2ML$

consists of a forward FFT, an D sample dot product, and an inverse FFT, totaling $8D \lg N + 2D$ operations per block. Per sample, this is $(8 \lg D + 2)/(1 - n/D)$ operations.

The FIR filter implementation has the advantage that it has no intrinsic latency, whereas the FFT convolution has at least the latency of the FFT block size $D - n \geq n$. For example, for a $1.4 - 1.4\,M_\odot$ template with duration $\sim 1\,\mathrm{ks}$, the FFT convolution has a latency $\geq 2\,\mathrm{ks}$. However, the FIR filter implementation has the disadvantage of much greater overhead per sample than the FFT convolution. For a 1 ks template sampled at 4096 Hz, the FIR implementation requires about about $n/8 \lg 2n = 2.2 \times 10^4$ times more operations per sample than the FFT implementation.

Drew: Why don't we change this to an overlap of m samples so we can see what happens as we increase the overlap to reduce latency.

Appendix B. Filter bank for generating mock Advanced LIGO strain data

The filter bank described below reproduces the "zero detuning, high power" Advanced LIGO noise model of [37] very faithfully. Since it is composed of a small number of third order or lower linear filters, a digital implementation of it can produce mock data in realtime with negligibly few floating point operations.

First, generate 5 independent streams of white Gaussian noise, $x_1, \ldots, x_5$, sampled at 16384 Hz. Next, the apply the (F/I)IR filters described in equation (B.1) to generate $y_1, \ldots, y_5$ from $x_1, \ldots, x_5$ respectively. Finally, sum and scale all of the $y_1, \ldots, y_5$ together to obtain the final output y.

The power spectrum of y is the sum of squares of the magnitudes of the transfer functions of all of the filters. The output of the filter bank is compared with the noise model in Figure B1 below.

The mock Advanced LIGO noise generator is implemented by the GStreamer element `lal_fakeadvligosrc`, which is included with the analysis code.

dk: We should add the seismic wall to this PSD. I know this is waiting on switching to gsteamer 0.10.32 but people will worry about this otherwise.

$$x_i[n] \sim \mathcal{N}[0,1] \quad \forall\, i, n$$

$$y_1[n] = \left(4 \times 10^{-28}\right) x_1[n] - y_1[n-1] + 2 \cdot 0.99995 \cos\left(\frac{2\pi \cdot 9.103}{16384}\right) y_1[n-2] - 0.99995^2 y_1[n-3]$$

$$y_2[n] = \left(1.1 \times 10^{-23}\right) (x_2[n] - x_2[n-1])$$

$$y_3[n] = \left(10^{-27}\right) x_3[n] - y_3[n-1] + 2 \cdot 0.999 y_3[n-2] - 0.999^2 y_3[n-3]$$

$$y_4[n] = \left(4 \times 10^{-26}\right) x_4[n] - y_4[n-1] + 2 \cdot 0.87 \cos\left(\frac{2\pi \cdot 50}{16384}\right) y_4[n-2] - 0.87^2 y_4[n-3]$$

$$y_5[n] = \left(6.5 \times 10^{-24}\right) x_5[n] - y_5[n-1] - 2 \cdot 0.45 y_5[n-2] - 0.45^2 y_5[n-3]$$

$$y\,[n] = \frac{3\sqrt{16384}}{4} \left(y_1[n] + y_2[n] + y_3[n] + y_4[n] + y_5[n]\right) \tag{B.1}$$

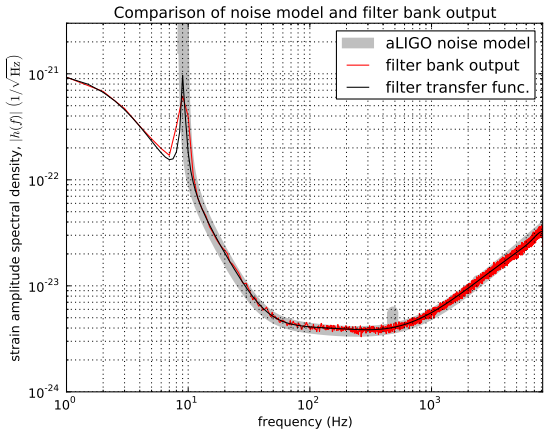


Figure B1: Power spectrum of 100 s of output from mock data filter bank compared with the "zero detuning, high power" Advanced LIGO noise model.

Note that the CPU overhead of this procedure will be entirely dominated by drawing the pseudorandom numbers $x_1, \ldots, x_5$.

References

[1] Abadie J *et al.* 2010 *Class. Quant. Grav.* URL `http://iopscience.iop.org/0264-9381/27/17/173001`

[2] Finn L and Chernoff D 1993 *Phys. Rev. D* **47** 2198–2219

[3] Poisson E and Will C M 1995 *Phys. Rev.* **D52** 848–855 (*Preprint* `arXiv:gr-qc/9502040`)

[4] Flanagan Éanna É and Hinderer T 2008 *Phys. Rev.* **D77** 021502 (pages 5) URL `http://link.aps.org/abstract/PRD/v77/e021502`

[5] Read J S, Markakis C, Shibata M, Uryū K, Creighton J D E and Friedman J L 2009 *Phys. Rev.* **D79** 124033–+ (*Preprint* `arXiv:0901.3258`)

[6] Will C M 2005 *Living Rev. Rel.* **9** 3 (*Preprint* `gr-qc/0510072`)

[7] Stavridis A and Will C M 2009 *Phys. Rev.* **D80** 044002

[8] Keppel D and Ajith P 2010 (*Preprint* `1004.0284`)

[9] Ellis J, Mavromatos N, Nanopoulos D, Sakharov A and Sarkisyan E 2006 *Astroparticle Physics* **25** 402–411 ISSN 0927-6505 URL `http://www.sciencedirect.com/science/article/B6TJ1-4JXY1RR-1/2/84c3810841cfe896d01a49e5b006ce61`

[10] Shibata M and Taniguchi K 2008 *Phys. Rev.* **D77** 084015–+ (*Preprint* `arXiv:0711.1410`)

[11] Buonanno A, Iyer B R, Ochsner E, Pan Y and Sathyaprakash B S 2009 *Phys. Rev. D* **80** 084043

[12] Owen B J 1996 *Phys. Rev. D* **53** 6749–6761

[13] Owen B J and Sathyaprakash B S 1999 *Phys. Rev. D* **60** 022002

[14] Abbott B *et al.* 2009 Advanced LIGO reference design Tech. Rep. LIGO-M060056-v1 LIGO Project URL `https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=1507`

[15] Acernese F *et al.* 2008 Advanced virgo preliminary design Tech. Rep. VIR-089A-08 The Virgo Collaboration URL `https://tds.ego-gw.it/ql/?c=2110`

[16] Sylvestre J 2003 *Astrophys. J.* **591** 1152–1156 URL `arXiv:astro-ph/0303512v1`

[17] Kanner J, Huard T L, Marka S, Murphy D C, Piscionere J, Reed M and Shawhan P 2008 URL `arXiv:0803.0312v4[astro-ph]`

[18] Brady P R and Ray-Majumder S 2004 *Classical and Quantum Gravity* **21** S1839–S1848 URL arXiv:gr-qc/0405036
[19] Heng I S 2008 URL arXiv:0810.5707v1[gr-qc]
[20] Wen L 2008 *Int. J. Mod. Phys. D* **17** 1095–1104 URL http://arxiv.org/abs/gr-qc/0702096v2
[21] Allen B A, Anderson W G, Brady P R, Brown D A and Creighton J D E 2005 (*Preprint* gr-qc/0509116)
[22] Marion F and the Virgo Collaboration 2004 *Proc. Rencontres de Moriond on Gravitational Waves and Experimental Gravity 2003*
[23] FBeauville, M-ABizouard, LBlackburn, LBosi, PBrady, LBrocco, DBrown, DBuskulic, FCavalier, SChatterji, NChristensen, A-CClapson, SFairhurst, DGrosjean, GGuidi, PHello, EKatsavounidis, MKnight, ALazzarini, NLeroy, FMarion, BMours, FRicci, AVicere and MZanolin 2006 *J. Phys. Conf. Ser.* **32** 212 URL arXiv:gr-qc/0509041v1
[24] Beauville F, Bizouard M A, Blackburn L, Bosi L, Brocco L, Brown D, Buskulic D, Cavalier F, Chatterji S, Christensen N, Clapson A C, Fairhurst S, Grosjean D, Guidi G, Hello P, Heng S, Hewitson M, Katsavounidis E, Klimenko S, Knight M, Lazzarini A, Leroy N, Marion F, Markowitz J, Melachrinos C, Mours B, Ricci F, Vicer A, Yakushin I and Zanolin M 2008 *Class. Quant. Grav.* **25** 045001 (*Preprint* http://www.iop.org/EJ/abstract/0264-9381/25/4/045001/) URL arXiv:gr-qc/0701027v1
[25] Buskulic D, the LIGO Scientific Collaboration and the Virgo Collaboration 2010 *Classical and Quantum Gravity* **27** 194013 URL http://stacks.iop.org/0264-9381/27/i=19/a=194013
[26] E K L, M W C and G W A 1992 *Class. Quantum Grav.* **9** L125–31
[27] Blanchet L 2002 *Phys. Rev. D* **65**
[28] Hanna C, Megevand M, Ochsner E and Palenzuela C 2009 *Class. Quantum Grav.* **26** 015009
[29] Cannon K, Chapman A, Hanna C, Keppel D, Searle A C and Weinstein A J 2010 *Physical Review D* **82** 44025 (c) : URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010PhRvD..82d4025C&link_type=ABSTRACT
[30] Chassande-Mottin E and Pai A 2006 *Phys. Rev. D* **73** 042003 URL http://arxiv.org/abs/gr-qc/0512137
[31] Cands E J, Charlton P R and Helgason H 2008 *Class. Quant. Grav* **25** 184020 URL http://arxiv.org/abs/0806.4417
[32] Buonanno A, Chen Y and Vallisneri M 2003 *Phys. Rev. D* **67** 024016 erratum-ibid. 74 (2006) 029903(E)
[33] Scharf 1994 *Signal Processing, IEEE Transactions on* **42** 2146 – 2157 URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=301849&queryText%253D%2528%2528Matched+subspace+detectors%2529+AND+%2528scharf%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All
[34] Anderson W G, Brady P R, Creighton J D E and Flanagan E E 2001 *Phys. Rev.* **D63** 042003 (*Preprint* gr-qc/0008066)
[35] Cannon K, Hanna C, Keppel D and Searle A C 2011 *ArXiv e-prints* (*Preprint* 1101.0584)
[36] GStreamer: open source multimedia framework URL http://gstreamer.freedesktop.org
[37] Shoemaker D 2009 Advanced ligo anticipated sensitivity curves Tech. rep. URL https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=2974
[38] Johnson S and Frigo M 2007 *Signal Processing, IEEE Transactions on* **55** 111 – 119 URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4034175&queryText%253D%2528%2528Document+Title%253Aa+modified+split-radix+fft+with+fewer%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All