An Aperçu about Gravitational Waves and Data Analysis

Eric CHASSANDE-MOTTIN
APC
AstroParticule et Cosmologie
Université Paris Diderot
CNRS/IN2P3
CEA/Irfu
Observatoire de Paris
Sorbonne Paris Cité, Paris, France

Abstract. On September 14, 2015 the US-based detectors LIGO made the first direct observation of gravitational waves. The signal extraction from the data of those instruments relies on a variety of search algorithms of which we review the key features in this article.

On September 14, 2015 at 9:50:45 UTC, the two detectors of the Laser Interferometer Gravitational-wave Observatory (LIGO) observed for the first time a gravitational-wave (GW) signal produced by a distant binary black-hole merger [1, 17].

This discovery is an experimental feat: it culminates decades of research and development leading to instruments [2, 17] that are sensitive enough to detect with high confidence the tiny strain of order 10^{-21} that gravitational waves exert on space-time when crossing the detectors.

Long-standing efforts have also been pursued for the development of data analysis methods in order to address a number challenging issues related to the detection and extraction of the gravitational-wave signal from the noisy data. Our focus here will be on transient gravitational waves.

GW data analysis is very much constrained by the nature of the detection problem: (*i*) five-sigma significance is conventionally requested to announce major discoveries which implies that the tail of the background noise has to be characterized to an extended depth, below p-value $\lesssim 10^{-7}$ (*ii*) it is impossible to shield the detector from gravitational waves and thus obtain signal-free data that can be used for the estimation of the background noise, (*iii*) because of the instrument complexity an accurate statistical model of the tails of the background noise dominated by impulsive, non-Gaussian disturbances (so-called "glitches") is out-of-reach.

The combination of those three facts, together with the typical size of the data set (time series of month duration sampled at few kHz, leading to billions of samples), put stringent computational requirements with direct implications on the analysis methods and their implementation.

Online searches enabling follow-up observations by other instruments (observing e.g., in the electromagnetic spectrum) are even more constrained as they intend to deliver results with about a minute latency.

A priori information plays a major rôle in detection problems. The GW waveform is intimately related to the dynamics of the source. For binary coalescences the orbital motion of the binary during the inspiralling phase and subsequent merger of the two bodies can be modelled to very good accuracy [8]. The predicted GW signal is quasi-periodic with an increasing frequency, i.e. a *chirp*. At leading order the chirp frequency sweeps towards high values according to a power law $f \propto t^{-3/8}$, where t is the time remaining before the final merger point.

This morphological information can help to distinguish the GW signal from the background noise perturbations. A first approach described in Sec. 2 is to perform searches targetting these waveform models using *matched filtering techniques* [18].

A second approach consists in "unmodelled" transient searches that don't rely on any prerequisite waveform model. These *excess power methods* discussed in the next section address non-parametric change point problems and identify departures from nominal noise fluctuations.

In the next sections, we give a "big picture" description of both approaches that condenses information reported in different places [1, 4, 3]. We also present the results obtained for the first GW event detected in September 2015.

1 Unmodelled transient searches

A whole range of data analysis pipelines has been developed for searching for unmodelled GW transients. These pipelines all rely on similar principles (see the recent reviews in [4, 7]). We concentrate here on one of those methods, Coherent WaveBurst (cWB) [14, 15].

In a nutshell cWB identifies outlier pixels in a time-frequency representation of the data and imposes that consistent pixels are identified at both detectors.

1.1 Trigger generation

In the cWB's scheme, the data are mapped to a time-frequency representation [13] by projecting onto a Wilson basis [9, 10]. Wilson bases are composed of wavelets obtained by modulating a base window with sine and cosine functions alternatingly. This alternating scheme allows to construct an orthonormal basis of Gabor-like wavelets that realize a short-term Fourier analysis, with good localization properties both in the time and frequency space.

The selected base window is the Meyer scaling function [12] which has a compact support in the Fourier space thus allowing a sharp decomposition in frequency subbands.

Following [16] a pair of in-phase and quadrature *dual* Wilson bases can be constructed from a single window. Time-frequency maps are computed by summing the powers in the in-phase and quadrature transforms. A collection of such maps is computed using windows of different durations thus allowing to match signal variability over different time scales.

Coefficients in this collection of time-frequency maps that fall outside the typical range expected for instrumental noise are retained. The consistency of the selected coefficients in the data from both observatories is evaluated by the "coherent" signal energy, a measure of *phase coherence* analog to beam-forming in sensor array analysis. A cluster of coefficients whose "coherent" energy dominates over "incoherent" energy form a GW candidate event. The coherent energy in the cluster also defines a statistic that ranks their likelihood of being a GW signal.

1.2 Assessing the event significance

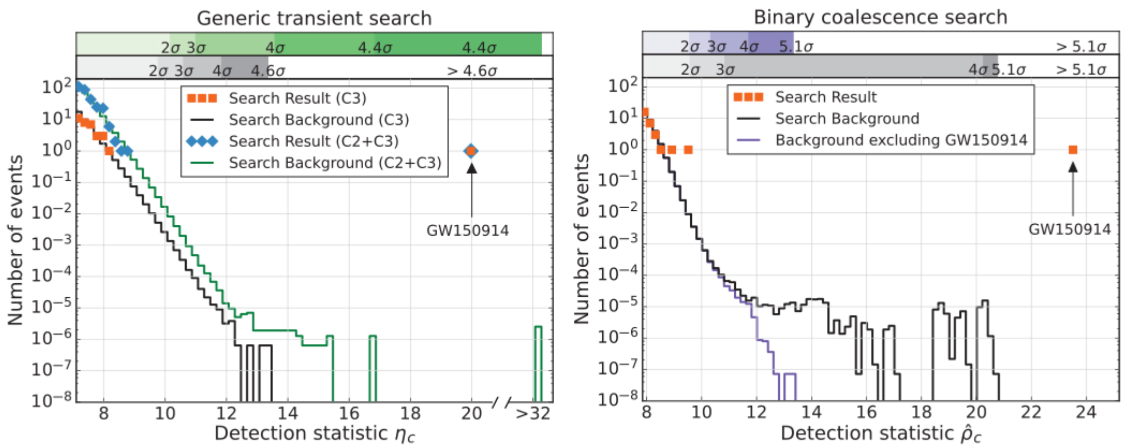

Figure 1: **Analysis backgrounds and search results for the unmodelled transient search (left) and the binary coalescence search (right).** These histograms show the number of candidate events (orange) and the mean number of background events (black lines) in the search class (see text) where GW150914 was found as a function of the search detection statistic (bin width of 0.2). The background associated with the black line in the right panel has been calculated with all single-detector triggers including those associated with GW150914. Random coincidences between GW150914 triggers in one detector with a glitch in the other detector create the long tails. The purple curve is the background excluding those coincidences. From [1].

The significance of a candidate event is determined by the rate at which the detector noise generates events with comparable or larger statistic. This rate is estimated empirically by repeating the analysis many times on surrogate data obtained by artificially shifting the time-stamps of one detector streams by a time offset ("time-lags") much larger than the inter-site propagation time. The distribution of candidate events obtained from this procedure constitutes the analysis background. The background is typically computed with a million of time-lags, corresponding to order of 100 000 years of surrogate data.

The background is evaluated independently for different classes of events based on their morphology (duration, central frequency, bandwidth, Q-factor, etc). The idea is to define classes that confine the contribution to the background due to identified noise perturbations and free the background of the other classes from the masking effect of those noise perturbations. While the introduction of multiple search classes significantly improves the overall sensitivity that results from the combined outcome from all classes, it has a price as the final significance needs to be decreased by a trials factor that accounts for running multiple searches. The conservative assumption that the backgrounds in the different classes are independent, results in a trials factor equal to the number of classes.

Three mutually exclusive classes are introduced: class C1 essentially collects events with a low Q-factor and gathers the dominating population of spurious noise glitches occurring in the LIGO data, class C2 collects events that does not display any increasing frequency pattern. Instead the last class C3 collects events with an increasing frequency. It is thus the class that is most sensitive to GW signals from binary mergers. The significance of candidate events is measured against the background from the same class.

The search sensitivity is also significantly improved by the thorough study [5] of the instrumental noise using more than 200 000 auxiliary channels that monitors the instrument behavior and environmental condition. This allows to identify and mitigate the impact of noise glitch sources that can be characterized and measured independently using one or several auxiliary channels.

Fig. 1 left panel shows the backgrounds obtained for the classes C2+C3 and C3 alone for the first 38 days of the O1 science run initiated on Sep 12 2015. The search results are also overlaid in this diagram. The event detected on Sep 14 2015, labeled GW150914, is associated with class C3 and stands clearly above the background in this class with a post-trial significance of 1 in 22 500 years which corresponds to 4.6σ.

This evidences that unmodelled searches can detect GW from binary mergers. But their scope is much larger as those searches can also detect GW from other transient astrophysical phenomena, including those that have not been anticipated.

2 Binary coalescence searches

The search for signal with *a priori* known shape is efficiently performed by *matched filtering techniques* [18]. Different matched filtering-based pipelines have been applied to LIGO data (see [3, 6] for a review). While there are significant differences in their implementation they all rely on the same core principle consisting in cross-correlating the data with the expected "template" waveforms obtained from the source model. Here again, we concentrate on one of those pipelines, PyCBC [19].

The targeted GW signal is known up to few parameters including the component masses and spins and other geometrical factors such as the binary position and orientation. The latter are merely overall scaling factors but the former are the source of intrinsic variability in the signal model as they modify the evolution of the signal phase.

A bank of templates is necessary in order to detect any signal in the space of possible signals spanned by the component masses and spins. The bank is laid out in such a way that there is always "close" templates in every part of the signal space. The signal space covered has component masses from 1 to 99 $M_\odot$ with total mass less than 100 $M_\odot$, and spins up to 0.99 (assumed aligned with the orbital angular momentum) [3]. A large fraction of the 250 000 template waveforms in the bank are computed with the effective-one-body formalism which combines results from the post-Newtonian approach with results from black hole perturbation theory and numerical relativity [8].

The data of each detector is whitened by dividing them by an estimate of the power spectral density and correlated with each template in the bank. The result of this correlation defines the signal-to-noise ratio. The search identifies local maxima of the signal-to-noise ratio with respect to the time of arrival of the signal. A goodness-of-fit measurement of the data with the matching template is evaluated by a chi-squared statistic. The signal-to-noise ratio is down-weighted for events with large chi-square.

Triggers at both detectors that occur in *coincidence* within 15 ms and from the same template are selected and marked as GW candidate events. The quadrature sum of the re-weighted signal-to-noise ratios of the triggers at both detectors defines a statistic that ranks the likelihood for the associated candidate event of being a GW signal.

The remaining of the search follows the steps done for unmodelled transient searches and described in the previous section, namely: the analysis background is calculated using the same method with 10 millions time-lags which corresponds to 500 000 years of surrogate time. The signal space is divided into three search classes based on the template duration [3].

Fig. 1 right panel shows the background of the class where GW150914 belongs, together with the search results obtained for the first 38 days of the O1 science run. Again, GW150914 stands clearly above the background of its class with a post-trial significance of 1 in 203 000 years, which corresponds to 5.1σ.

3 Concluding remarks

We presented two distinct and complementary approaches for transient GW detection: targeted searches based on matched filtering and "eyes wide open" searches for generic GW transients. Both have successfully detected the first GW event GW150914 with high significance. We summarized the results obtained for the first part of the O1 science run. Additional data have been collected and analyzed since, leading to more discoveries (see [6] for a full and final report about the detected events during the O1 science run).

This article mainly focuses on event detection. GW data analysis is much wider in scope and includes contributions that are left untold.

We very briefly mentioned detector characterization. This is an area of research with a significant margin for improvement and major potential impact on the search sensitivity, thus on science reach. Advanced machine learning techniques may play a key role in this domain.

Once a signal is detected, the GW source has to be characterized and its parameters estimated. For compact binary mergers this task is currently done using Bayesian estimation methods that provide joint or marginal posterior distributions for those parameters. Bayesian samplers currently requires a significant amount of computing resources and provides results with a latency of a few days. The reduction of this latency is an area of active research and is particularly relevant for guiding follow-up observations of GW events.

References

[1] Abbott, B., et al.: Observation of gravitational waves from a binary black hole merger. Phys. Rev. Lett. **116**, 061102 (2016).

[2] Abbott, B., et al.: GW150914: The Advanced LIGO Detectors in the Era of First Discoveries Phys. Rev. Lett. **116**, 131103 (2016).

[3] Abbott, B., et al.: GW150914: First Results from the Search for Binary Black Hole Coalescence with Advanced LIGO Phys. Rev. D **93**, 122003 (2016).

[4] Abbott, B., et al.: Observing gravitational-wave transient GW150914 with minimal assumptions Phys. Rev. D **93**, 122004 (2016).

[5] Abbott, B., et al.: Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914 Class. Quant. Grav. **33**, 134001 (2016).

[6] Abbott, B., et al.: Binary Black Hole Mergers in the First Advanced LIGO Observing Run Phys. Rev. X **6**, 041015 (2016).

[7] Abbott, B., et al.: All-sky search for short gravitational-wave bursts in the first Advanced LIGO run. arXiv:1611.02972, (2016).

[8] Damour, T.: Gravitational Waves and Binary Black Holes. This volume, 2016.

[9] Daubechies, I.: *Ten Lectures on Wavelets*. Number 61. SIAM, 1992.

[10] Daubechies, I., Jaffard, S., and Journé, J.-L.: A Simple Wilson Orthonormal Basis with Exponential Decay, SIAM J. Math. Anal. **22** (2), 554–573 (1991).

[11] Chassande-Mottin, E., and Flandrin, P.: On the time-frequency detection of chirps. Appl. Comp. Harm. Anal. **6**, 252–281 (1999).

[12] Jaffard, S., Meyer, Y., et Ryan, R.: *Wavelets: Tools for science and technology.* SIAM, 2001.

[13] Flandrin, P.: *Time-frequency/time-scale Analysis.* Academic Press, 1999.

[14] Klimenko, S., et al.: Constraint likelihood analysis for a network of gravitational wave detectors. Phys. Rev. D **72**, 122002 (2005).

[15] Klimenko, S., et al.: Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. Phys. Rev. D **93**, 042004 (2016).

[16] Necula, V., et al.: Transient analysis with fast Wilson-Daubechies time-frequency transform. J. Phys. Conf. Ser. **363**, 012032 (2012).

[17] Shoemaker, D. H., Man, C. N., Vinet, J.-Y.: Detecting Gravitational Waves: Theory, Experiments and Results. This volume, 2016.

[18] Schutz, B. F.: *The detection of Gravitational Waves*, Cambridge University Press, Cambridge, England, 1991, chap. Data processing, analysis and storage for interferometric antennas, pp. 406–452, D. Blair, ed.

[19] Usman, S. A., et al.: The PyCBC search for gravitational waves from compact binary coalescence Class. Quant. Grav. **33**, 215004 (2016).