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 LS and NF

Based on arXiv:1107.2665

LIGO Laboratory, California Institute of Technology

December 12, 2011

Motivation

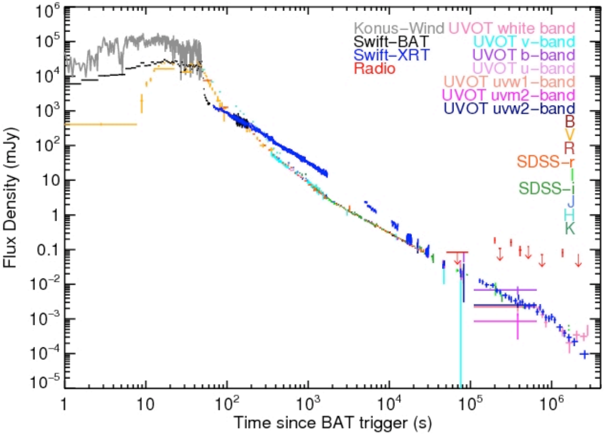
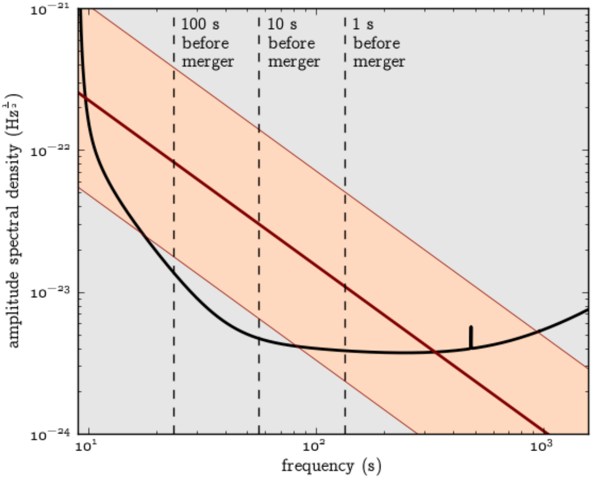


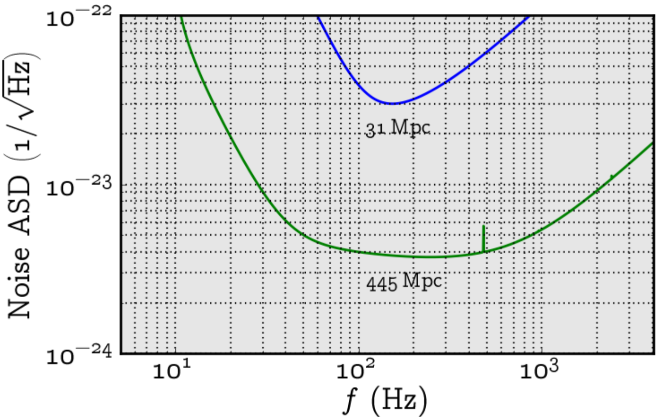
Image from Racusin et al., "Broadband observations of the naked-eye γ-ray burst GRB 080319B," Nature 455 183–188 (2008), `http://www.nature.com/nature/journal/v455/n7210/pdf/nature07270.pdf`.

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.

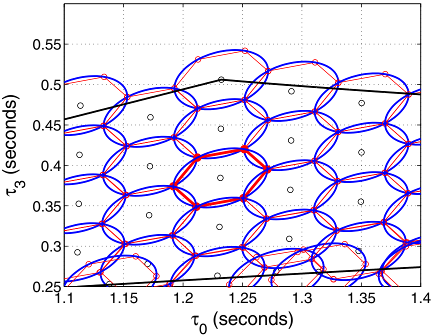
Advanced LIGO challenges



- 4× lower low-frequency cutoff ⇒ 40× longer waveforms
- 2× wider bandwidth ⇒ 10× more search templates

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).

Time domain method: FIR filter

The most straightforward way to build a matched filter bank is using FIR filters. FIR filters can be understood as sliding dot products or cross-correlation with one function time-reversed.

$$y[k] = \sum_{n=0}^{N-1} b[n]x[k-n]$$

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 CPU architectures

Cons:

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

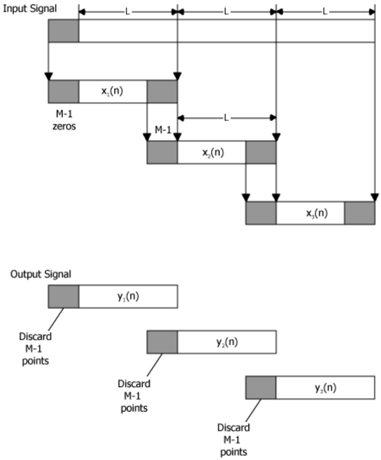
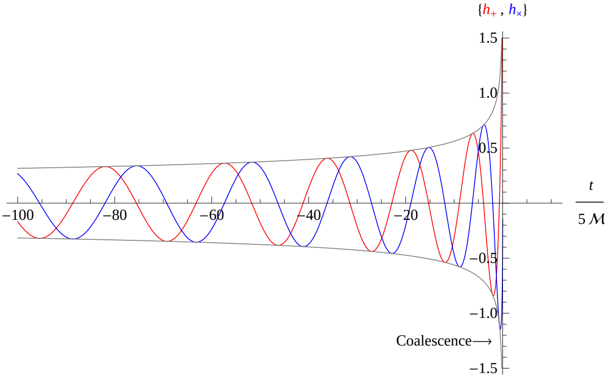


Image courtesy of Douglas Jones, "Fast Convolution," Connexions, June 21, 2004, http://cnx.org/content/m12022/1.5/.

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

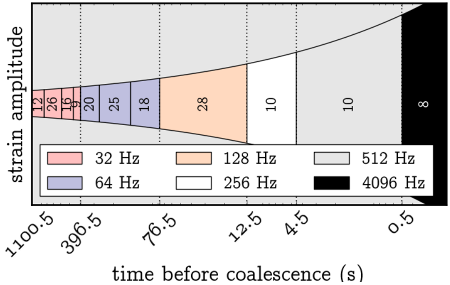
resulting in an overall speedup by a factor of $\sim 10^5$–10^6 over the conventional TD method.

First trick: time slices

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

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

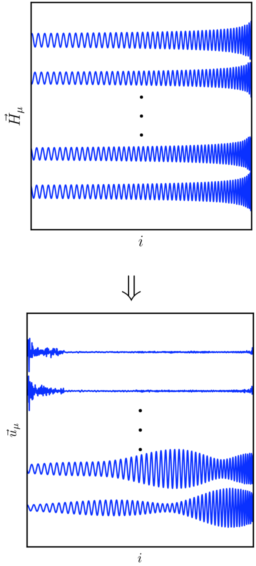
split templates into orthogonal "time slices": $h_i[k] = \sum_{s=0}^{S-1} \begin{cases} h_i^s[k] & \text{if } t^s \leqslant \frac{k}{f^0} < t^{s+1} \\ 0 & \text{otherwise.} \end{cases}$



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



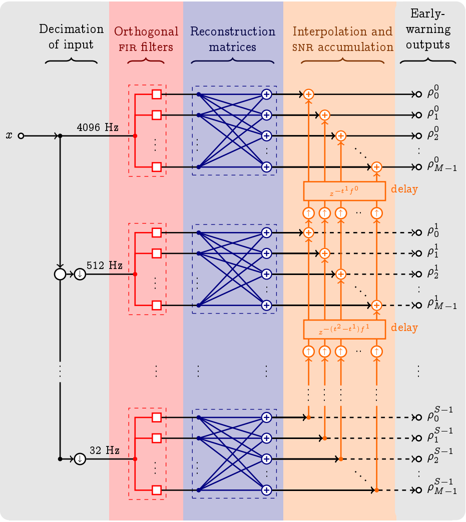
Time slices represent orthogonal subspaces, but within one time slice the templates are still highly correlated.

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).



SNR from previous time slices — orthogonal FIR filters

$$\rho_i^s[k] = \left(H^{\uparrow}\rho_i^{s+1}\right)[k] + \sum_{l=0}^{L^s-1} v_{il}^s \sigma_l^s \sum_{n=0}^{N^s-1} u_l^s[n]\, x^s[k-n]$$

early-warning output — reconstruction — decimated $h(t)$

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

method	flop/s (sub-bank)	latency (s)	flop/s (NS–NS)	number of machines
time domain	4.9×10^{13}	0	3.8×10^{15}	$\sim 3.8 \times 10^{5}$
frequency domain	5.2×10^{8}	2×10^{3}	5.9×10^{10}	~ 5.9
LLOID (theory)	6.6×10^{8}	0	1.1×10^{11}	~ 11
LLOID (prototype)	(0.9 cores)	0.5	———	$\gtrsim 10$

Related efforts

- "Reduced basis catalogs" arXiv:1101.3765
- "Summed parallel infinite impulse response" arXiv:1101.3765
- Porting of algorithms to GPUs

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.

arXiv:1107.2665

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.