Used percentage veto for LIGO and virgo binary inspiral searches

2010 J. Phys.: Conf. Ser. 243 012005

(http://iopscience.iop.org/1742-6596/243/1/012005)

14th Gravitational Waves Data Analysis Workshop IOP Publishing
Journal of Physics: Conference Series **243** (2010) 012005 doi:10.1088/1742-6596/243/1/012005

Used Percentage Veto for LIGO and Virgo Binary Inspiral Searches

Tomoki Isogai (for the LIGO Scientific Collaboration and the Virgo Collaboration)

Physics and Astronomy, Carleton College, Northfield, MN 55057, USA

E-mail: `isogait@carleton.edu`

Abstract. A challenge for ground-based gravitational wave detectors such as LIGO and Virgo is to understand the origin of non-astrophysical transients that contribute to the background noise, obscuring real astrophysically produced signals. To help this effort, there are a number of environmental and instrumental sensors around the site, recording data in "channels". We developed a method called the used percentage veto to eliminate corrupted data based on the statistical correlation between transients in the gravitational wave channel and in the auxiliary channels. The results are used to improve inspiral binary searches on LIGO and Virgo data. We also developed a way to apply this method to help find the physical origin of such transients for detector characterization. After identifying statistically correlated channels, a follow-up code clusters coincident events between the gravitational wave channel and auxiliary channels, and thereby classifies noise by correlated channels. For each selected event, the code also gathers and creates information that is helpful for further investigations. The method is contributing to identifying problems and improving data quality for the LIGO S6 and Virgo VSR2 science runs.

1. Introduction

LIGO [1] and Virgo [2] started science runs called S6 and VSR2, respectively, in July 2009 (GEO 600 [3] is also participating in this run). The LIGO Scientific Collaboration (LSC) and Virgo Collaboration are analyzing the data for direct observations of gravitational waves (GWs) of cosmic origin. The LIGO and Virgo detectors are, however, susceptible to various kinds of noise. Their sensitivity to GWs depends on how well such noise can be controlled. Naturally much effort in LIGO and Virgo goes into understanding noise and characterizing the data obtained. To maintain the quality of data and minimize the effect of noise transients, or so called “glitches”, we exclude, or more commonly referred to as “veto”, contaminated data that are not suited to be analyzed for GW detection. Since a number of environmental and instrumental noise sources can potentially couple into the main detector output, many of them are continuously recorded along with the data from the main detector output. The measurement points for time-series data from various monitors are referred to as “channels”. The channels include interferometer control signals such as length and angle control signals for optical cavities and physical environmental monitors (PEMs) such as seismometer, magnetometer and microphones around the interferometer.

Conventionally, we have only vetoed noise whose cause is fairly well understood with the consequence that many glitches remained in the science data due to their unknown origin [4, 5]. To eliminate the remaining glitches, we have developed a method called the used percentage veto (UPV). It utilizes the information from various instrumental and PEM channels, and uses their statistical correlation with the GW channel to exclude noisy times from our science data, as opposed to using knowledge about the deterministic causes of the noise. The result of this method is being used to improve the search for GWs from the inspiral of compact binary systems. In addition, the statistical data from UPV is used to help identify the physical cause of noise.

In section 2, we describe the method of UPV. Section 3 explains the application of UPV for detector characterization. Finally, a summary is given in section 4.

2. Used Percentage Veto

UPV finds auxiliary channels that are statistically correlated with the GW channel. To measure the correlation, we use KleineWelle (KW) triggers [6] (UPV can be applied to any trigger generation algorithm; we use KW for its availability on many channels and its relatively low latency). KW is a multiresolution method to find and characterize transients in an input timeseries. KW makes use of the dyadic wavelet transform to search for regions of excess energy in the time-scale decomposition. KW is applied to instrumental and environmental channels to find brief periods of excess power in each channel in near real time. Each trigger has a peak time and a significance (which describes the amplitude of triggers). The UPV code finds time-coincident triggers between the GW channel and an auxiliary channel with a ±1s coincidence window using triggers that 1) have at least a KW significance of 50 and 2) are in time segments after severely corrupted times with well-known causes are excluded [4, 5, 7]. Starting from a KW significance of 50, the code raises the threshold at a step interval of 50 up to 5000 (these values are chosen empirically considering accidentals and computational cost), and at each threshold the code calculates various metrics, or “figures of merit”, that measure how well the channel would work as a veto. These figures of merit are defined as follows:

- Used Percentage

$$\text{Used Percentage}(\rho) \equiv \frac{100 \times N^{\text{aux}}_{\text{coinc}}(\rho)}{N^{\text{aux}}_{\text{Total}}(\rho)} \tag{1}$$

where $N^{\text{aux}}_{\text{coinc}}$ is the number of the auxiliary channel KW triggers coincident with the GW channel above the KW significance threshold ρ and $N^{\text{aux}}_{\text{Total}}$ is the number of the total auxiliary channel KW triggers above the KW significance threshold ρ. This value tells how well the channel is correlated with the GW channel. For an ideal veto, the used percentage would be 100%, but in reality the used percentage remains under 50% for most vetoes. UPV requires a used percentage of at least 50% with $N^{\text{aux}}_{\text{coinc}}$ larger than 10 over the period of time analyzed (a week for S6/VSR2) for channels to be a veto candidate (see Fig. 1). These parameters are determined empirically to ensure sufficient correlation between the GW and auxiliary channel.

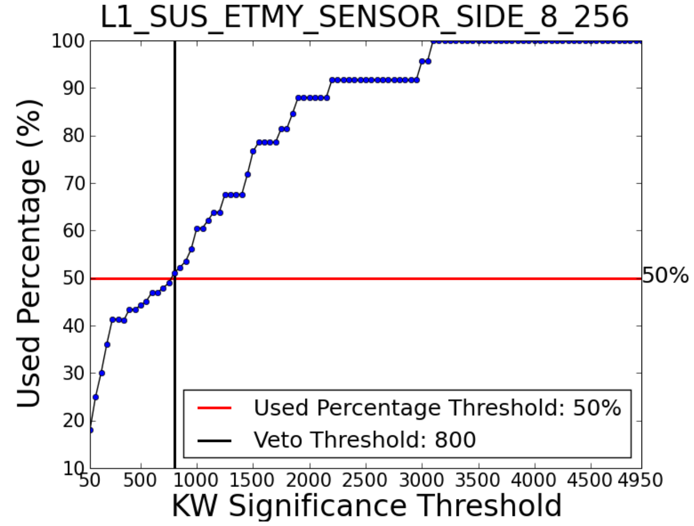


Figure 1: A plot showing used percentage as a function of KW significance for a particular suspension signal. UPV picks a significance threshold where the used percentage exceeds 50%, shown as a horizontal line in the plot.

- Efficiency

$$\text{Efficiency}(\rho) \equiv \frac{100 \times N^{\text{GW}}_{\text{vetoed}}(\rho)}{N^{\text{GW}}_{\text{Total}}} \tag{2}$$

where $N^{\mathrm{GW}}_{\mathrm{vetoed}}(\rho)$ is the number of vetoed GW channel triggers at the KW threshold ρ and $N^{\mathrm{GW}}_{\mathrm{Total}}$ is the total number of the GW channel triggers analyzed. This value tells how well the veto would eliminate glitches in GW channel.

- Dead Time

$$\text{Dead Time}(\rho) \equiv \frac{100 \times T_{\mathrm{vetoed}}(\rho)}{T_{\mathrm{Total}}} \tag{3}$$

where T_{vetoed} is the time vetoed at the KW threshold ρ and T_{Total} is the total time analyzed. A good veto has a high efficiency with a low dead time. Thus, the ratio

$$\frac{\text{Efficiency}(\rho)}{\text{Dead Time}(\rho)} \tag{4}$$

is often used to characterize the veto (see Fig. 2).

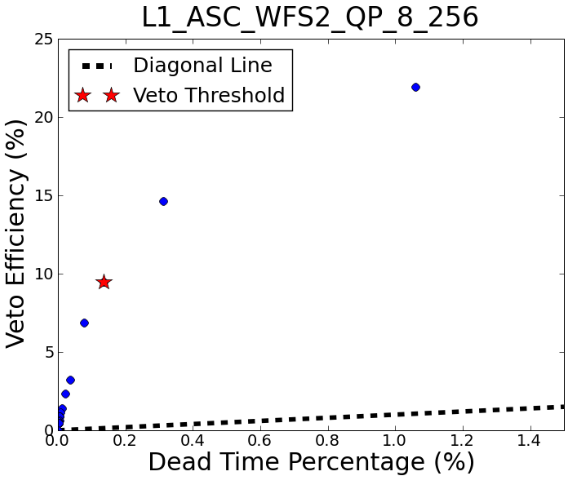


Figure 2: A plot showing the veto efficiency vs dead time for a particular wave front sensor. The dots correspond to each threshold, and the dashed line is the diagonal (efficiency = deadtime). A "good veto" has a high efficiency and a low dead time, so we would expect the points to be well above the diagonal line. In this case, we can see that the point at our threshold of choice (star) makes a very effective veto.

- Random Used Percentage

$$\text{Random Used Percentage}(\rho) \equiv \frac{\text{Dead time}(\rho) \times N^{\mathrm{GW}}_{\mathrm{total}}}{N^{\mathrm{aux}}_{\mathrm{total}}(\rho)} \tag{5}$$

This is the used percentage we expect when triggers in auxiliary channel are randomly distributed, i.e. not correlated with triggers in the GW channel. Therefore, the ratio

$$\frac{\text{Used Percentage}(\rho)}{\text{Random Used Percentage}(\rho)} \tag{6}$$

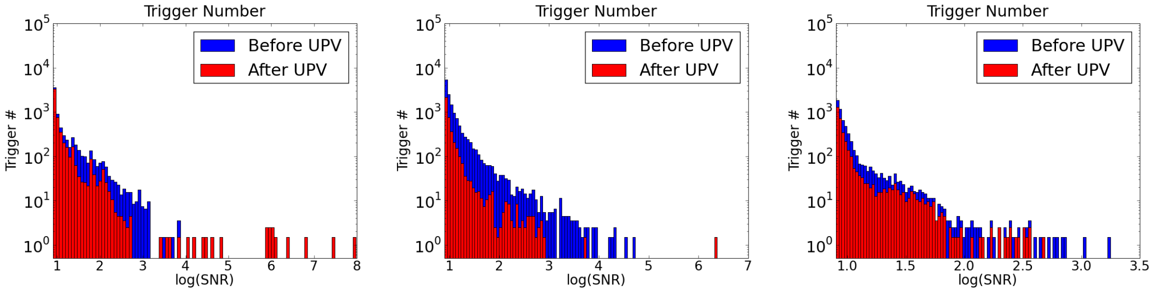
tells how well the channel is correlated with the GW channel above the random chance value. Note that this value is related to Efficiency/Deadtime by the factor

$N^{\mathrm{GW}}_{\mathrm{vetoed}}/N^{\mathrm{aux}}_{\mathrm{coinc}}$, so comparing the two ratios gives a rough idea as to what extent glitches are clustered. Even though UPV does not use these two ratios directly, they provide a useful means for sanity check and comparison with other vetoes.

Once all the figures of merit are calculated at each threshold, the UPV code creates veto segments for channels that pass the criteria described above. Using the KW significance threshold at which the channel exceeds a 50% used percentage, the veto segments are defined by placing a time window (±1s) around the peak times of auxiliary channel triggers above the threshold. To follow convention, all the veto segment times are rounded (widened only) to integer second values. For the inspiral searches, the veto segments are often further padded with extra time windows, because large SNR glitches tend to trigger various templates and produce cascade of triggers that could last longer than our original veto window [7, 8].

The UPV code is written in Python, using libraries that the LSC and Virgo have developed to manipulate the data. To avoid memory problems due to a large number of triggers, the code uses the SQLite database and works on a local disk, instead of manipulating data on memory. Also, some computationally heavy calculations are done on C-based libraries for better performance. The code is run on the LIGO data grid computers [9].

The code defines vetoes online for S6/VSR2 on a weekly basis. The veto results are used to improve the search for GWs emitted from compact binary coalescence systems [10]. Fig. 3 and Table 1 shows UPV performance on single detector inspiral triggers for the first 3 months of S6/VSR2 for each LIGO and Virgo detector (the 4 km interferometer at Hanford is called H1, the 4 km interferometer at Livingston is called L1, and the 3 km Virgo interferometer is called V1). We see that UPV is helping to clean up high SNR "outliers".



(a) H1 (b) L1 (c) V1

Figure 3: Histograms of single detector inspiral triggers for the first 3 months of S6/VSR2 before and after UPV is applied. Each histogram corresponds to the detectors called H1, L1 and V1 respectively. We can see that UPV removes many "outliers".

The search for GW bursts uses a related veto method called "hveto" that searches for statistically significant auxiliary channels [11]. Even though hveto uses a different ranking statistics, the results agree well with UPV, providing assurance for our method.

	H1	L1	V1
Deadtime	0.604%	1.062%	0.475%
Efficiency, SNR> 8	23.0%	70.4%	42.1%
Efficiency, SNR> 10	42.4%	79.8%	51.6%
Efficiency, SNR> 20	58.2%	85.9%	40.1%

Table 1: Efficiencies of UPV on single inspiral triggers for the first 3 months of S6/VSR2.

2.1. Safety

Since the UPV is based on statistical measurements without regards to causality, we need to ensure that the auxiliary channels have negligible sensitivity to GWs. In particular, GW signals could couple into some auxiliary channels, leading us to veto GW signals rather than noise by finding coincidences between the GW channel and the auxiliary channels. Auxiliary channels with a non-negligible coupling from GW signals are called "unsafe". Some channels are already identified as unsafe from the past analysis and input from the commissioning team. We disregard these as veto channels *a priori.* (There are, however, some suggested veto methods to make use of such unsafe channels [13].)

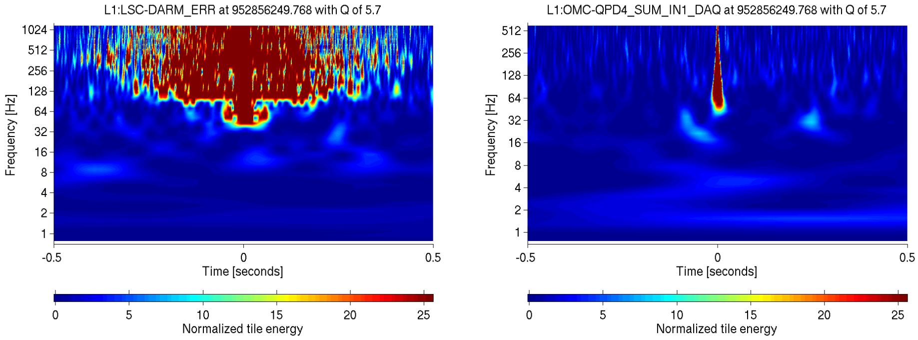
Since the detectors are highly complicated, there is always a possibility of unknown coupling mechanisms. Therefore, we take another safety precaution using simulated GW signals inserted into the data by physically shaking the mirrors (hardware injections). The number of false dismissals of hardware injections should be consistent with an accidental coincident rate calculated from the deadtime, and with a sufficient number of injections, we can establish the safety of vetoes. We calculate the probability that the channel would veto at least N injections via accidental coincidence:

$$\text{Safety Probability} \equiv 1 - F(N^{\text{inj}}_{\text{vetoed}} - 1; N^{\text{inj}}_{\text{exp}}) \tag{7}$$

where F is Poisson cumulative density function, $N^{\text{inj}}_{\text{vetoed}}$ is the number of injections actually vetoed, and $N^{\text{inj}}_{\text{exp}}$ is the expected number of injections to be vetoed from the deadtime assuming triggers are randomly distributed. We do not use channels that have a safety probability smaller than 10^{-5}. This value was chosen empirically; the past analysis has shown that statistical fluctuations would not give probability lower than 10^{-3} for channels we believe safe, while known unsafe channels usually show up with probability $< 10^{-8}$. Finally, all the veto candidate channels are reviewed by the LSC's detector characterization group and glitch group [12] for safety before being implemented as vetoes; Virgo vetoes are similarly reviewed by the Virgo data quality group.

3. Application to Detector Characterization

Even though the UPV was initially developed for defining vetoes, it can be a useful diagnostic tool for the detectors. For example, while we were following up the outcome of the UPV for the last science run (S5), we identified that some noise was due to sledge hammers used in construction work near the observatory. Motivated by this example, we further developed the code so that we could apply the UPV for detector characterization in a more systematic way. To apply the output of the UPV for detector characterization, we tuned some of the criteria differently from defining vetoes; we required a 30% used percentage instead of 50% and also required only 5 $N^{\mathrm{aux}}_{\mathrm{coinc}}$ in a day instead of 10 in a week, in order to allow more channels to be analyzed. We developed a postprocessing code that gathers all the coincident triggers from veto candidate channels, clusters the events with 1s window, reorders them according to the corresponding significance of the GW trigger, and creates a webpage for further human investigations for these events that have at least two channels in coincidence with the GW channel. The idea is to classify noise by corresponding channels, and analyze each noise group to identify the physical source. The code uses information from "data quality (DQ) flags" [4, 5, 7] and launches "Omega Scans" (formerly known as "Q Scans" [6]). DQ flags are defined and stored in a database to provide information on data quality. They mark noise with well-known origins, such as seismic noise from train or wind, as well as the state of the detectors, and thus provide useful information for the investigation. Omega Scan is a tool to perform a detailed study of the LIGO and Virgo data stream around a specific time of interest. In addition to displaying the time series and time-frequency spectrograms of the GW channel data, Omega Scan can also efficiently search a large number of auxiliary channels for statistically significant signal content (see Fig. 4 for an example of the Omega Scan). Both DQ flags and Omega Scans are linked to the webpage that the postprocessing code creates. The basic work flow of the UPV is shown in Fig. 5. The UPV and the postprocessing codes are launched every day, and the results are posted for further investigation.



(a) GW channel (b) Auxiliary channel

Figure 4: Examples of Omega Scans [6]. These are time-frequency representation of a particular glitch in the GW channel (a) and an auxiliary channel (b). UPV removed this kind of glitch using an output mode cleaner channel. We are using information obtained through the UPV process to track down the actual cause of glitches.

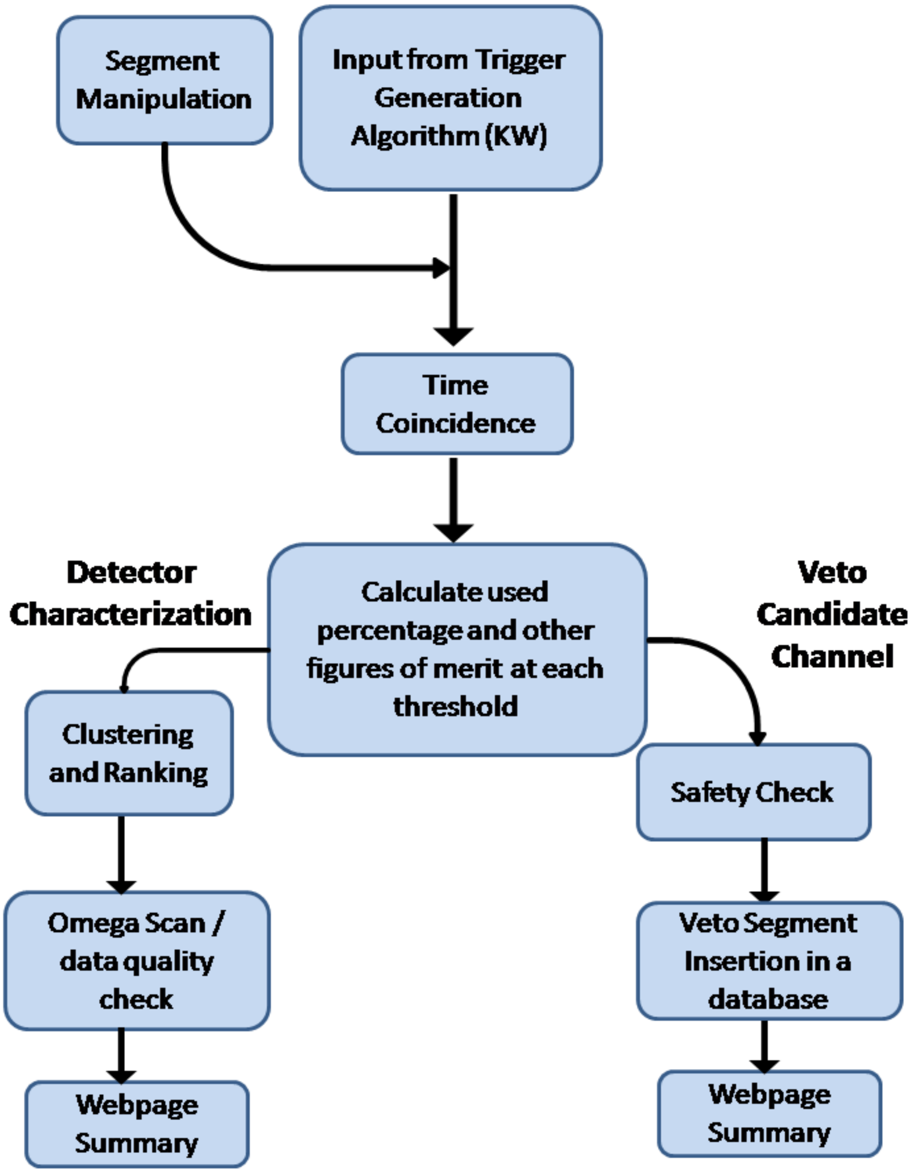


Figure 5: A diagram showing the workflow of UPV for a single auxiliary channel. Figures of merit including used percentage are calculated at each threshold after particularly noisy segments have been excluded. We define vetoes for channels with a strong correlation with the GW channel. In addition, we use information obtained from this process for detector characterization. This process is done on over 150 auxiliary channels for each of LIGO detectors and over 500 channels for the Virgo detector.

4. Summary

The UPV is contributing to the elimination of transients in LIGO and Virgo data that are difficult to remove by other methods. For the ongoing S6 and VSR2 science run, the

UPV is running week by week automatically, defining vetoes for binary inspiral searches, and helping to identify and eliminate a large number of problematic noise triggers.

In addition, the UPV is running every day for detector characterization. It is monitoring day-to-day changes of the detectors with respect to their auxiliary channels, and providing quick feedback about the malfunctioning of some components of the detectors. It also supplies useful information on identifying the physical origin of noises, thereby contributing to the improvement of the detectors.

5. Acknowledgments

This project is funded by NSF Grant PHY-0854790, Summer Undergraduate Research Fellowship and the Kolenkow-Reitz Fund at Carleton College. The authors gratefully acknowledge the support of the United States National Science Foundation for the construction and operation of the LIGO Laboratory, the Science and Technology Facilities Council of the United Kingdom, the Max-Planck-Society, and the State of Niedersachsen/Germany for support of the construction and operation of the GEO600 detector, and the Italian Istituto Nazionale di Fisica Nucleare and the French Centre National de la Recherche Scientifique for the construction and operation of the Virgo detector. The authors also gratefully acknowledge the support of the research by these agencies and by the Australian Research Council, the Council of Scientific and Industrial Research of India, the Istituto Nazionale di Fisica Nucleare of Italy, the Spanish Ministerio de Educacion y Ciencia, the Conselleria dEconomia Hisenda i Innovacio of the Govern de les Illes Balears, the Foundation for Fundamental Research on Matter supported by the Netherlands Organisation for Scientific Research, the Polish Ministry of Science and Higher Education, the FOCUS Programme of Foundation for Polish Science, the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, The National Aeronautics and Space Administration, the Carnegie Trust, the Leverhulme Trust, the David and Lucile Packard Foundation, the Research Corporation, and the Alfred P. Sloan Foundation.

Reference

[1] Abbott B *et al* (the LIGO Scientific Collaboration) 2009 LIGO: the Laser Interferometer Gravitational-Wave Observatory *Reports on Progress in Physics* **72** 076901

[2] Acernese F *et al* (the Virgo Collaboration) 2009 Status of Virgo *Class. Quantum Grav.* **25** 114045

[3] Grote H (the LIGO Scientific Collaboration) 2008 The Status of GEO 600 *Class. Quantum Grav.* **25** 114043

[4] Christensen N (for the LIGO Scientific Collaboration and the Virgo Collaboration) 2010 LIGO S6 Detector Characterization Studies submitted to these GWDAW14 proceedings, LIGO Document P1000045

[5] Robinet F (for the LIGO Scientic Collaboration and the Virgo Collaboration) 2010 Data Quality in Gravitational Wave Burst and Inspiral Searches in the Second Virgo Science Run submitted to these GWDAW14 proceedings, LIGO Document P1000047

[6] Chatterji S, Blackburn L, Martin G and E. Katsavounidis E 2004 Multiresolution techniques for the detection of gravitational-wave bursts *Class. Quantum Grav.* **21** S1809

[7] Slutsky J *et al* 2010 Methods for Reducing False Alarms in Searches for Compact Binary Coalescences in LIGO Data *arXiv:1004.0998*

[8] Christensen N, Shawhan P and Gonzalez G (for the LIGO Scientific Collaboration) 2004 *Class. Quantum Grav.* **21** S1747-S1755

[9] `https://www.lsc-group.phys.uwm.edu/lscdatagrid/overview.html`

[10] Abbott B *et al* (the LIGO Science Collaboration) 2009 Search for gravitational waves from low mass binary coalescences in the first year of LIGO psila s S5 data *Phys. Rev. D* **79** 122001

[11] Abbott B *et al* (the LIGO Science Collaboration) 2009 Search for gravitational-wave bursts in the first year of the fifth LIGO science run Phys. Rev. D **80**, 102001

[12] Blackburn L *et al* 2008 The LSC glitch group: monitoring noise transients during the fifth LIGO science run *Class. Quantum Grav.* **25** 184004

[13] Hild S *et al* 2007 A statistical veto method employing an amplitude consistency check *Class. Quantum Grav.* **24** 3783-3798