Detection Confidence Tests for Burst and Inspiral Candidate Events

Romain Gouaty for the LIGO Scientific Collaboration

Louisiana State University, Baton Rouge, LA 70803, USA

E-mail: `romain@phys.lsu.edu`

Abstract. In order to detect gravitational-wave transients, such as compact binary mergers or core-collapsed supernovae, the LIGO Scientific Collaboration (LSC) is using analysis pipelines which aim to reduce the rate of false-alarm triggers without rejecting gravitational-wave signals. However, because of the non-Gaussian, non-stationary nature of the noise exhibited by the LIGO detectors [1], residual false alarms might be found at the end of the pipelines. A critical aspect of the search is then to assess our confidence for gravitational waves and to distinguish them from those false alarms. Both the "Compact Binary Coalescence" and the "Burst" working groups have been developing a detection checklist for the validation of candidate-events, consisting of a series of tests including data quality checks, analysis of the candidate appearance, parameter consistency studies, coherent analysis, which aim to corroborate a detection or to eliminate a false alarm. In this paper, the general methodology used for candidate validation is presented. The method is illustrated with an example of simulated gravitational wave signal and a false alarm.

1. Introduction

The Compact Binary Coalescence (CBC) working group of the LIGO-Virgo [1, 2] joint collaboration is a data-analysis group looking for gravitational-wave signals emitted by inspiralling compact binary systems [4, 5]. The duration that such signal spends in the frequency bandwidth of the current ground-based gravitational-wave interferometers lies in a range which goes from a few tenths up to several tens of seconds.

The Burst working group of the LIGO-Virgo joint collaboration is a data-analysis group looking for transient gravitational-wave signals of typical duration of the order of the milli-second or the tenth of second, usually without specific assumptions on the expected waveform.

Both of the working groups previously mentioned execute analysis that are sensitive to the non-stationary and non-gaussian noise of the detectors. Because of the time-frequency properties of the signals that are being looked for, noise transients can actually induce false alarm triggers in the analysis pipelines that may result in accidental coincidences between the interferometers. It is therefore crucial to submit each gravitational-wave candidate identified by the analysis to a detection checklist which aims to estimate confidence in this candidate.

The detection checklist is made of a list of standard tests in different stages of development that are used to review the gravitational-wave candidates. An overview of this checklist is provided in section 2. Section 3 shows a method implemented by the Burst and CBC groups to estimate the statistical significance of the candidates. In section 4, the paper will describe with more details a few items of the detection checklist, using an example of simulated gravitational-wave signal and false-alarm trigger for illustration purposes.

2. Overview of the detection checklist

A detection checklist to evaluate the significance of candidate-events has been developed by each of the CBC and Burst groups. Despite some specificities inherent to the kind of signals that are being looked for by the CBC and Burst searches, the method implemented to estimate confidence in a gravitational-wave candidate is very similar between the two data-analysis groups. Thus the summary of the checklist provided in this section applies both to the CBC and Burst groups, unless otherwise noted. The list presented below is a short synthesis of the tests implemented in the CBC and Burst detection checklists. As many of these tests are still under development or refinement, the checklist is rapidly evolving, and the following list should not be considered as exhaustive. Here we outline the main tests that are currently part of the detection checklist for candidate-events or in the process of implementation:

- **Statistical significance** The first step of the candidate validation procedure consists of determining the statistical significance of the candidates identified by the analysis pipeline, that is to say the probability of coincident triggers arising

from random coincidences of noise triggers. The general method is described in section 3, which will also explain how the statistical significance of the candidates affects the way the other tests of the checklist are addressed.

- **Data integrity:** sanity checks to verify that the data set containing the candidate is not corrupted.

- **Status of the interferometers:** The state of the LIGO interferometers and their sensitivity near the time of the candidate are checked. This test also includes a verification of the data quality flags recorded in the database. Section 4.1 will show how this test can allow the identification of noisy data segment containing false-alarm triggers.

- **Environmental or instrumental causes:** We analyze the auxiliary channels of the interferometers, such as the environmental sensors or the signals involved in the mirror control loops, to check for the presence of possible noise transients which could be the cause of a false alarm identified by the analysis. This effort is also part of the "glitch group" activities [6]. More details on this part of the detection checklist are provided in section 4.2.

- **Candidate's appearance:** Part of the detection checklist consists of tests of the candidate's appearance which aim to confirm the presence of a gravitational-wave signal in the data or to identify obvious excess of noise responsible for a false alarm. A variety of tools are used to examine the data containing the candidate-event, such as time series, time-frequency spectrograms, or the outputs of the search pipeline, namely the signal to noise ratio (SNR) or the χ^2 [8] time series in the case of the CBC search [7]. The analysis of this graphical information is performed by comparing the results obtained at the time of the candidate-event to the expectations for simulated gravitational-wave signals or for known instances of false-alarm triggers. Examples of tests of the candidate's appearance will be presented in section 4.3. A complementary and quantitative test based on the candidate's parameters is in preparation as mentioned in the next item.

- **Consistency of the candidate's estimated parameters:** Automated tools are being developed in order to establish a likelihood ranking of the candidates given their estimated physical parameters. This relies on the comparison of parameters' distribution for simulated gravitational-wave signals and for known accidental coincident triggers.

- **Detection robustness:** The Burst group has developed many independent search algorithms looking for unmodelled gravitational-wave signals, a few instances of which are `Block-Normal` [9], `Kleine-Welle` [10], `Q-online` [11] and `Waveburst` [12]. The robustness of a detection for burst searches is checked by verifying that the candidate-event is identified by different independent algorithms. In the case of the CBC search, the list of candidate-events is obtained from a single analysis pipeline based on match filter algorithm [7] using specific inspiral waveforms [13]. However a Bayesian analysis dedicated to the search for inspiral signals [14] has also been

developed and the CBC group has started using it as an independent algorithm to check the robustness of the detection. Part of the detection robustness' test also consists in verifying the accuracy of the detectors' calibration [15, 16] at the time of the detection. A current effort of the CBC group aims to automate a test that will check the impact of the possible errors in the calibration for the analysis. To this purpose the data will be reanalyzed using a calibration function that will be modified according to the possible uncertainties on the measurement.

- **Correlation between interferometers:** Other tests of the detection checklist consist in checking for the correlation between the signals measured in the different detectors of the network composed of the three LIGO interferometers [1], GEO600 [3] and Virgo [2]. The Burst group has developed many network or coherent analysis [24], one example of which is `Coherent Waveburst`, based on a constraint likelihood method [17, 18]. The CBC group is developing two tests that check for the expected signal correlations in the data containing coincident triggers from multiple detectors. One of these tests uses the matched filtering algorithm [7] to compute the multi-interferometer coherent SNR [19, 20] and compares it with the null-stream statistic [21] for inspiral waveforms to assess the significance of a trigger being an inspiral event. The essential idea behind this test is that above a certain threshold value for the coherent SNR, real gravitational-wave signals will yield a smaller value for the null-stream statistic than instrumental or environmental glitches of the same coherent SNR. The second test uses a Bayesian approach to infer the posterior distributions of the signal parameters [22]. Both are currently tested as part of the detection checklist. In addition to these checks for correlation between interferometers, the analysis groups also verify if the candidate-event is identified by the resonant bar detectors [23].

- **Coincidence with external searches:** The LIGO Scientific Collaboration is implementing procedures to check for possible coincidences between gravitational-wave candidates and triggers from external searches, such as Gamma-Ray Bursts, optical transients or neutrinos observations [24].

3. Statistical significance of the candidates

In order to estimate the statistical significance of the candidate-events, the CBC and Burst searches compare the triggers found in coincidence between at least two of the LIGO interferometers to an expected background of accidental coincidences. This background is estimated by repeating the analysis after time-shifting the data of each interferometer with respect to each other. This method, called the time-slides analysis, has already been described in previous publications such as [25]. Fig. 1 shows an example of comparison between the in-time coincident candidate-events and the expected background for the Binary Neutron Star search run over the data taken during the LIGO fourth (S4) science run [25]. The goal of the candidate's follow-up with the checklist differs whether the candidate has a low probability of being an accidental
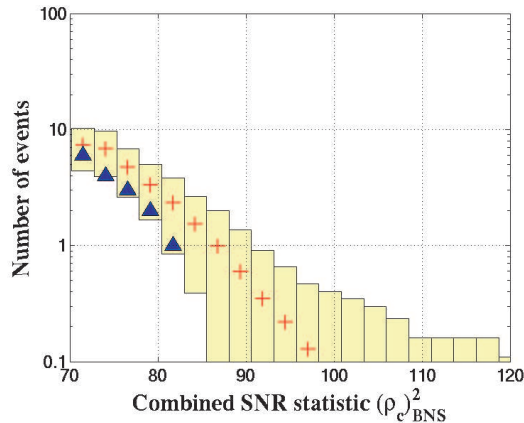
Figure 1. Cumulative histogram of the combined SNR, ρ_c, for the S4 Binary Neutron Star search: for the in-time coincident candidate-events (triangles), and for the estimated background of accidental coincidences (crosses and 1 standard-deviation ranges). This plot has been extracted from [25]. All candidates were found consistent with the background.

coincidence or not. If the candidate is consistent with the estimated background (as the candidate-events shown in Figure 1), then the goal of the follow-up is to perform a sanity check to make sure that there is no obvious gravitational-wave signal hidden in the estimated background. Such candidate shall not lead to a detection claim unless strong evidence of an error in the background estimate. In case the candidate has a low probability of accidental coincidence (i.e. if the candidate is lying above the estimated background), the goal of the follow-up is then to strengthen our confidence in a possible detection by submitting the candidate to the detection checklist that gravitational-wave signals should pass succesfully.

4. Detailed examples of the checklist

This section will highlight a few examples of tests used for the review of candidate-events. In order to illustrate the expected results for an inspiral gravitational-wave signal we will refer to a simulated inspiral signal. This simulation was performed by acting on one of the arm test masses of the interferometers to generate a differential motion of the interferometer arm cavities approximately equivalent to the expected effect of a gravitational wave. This simulated gravitational-wave signal was detected by the CBC analysis and stands as an outlier above the estimated background. We will also illustrate the behaviour of the detection checklist when it is applied to an example of false alarm. In the following subsections, we will refer to the simulated gravitational wave as *Candidate G* and to the false alarm as *Candidate F*.
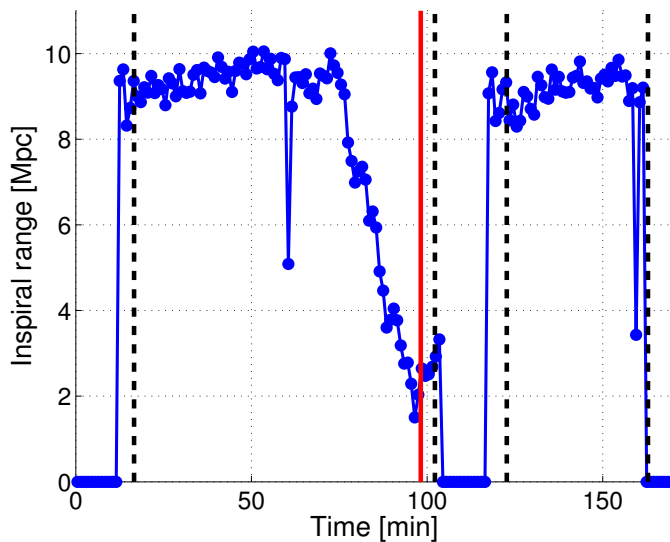
Figure 2. Inspiral range as a function of time at Livingston (the origin of time is arbitrary). The dashed vertical lines indicate the start and end times of the segments when the interferometer was in science mode. Two science segments are represented in this figure (corresponding to the time windows 16.5-102.2 min and 122.7-162.9 min). The *Candidate F* was found in the first science segment, at t≃98.2 min. This time is underlined by a solid vertical line.

4.1. Status of the interferometers

A part of the detection checklist consists in verifying the status of the interferometers and the data quality in the segment containing the candidate. This includes examining the figures of merit (detectors state, sensitivity, seismic trends) posted in the detectors log, scanning the database which contains the list of data quality flags, as well as checking the information reported by the "glitch group" [6]. The goal of this study is to check for a possible misbehaviour of the detectors or an unsual excess of noise which could translate into a higher rate of false-alarm triggers and thus reduce our condifence in the candidate-event. For instance we check how the detectors sensitivity varies in time and how it might affect the performances of our searches.

Figure 2 shows an example of figure of merit displaying the minute trends of the inspiral search horizon distance at the Livingston's site, called *inspiral range*, for a time window of about 3 hours which includes the *Candidate F*. The *inspiral range* is defined as the horizon of the search for 1.4-1.4 $M_\odot$ systems (binary neutron stars systems), averaged over all sky positions and orientations, assuming a detection with a conventional SNR equal to 8. During the day from which the three hours of data shown in figure 2 have been extracted, the typical *inspiral range* measured in science mode was fluctuating between 8.5 and 10.5 Mpc. However one can notice that the first science segment shown in figure 2 (between t=16.5 min and t=102.2 min) terminates with a dropping *inspiral range* for about twenty five minutes. The inspiral trigger associated with *Candidate F*

(false alarm) in the Livingston data was found inside this segment (it is highlighted by the solid vertical line in figure 2) while the *inspiral range* was about 2.5 Mpc, that is to say well below the averaged sensitivity reached by the interferometer during this day. This indicates that the *Candidate F* was detected while the Livingston interferometer was exceptionally noisy. Such an observation reduces our confidence in *Candidate F*, although this check does not prove that the candidate itself is due to detector noise. A confirmation of the nature of the candidate is brought by its SNR or χ^2 [8] time series as discussed in section 4.3.

4.2. Environmental and instrumental causes
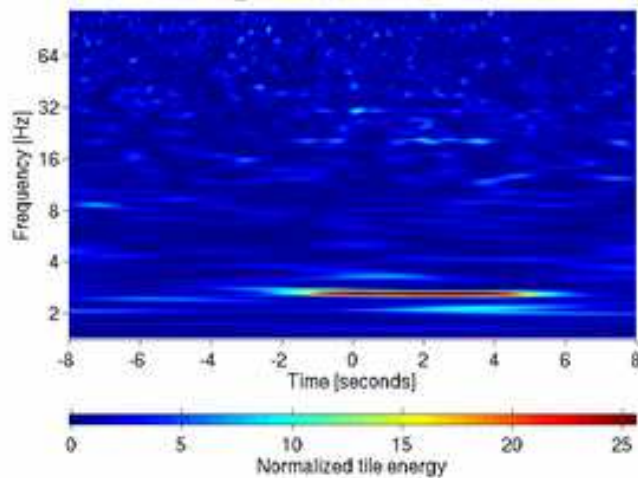
Figure 3. *Q spectrogram* of a transient in a seismometer located near the end mirror test mass of the Hanford 2 km interferometer. This low frequency transient (f$\simeq$2.6 Hz) is time-coincident with the inspiral trigger associated to the *Candidate G* (simulated gravitational-wave signal) whose position corresponds to the origin of the x axis (t=0s).

In order to check for possible instrumental artefacts that could be responsible for false alarm triggers, we examine the auxiliary channels of the detectors in a few seconds long window around the candidate-events. For this purpose time-frequency maps of auxiliary channels are being analysed, using an event visualization tool called QScan, which is based on a Q-transform. More details on this tool can be found in [11]. *Qscan* produce time-series and *Q spectrograms* of the auxiliary channels in which transients are detected. The *Q spectrograms* correspond to time-frequency decompositions using sinusoidal Gaussians characterized by a central time, central frequency, and a quality factor Q. An example of *Q spectrogram* is provided in figure 3, where the examined channel is a seismometer located near the end mirror test mass of the Hanford 2 km interferometer. This kind of channel can measure seismic disturbances at frequencies below a few Hertz. In figure 3 a low frequency transient (approximately at 2.6 Hz) lasting for a
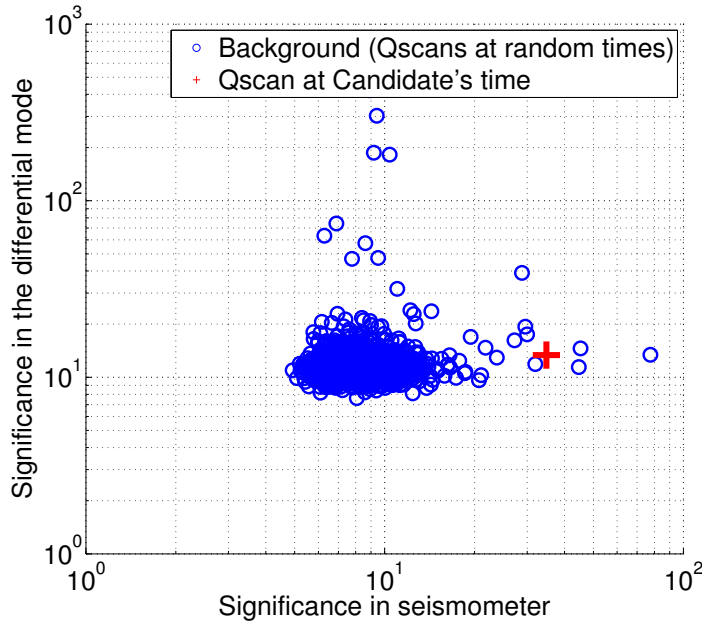
Figure 4. Scattered plot of *QScan significance* in a seismometer channel (x axis) and in the error signal of the differential mode control loop (y axis) of the H2 interferometer. The "circles" refer to the *QScan significance* measured at times randomly distributed over the first calendar year of the fifth LIGO science run ("S5") [1] to estimate the seismometer background. The "plus" symbol refers to the *QScan significance* measured at the time of the *Candidate G* (simulated gravitational-wave signal).

few seconds (between t≃-2s and t≃6s along the time axis) is visible. Moreover this seismometer's transient is coincident with the *Candidate G* (simulated gravitational-wave signal) whose time corresponds to t=0s in figure 3. When a candidate-event is simultaneous with a transient in an auxiliary channel, further investigations are performed as explained below.

Statistical significance of instrumental transients

The statistical relevance of the transient found in the auxiliary channel is calculated by comparing its strength (which is characterized by a parameter called *Z significance* [11], equivalent to a SNR squared for the Q-transform) to an estimated background. For each auxiliary channel the background is estimated by running `QScan` at times which are randomly distributed over epochs of the data-taking. This provides a distribution of *Z significances* corresponding to the background of the auxiliary channel. Since the background of auxiliary channels might change during a long data-taking due for example to seasonal variations of the environmental noises or to instrumental drifts, we perform several estimations of the background, over large (such as a year) and short (such as a few days) epochs, and compare the results.
Figure4 shows a scattered plot of the *Z significance* measured in the seismometer versus

the *Z significance* measured in the error channel of the differential mode control loop of the H2 interferometer (which is the main port sensitive to gravitational waves). The "circles" refer to the estimated background over the first calendar year of the fifth LIGO science run ("S5") [1] while the 'plus' symbol shows the significances measured in both channels at the time when the *Candidate G* was detected. A seismic transient whose *Z significance* would be comparable with the median of the background distribution could be ignored as irrelevant. On the contrary the seismic transient shown in Figure 3 has a higher significance than the background, which makes it statistically relevant. Another estimation of the seismometer background was obtained by analyzing only two days of data including the time of the *Candidate G* and led to a similar conclusion. If we faint to ignore the nature of the *Candidate G*, the next question that needs to be addressed is to identify whether or not this environmental transient couples to the interferometer output port.

Coupling of environmental disturbances into the interferometer output port

The coupling of an environmental distrubance into the interferometer output port can be proven by comparing the *Q spectrogram* of the auxiliary channel to the *Q spectrogram* of the interferometer output port and by looking for possible correlations between these two channels. A high frequency disturbance might couple linearly into the gravitational-wave bandwidth of the output port. When a measured transfer function from the auxiliary channel to the output port is available one can then compare it to the amplitude ratio measured in the *Q spectrograms* of the two channels. If such a coupling is proven, this leads to the rejection of the candidate as a possible detection.

In the case of a low frequency seismic transient, noise upconversion mechanisms might induce a false-alarm trigger in the gravitational-wave bandwidth. One can notice in Figure 4 that the *Z significance* measured in the error signal of the interferometer differential mode at the time of the *Candidate G* is comparable to the significances obtained at random times. Therefore, despite the statistical relevance of the transient in the seismometer, Figure 4 does not argue in favor of a possible coupling into the interferometer output port. Knowing that the *Candidate G* is a simulated gravitational-wave signal, the presence of an inspiral trigger is indeed not related to the seismic transient.

4.3. Candidate appearance

In this section two examples of qualitative checks of the candidate's appearance are illustrated: a check of the candidate's time-frequency map, and a check of the output of the match-filtering algorithm [7] used to search for inspiral gravitational-wave signals.

Q spectrograms of the candidate

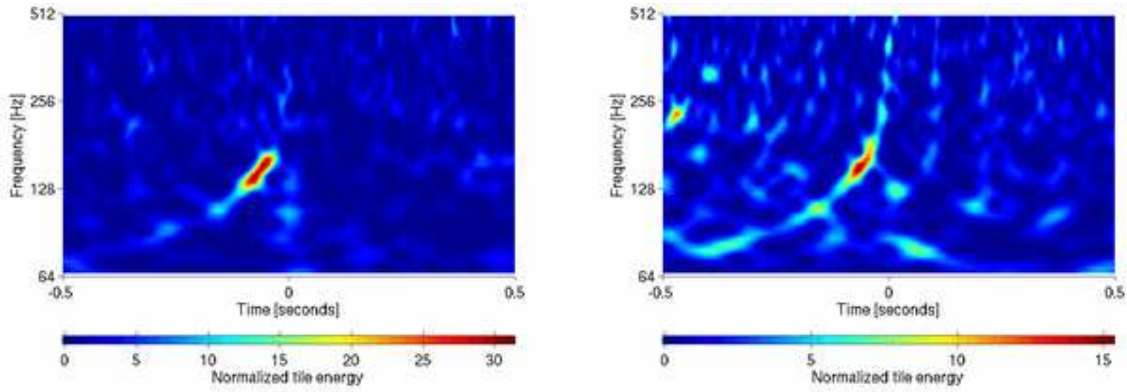A `QScan` of the data in which a candidate-event as been detected is examined in or-

Figure 5. *Q spectrograms* of the data containing the *Candidate G* (simulated gravitational-wave signal) in the two more sensitive detectors: Hanford 4 km (left) and Livingston 4 km (right). A "chirp" waveform (frequency increasing with time) is visible in the data of the two detectors.
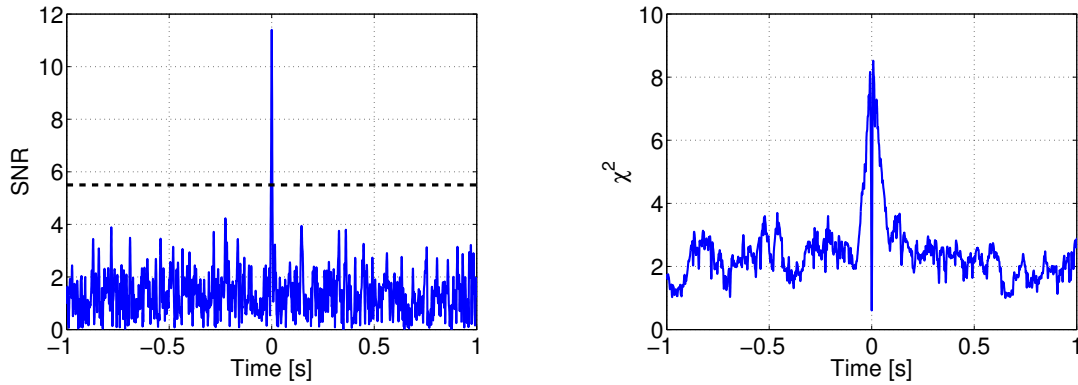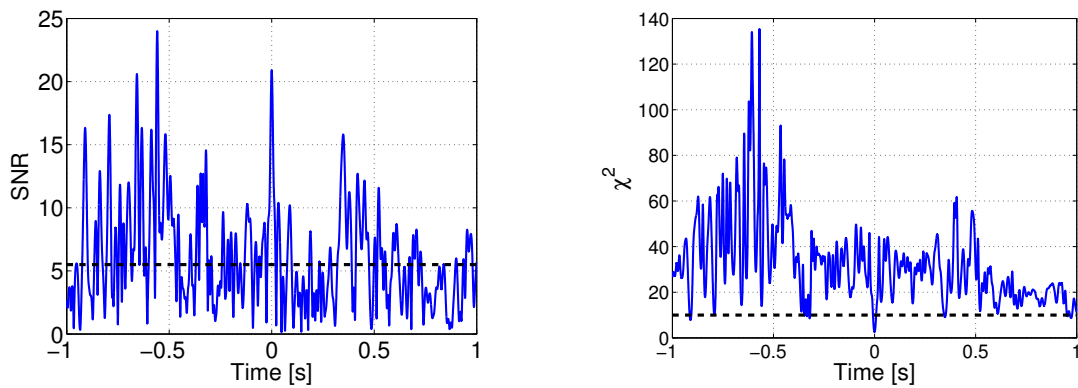

Figure 6. SNR (left) and χ^2 (right) time series obtained after match-filtering the Livingston data containning the *Candidate G* (simulated inspiral gravitational-wave signal). The time origin on the x axis coincides with the time of the inspiral trigger.

Figure 7. SNR (left) and χ^2 (right) time series obtained after match-filtering the Livingtson data containning the *Candidate F* (false alarm). The time origin on the x axis coincides with the time of the inspiral trigger.

der to perform the following checks:

- The presence of a possible known signal waveform that might confirm the detection is verified. However, a low SNR inspiral signal is not expected to be visible in a *Q spectrogram*. Therefore the absence of visible known waveform in the spectrogram does not rule out a possible detection.
- The presence of an obvious excess of noise in the data is also checked.

Figure 5 shows the *Q spectrograms* of the *Candidate G* in the two more sensitive interferometers where this simulated gravitational-wave signal was injected. The transient visible in the H1 and L1 data corresponds to the typical "chirp" pattern (frequency increasing with time) that is characterestic of an inspiral signal. The simulated gravitational-wave signal is thus visible in these two spectrograms. It is actually not clearly visible in the *Q spectrogram* of the H2 data (not represented here) because of the lower SNR in this interferometer.

Output of the match-filtering algorithm

Another example of check for the candidate's appearance that is used by the CBC group consists in examining the time-series of the SNR obtained after match-filtering [7] the data with inspiral waveforms [13], as well as the time-series of a χ^2 which aims to test the consistency between the triggered waveform and the signal present in the data.

An example of the expected time-series for a simulated gravitational-wave signal is shown in Figure 6. On the left plot, the SNR time series shows a short central peak corresponding to the time of the trigger associated with the simulated inspiral signal. On the right plot, the χ^2 time-series presents a very characteristic shape for a few tens of milli-seconds around the inspiral trigger, which corresponds to the expectations for gravitational-wave signal in stationary gaussian noise.

Figure 7 shows the SNR and χ^2 time-series around the time of the *Candidate F* at Livingston. Multiple peaks of SNR are visible, which indicates highly non-stationary data. Moreover the χ^2 time series shows large values for the whole two seconds window surrounding the candidate, which again indicates a very noisy stretch of data. Accordingly the *Candidate F* can be ruled out as a possible detection, which confirms the first suspicions born from the analysis of the inspiral horizon in section 4.1.

5. Conclusions and perspectives

The Burst and CBC groups are pursuing the refinement of the detection checklist for candidate-event validation. Part of these tests are still under development. Efforts of the groups are currently aiming to automate this detection checklist in order to build a candidate follow-up pipeline which will improve the swiftness of the analysis. The detection checklist is presently being applied to the candidate-events obtained by the searches analyzing the data taken during the fifth LIGO science run ("S5") [1].

The detection checklist should play an even more crucial role in the analysis of the future LIGO science runs, for which we expect better detectors' sensitivities and higher probabilities of gravitational-wave detections.

6. Acknowledgments

7. References

[1] Abbott B *et al* (LIGO Scientific Collaboration), 2007 "LIGO: The Laser Interferometer Gravitational-Wave Observatory", *Preprint* arXiv:0711.3041
[2] Acernese F *et al* (Virgo Collaboration), 2007 Class. Quantum Grav. **24** S381
[3] Lück H *et al*, 2006 Class. Quantum Grav. **23** S71
[4] Thorne K S, 1987 "Gravitational radiation" in "300 years of gravitation", Cambridge University Press
[5] Blanchet L, Damour T, Iyer B, Will C, Wiseman A, 1995 Phys. Rev. Lett. **74** 3515
[6] Blackburn L *et al*, 2008 "The LSC Glitch Group: Monitoring Noise Transients During the 5th LIGO Science Run", Submitted to Class. Quantum Grav., *Preprint* arXiv:0804.0800
[7] Allen B, Anderson W, Brady P, Brown D, Creighton J, 2005 "FINDCHIRP: an algorithm for detection of gravitational waves from inspiraling compact binaries", arXiv:gr-qc/0509116
[8] Allen B, 2005 Phys. Rev. D **71** 062001
[9] McNabb J *et al*, 2004 Class. Quantum Grav. **21** S1705
[10] Chatterji S, Blackburn L, Martin G, Katsavounidis E, 2004 Class. Quantum Grav. **21** S1809
[11] Chatterji S, 2005 Ph.D thesis, MIT
[12] Klimenko S, Mitselmakher G, 2004 Class. Quantum Grav. **21** S1819
[13] Blanchet L, Iyer B, Will C, Wiseman A, 1996 Class. Quantum Grav. **13** 575
[14] Röver C, Meyer R, Christensen N, 2006 Class. Quantum Grav. **23** 4895
[15] Landry M for the LIGO Scientific Collaboration, 2005 Class. Quantum Grav. **22** S985
[16] Siemens X, Allen B, Creighton J, Hewitson M, Landry M, 2004 Class. Quantum Grav. **21** S1723
[17] Klimenko S, Mohanty S, Rakhmanov M, Mitselmakher G, 2005 Phys. Rev. D **72** 122002
[18] Mercer A, Klimenko S, 2008 "Visualising GW Event Candidates Using the Coherent Event Display", in preparation for submission to Class. Quantum Grav.
[19] Bose S, Pai A, Dhurandhar S V, 2000 Int. J. Mod. Phys. D **9**, 325, arXiv:gr-qc/0002010
[20] Pai A, Dhurandhar S, Bose S, 2001 Phys. Rev. D **64**, 042004, arXiv:gr-qc/0009078

[21] Guersel Y, Tinto M, 1989 Phys. Rev. D **40**, 3884
[22] Röver C, Meyer R, Christensen N, 2007 Phys. Rev. D **75** 062004
[23] Baggio L *et al* (AURIGA Collaboration and LIGO Scientific Collaboration), 2007 "A Joint Search for Gravitational Wave Bursts with AURIGA and LIGO", arXiv:0710.0497
[24] Abbott B *et al* (LIGO Scientific Collaboration and Virgo Collaboration), 2008 "Astrophysically Triggered Searches for Gravitational Waves: Status and Prospects", accepted for publication in Class. Quantum Grav., arXiv:0802.4320
[25] Abbott B *et al* (LIGO Scientific Collaboration), 2008 Phys. Rev. D **77** 062002