Toward early-warning detection of compact binary coalescence

Kipp Cannon, Romain Cariou, Adrian Chapman, Mireia Crispin-Ortuzar, Nickolas Fotopoulos, Melissa Frei, Chad Hanna, Erin Kara, Drew Keppel, Laura Liao, Stephen Privitera, Antony Searle, Leo Singer, and Alan Weinstein

Presentation by Leo Singer

LIGO Laboratory, California Institute of Technology

August 2, 2011

Outline

1. Motivation
2. Prospects for early-warning detection
3. Method
4. Implementation
5. Conclusion

Motivation

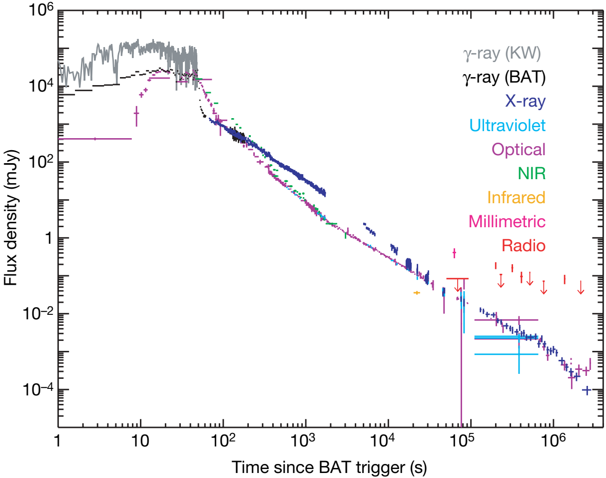
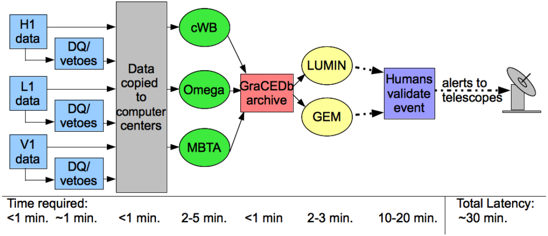
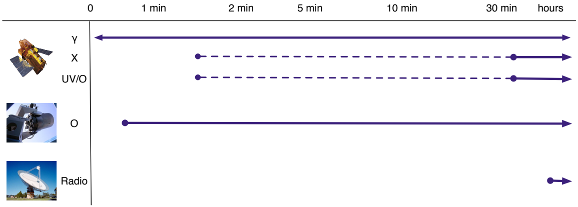


Image from Racusin et al., "Broadband observations of the naked-eye γ-ray burst GRB 080319B," Nature 455 183–188 (2008), `http://www.nature.com/nature/journal/v455/n7210/pdf/nature07270.pdf`.

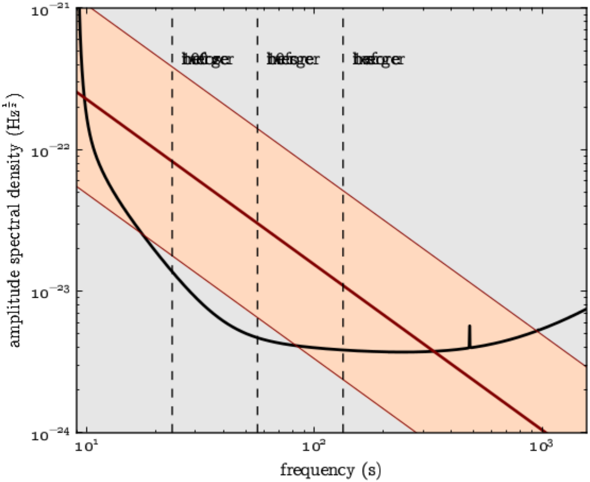
Latency associated with EM and GW observation





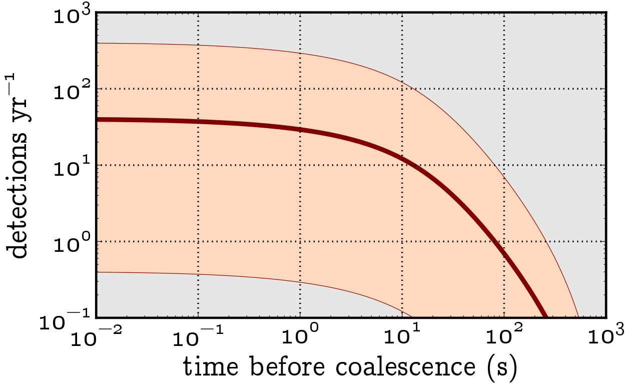
Upper image from Brennan Hughey, "Triggers in the recent LIGO an Virgo science runs," GWPAW 2011, http://www.gravity.phys.uwm.edu/conferences/gwpaw/talks/hughey.pdf.

Detectability before merger



In advanced LIGO, inspiral signals are in principal detectable tens or hundreds of seconds before the GW from the merger have reached the earth.

Prospects for early-warning detection

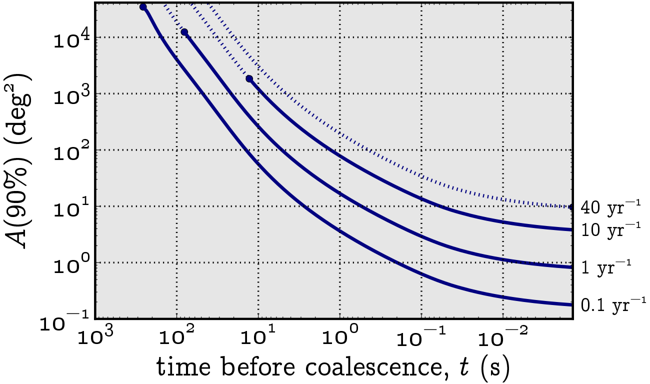


Although rates are highly uncertain due to our small sample of confirmed binary neutron stars, of a predicted 40 detections yr^{-1},

- 10 yr^{-1} will be detectable 10 s before merger, and
- 1 yr^{-1} will be detectable 1 s before merger.

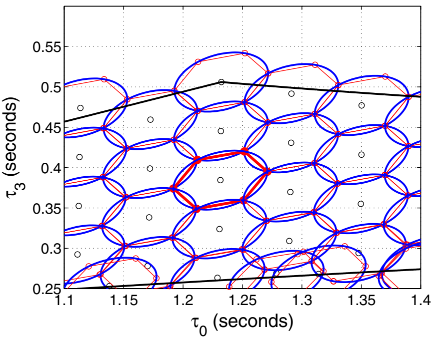
Sky localization accuracy

Sky localization accuracy improves rapidly as the signal progresses toward merger. Although a near real-time GW search pipeline might make it possible to point telescopes minutes after merger, one might hope to image counterparts to a few exceptional events before merger.



Conventional inspiral searches: matched filter banks

General relativity predicts the GW signal due to the inspiral of a system with known intrinsic source parameters (mass, eccentricity, spin).



To detect any signal that nature may provide, we can build banks of filters each of which has optimal signal to noise for a given source.

These matched filters tile the parameter space discretely, for example in a hexagonal grid.

Image from Cokelar, T, Phys. Rev. D 76, 102004 (2007).

Matched filter banks

For systems in which the effects of can be ignored, the intrinsic source parameters are just the component masses of the binary, $\theta = (m_1, m_2)$.

Strain observed by the detector is a linear combination of two orthogonal signals corresponding to the '$+$' and '$\times$' polarizations.

M templates are chosen for $M/2$ sources $\theta_0, \theta_1, \ldots, \theta_{M/2-1}$.

For $i \in [0, M)$, unit normalized template $h_i[k]$, whitened detector data $x[k]$, the filter outputs are just cross-correlations:

$$\rho_i[k] = \sum_{n=0}^{N-1} h_i[n] x[k-n].$$

Time domain method: FIR filter

The most straightforward way to build a matched filter bank is using FIR filters, which are just sliding dot products.

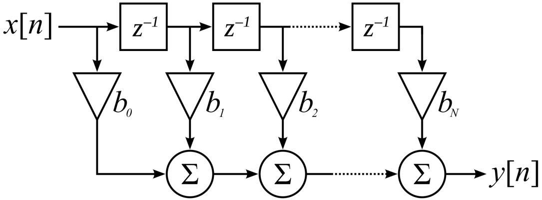


Image courtesy of Jonathan Blanchard, "FIR Filter Canonical Realization," Wikipedia, February 23, 2008, `http://commons.wikimedia.org/wiki/File:FIR_Filter.svg`.

Pros:

- Easy to implement
- Zero latency

Cons:

- Expensive if templates contain many samples

Frequency domain method: overlap-save

An alternative to the time domain method is frequency domain convolution via the FFT.

Pros:

- Computationally efficient even for very long templates
- Highly tuned FFTs available for most architectures

Cons:

- Input must be zero-padded, output must be clipped
- High latency: typically comparable to length of templates

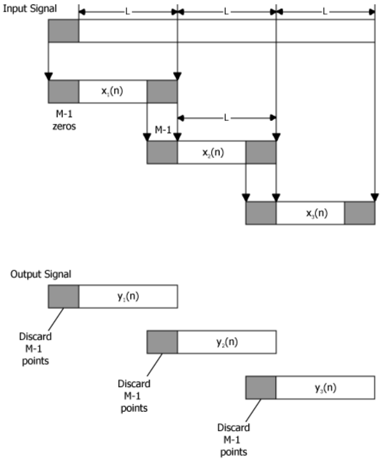
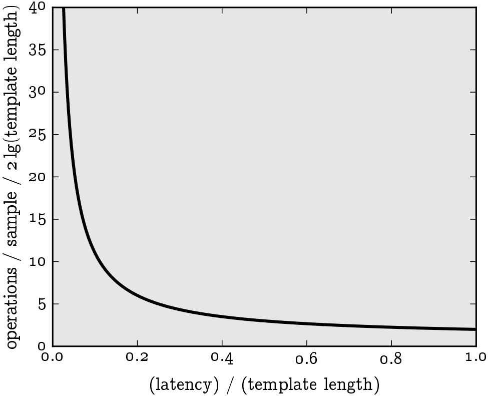


Image courtesy of Douglas Jones, "Fast Convolution," Connexions, June 21, 2004, http://cnx.org/content/m12022/1.5/.

Frequency domain method: overlap-save

The frequency domain (FD) method's latency is determined by the overlap between blocks.

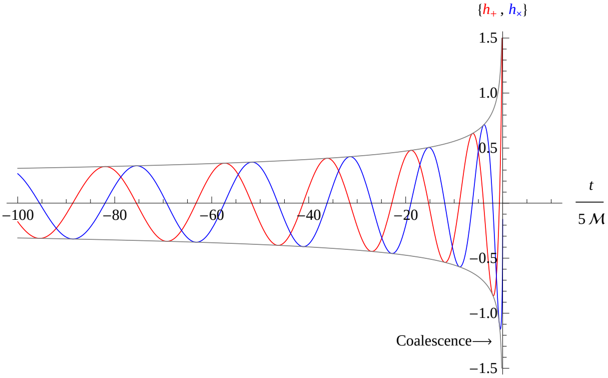


The latency can be made arbitrary small, but at the price of a *divergent* computational cost of

$$\approx \left(1 + \frac{\text{filter length}}{\text{latency}}\right)$$

operations / sample / $2\lg(\text{template length})$.

Novel methods can exploit properties of CBC signals



- Inspiral signals are chirps: "slowly" evolving in frequency
- Templates in inspiral filter banks are by design highly correlated

Novel method: LLOID

Low Latency Online Inspiral Detection

We exploit the chirp-like nature of inspiral signals by

- Partitioning and downsampling the template coefficients
  reducing number of filter coefficients by a factor of $\sim 10^2$
- Decimating the detector data in several stages
  reducing sample rate by a factor of $\sim 10^2$
- Decomposing templates further using the singular value decomposition (SVD)
  reducing the number of filters by a factor of $\sim 10^1$—10^2

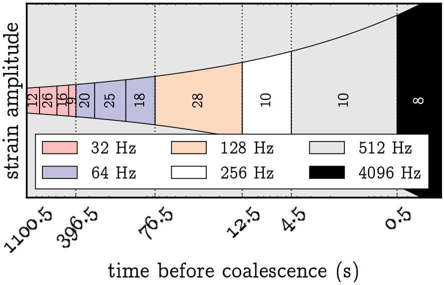
resulting in an overall speedup by a factor of $\sim 10^5$—10^6 over the conventional TD method.

First trick: time slices

Inspiral signals are chirps $\Rightarrow$ truncating the waveform at some time t before merger results in a bandlimited signal.

Using known time-frequency relationship, e.g. $f(t) = \frac{1}{\pi \mathcal{M}} \left[\frac{5}{256} \frac{\mathcal{M}}{t} \right]^{3/8}$,

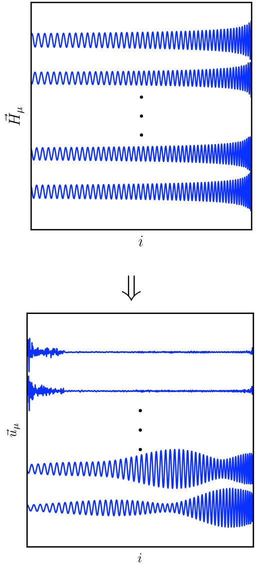
split templates into orthogonal "time slices": $h_i[k] = \sum_{s=0}^{S-1} \begin{cases} h_i^s[k] & \text{if } t^s \leqslant \frac{k}{f^0} < t^{s+1} \\ 0 & \text{otherwise.} \end{cases}$



Can downsample time slices w/o aliasing:

$$h_i^s[k] \equiv \begin{cases} h_i\left[k \frac{f}{f^s}\right] & \text{if } t^s \leqslant k/f^s < t^{s+1} \\ 0 & \text{otherwise.} \end{cases}$$

Second trick: singular value decomposition



Time slices represent orthogonal subspaces, but within one time slice the templates are still highly correlated.

We decompose the time-sliced templates further using the singular value decomposition,

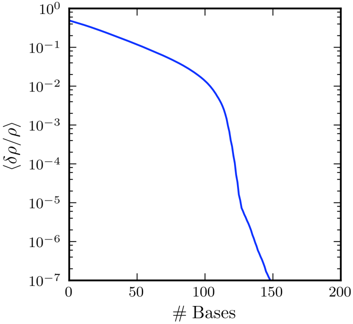
$$h_i^s[k] = \sum_{l=0}^{M-1} v_{il}^s \sigma_l^s u_l^s[k] \approx \sum_{l=0}^{L^s-1} v_{il}^s \sigma_l^s u_l^s[k].$$

This gives us orthonormal *basis templates* $u_l^s[k]$, related to the original templates through a *reconstruction matrix* $v_{il}^s \sigma_l^s$.

Images from Phys. Rev. D 82, 044025 (2010).

Second trick: singular value decomposition

The SVD is an exact matrix factorization, but we can greatly reduce the number of filters by keeping only $L \ll M$ basis templates.

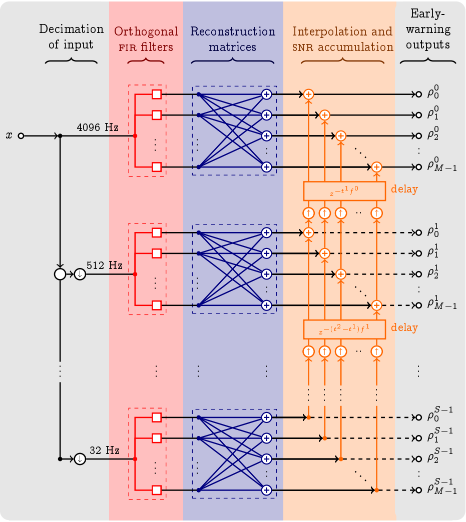


L determines the SVD tolerance,

$$\left[\sum_{l=0}^{L^s-1} \left(\sigma_l^s\right)^2\right] \left[\sum_{l=0}^{M-1} \left(\sigma_l^s\right)^2\right]^{-1},$$

which equals the expectation value of the fractional loss in SNR.

Image from Phys. Rev. D 82, 044025 (2010).

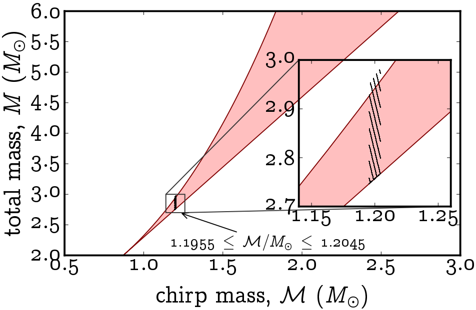


$$\underbrace{\rho_i^s\,[k]}_{\text{early-warning output}} = \overbrace{\left(H^{\uparrow}\rho_i^{s+1}\right)[k]}^{\text{SNR from previous time slices}} + \underbrace{\sum_{l=0}^{L^s-1} v_{il}^s \sigma_l^s}_{\text{reconstruction}} \overbrace{\sum_{n=0}^{N^s-1} u_l^s[n]}^{\text{orthogonal FIR filters}} \underbrace{x^s[k-n]}_{\text{decimated } h(t)}$$

Implementation

Offline planning stage

1. Place templates in parameter space (`lalapps_tmpltbank`)
2. Partition bank into *sub-banks* of comparable $\mathcal{M}$, T-F evolution
3. Divide each sub-bank into bandlimited time slices
4. Perform SVD on each time slice
5. Save everything to disk:
   - basis filter coefficients
   - reconstruction matrices
   - time slice boundaries



Implementation

Online filtering stage

We built a prototype LLOID filtering stage using *GStreamer*, an open source signal processing framework.

A GStreamer element consists of one or more *pipelines* of interconnected, reusable signal processing *elements*. Elements exchange both *events* and *buffers* of binary data.

GStreamer:

- is ubiquitous (present on most Linux desktops, many mobile devices)
- comes with C, Python, and shell interfaces
- is implicitly multithreaded
- provides useful stock signal processing elements

GStreamer pipeline

Computational costs and latency

Table : Cost and latency of the TD method, the FD method, and LLOID.

method	flop/s	latency (s)
time domain	4.9×10^{13}	0
frequency domain	5.2×10^{8}	2×10^{3}
LLOID (theory)	6.6×10^{8}	0
LLOID (prototype)	——	0.5

Table : Expressions for floating point operations per second (flop/s).

method	flop/s
time domain	$2f^0 MN$
frequency domain	$\approx 2f^0 M \lg D$
LLOID	$\approx 2\sum_{s=0}^{S-1} f^s L^s \left(N^s + M\right)$

Conclusion

- We have demonstrated a computationally feasible filtering algorithm for the rapid or possibly even early-warning detection of GWs emitted during the coalescence of neutron stars and stellar-mass black holes.
- The LLOID algorithm employs downsampled time-slices and rank-reduced basis templates given by the SVD.
- We have shown a prototype implementation of the LLOID algorithm using GStreamer, an open source signal processing platform.

Acknowledgements

LIGO was constructed by Caltech and MIT with funding from the NSF and operates under cooperative agreement PHY-0107417. LS is supported by the NSF through a Graduate Research Fellowship.