Strategy for signal classification to improve data quality for Advanced Detectors gravitational-wave searches

Elena Cuoco [12], Jade Powell[3], Alejandro Torres-Forné[4], Ryan Lynch[5], Daniele Trifirò[6] [7], Marco Cavaglià[7], Ik Siong Heng[3] and José A. Font[4] [8]

[3] *SUPA, Institute for Gravitational Research, School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, Scotland, United Kingdom.*
[4] *Departamento de Astronomía y Astrofísica, Universitat de València, Dr. Moliner 50, 46100, Burjassot (València), Spain.*
[5] *Massachusetts Institute of Technology, 185 Albany St, 02139 Cambridge USA*
[6] *Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa 56127, Italy.*
[7] *Department of Physics and Astronomy, The University of Mississippi, University, MS 38677, USA.*
[1] *European Gravitational Observatory (EGO), Via E Amaldi, I-56021 Cascina, Italy*
[2] *Istituto Nazionale di Fisica Nucleare (INFN) Sez. Pisa Edificio C - Largo B. Pontecorvo 3, 56127 Pisa, Italy.*
[4] *Observatori Astronòmic, Universitat de València, C/ Catedrático José Beltrán 2, 46980, Paterna (València), Spain.*

Summary. — Noise of non-astrophysical origin contaminates science data taken by the Advanced Laser Interferometer Gravitational-wave Observatory and Advanced Virgo gravitational-wave detectors. Characterization of instrumental and environmental noise transients has proven critical in identifying false positives in the first aLIGO observing run O1. In this talk, we present three algorithms designed for the automatic classification of non-astrophysical transients in advanced detectors. Principal Component Analysis for Transients (PCAT) and an adaptation of LAL-Inference Burst (PC-LIB) are based on Principal Component Analysis. The third algorithm is a combination of a glitch finder called Wavelet Detection Filter (WDF) and unsupervised machine learning techniques for classification.

1. – Introduction

The advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) detectors are two 4km interferometers at Hanford, Washington (H1) and Livingston, Louisiana (L1) [10, 1]. The Italian 3km interferometer Virgo is expected to join the advanced detector network early next year [3]. The detector duty cycle and sensitivity to astrophysical signals will be determined by noise sources created by the instruments and the environment. In particular, as the detector noise is non-Gaussian short-duration transients

will limit the sensitivity of searches for transient astrophysical sources such as compact binary coalescences [5].

The detectors contain many environmental and instrumental sensors, which produce auxiliary channels of data that can be used to monitor the detector behaviour and track the causes of short-duration noise artifacts. Auxiliary channels that are not sensitive to gravitational waves can be used to identify noise transients, also known as "glitches", in the detector output and veto those events [20, 4, 22]. Classification and categorization of transients using individual channels of data may provide valuable clues for the identification of their sources, which can aid in efforts to eliminate them [17, 14]. So far classification has mainly been achieved by visual inspection of spectrograms of the transients, but automatic classification is essential for future detections of astrophysical gravitational-wave signals.

Three methods for fast classification of transients have been developed for the analysis of aLIGO and Virgo data. They are Principal Component Analysis for Transients (PCAT), Principal Component LALInference Burst (PC-LIB) and Wavelet Detection Filter with Machine Learning (WDF-ML). We will report on the comparison of these 3 methods on simulated data as in [17]. The same methods have been successfully applied to data from the 7th aLIGO engineering run (ER7), which began on the 3rd of June 2015 and finished on the 14th of June 2015 [?].

2. – Transient classifying algorithms

Three different classifying algorithms were developed for the fast classification of noise transients in the detectors. Most of the technical details have been described in [17] and application on real data have been reported in [18]. Here we give a brief outline of the three methods and some examples of their classification efficiency on simulated and real data.

To find transients in the data we use event trigger generators (ETGs). ETGs typically search for excess power in individual interferometers and output the time, SNR, frequency, duration and other parameters of transients found in the data. PC-LIB uses Omicron, the main ETG used by the LIGO Scientific Collaboration's (LSC) detector characterization group [19, 8]. WDF-ML[2] and PCAT have their own internal ETGs.

$2^\cdot 1$. *PCAT*. – PCAT uses a technique called Principal Component Analysis (PCA) that allows for dimensional reduction of large data sets [21, 17]. In the first stage of the PCAT analysis, the data are downsampled to $8192\,\mathrm{Hz}$, whitened and high-pass filtered at $10\,\mathrm{Hz}$. Then PCA is applied to all of the noise transients found by the ETG in all the analyzed segments of data.

A projection of the original waveforms on to the Principal Components (PCs) allows for the calculation of scale factors for each PC called PC coefficients. Noise transients of different types are separated in the PC coefficient parameter space. This allows PCAT to classify the transients by applying a Gaussian Mixture Model (GMM) machine learning classifier to the PC coefficients [16].

$2^\cdot 2$. *PC-LIB*. – LALInference Burst (LIB) is a Bayesian parameter estimation and model selection tool, which uses a sine-Gaussian signal model to estimate parameters of gravitational-wave bursts [9]. It can also be combined with Omicron to be run as a search [13]. PC-LIB adapts LIB for the classification of transients by replacing the LIB sine-Gaussian signal model with a new signal model created from a linear combination of PCs

TABLE I. – *The table shows the LIB, PCAT and WDF-ML results for simulated data set. The values show the percentage of the different morphologies classified in each type. The total number of simulated waveforms was 1000 of each type. The total number of glitches analysed were 1309 for PCAT, 1452 for LIB and 1814 for WDF-ML.*

Classifier	SG	G
PCAT Type 1	99%	0%
PCAT Type 2	1%	100%
LIB Type 1	99.9%	5%
LIB Type 2	0.1%	95%
WDF Type 0	99.5%	2.4%
WDF Type 1	0.3%	46.1%
WDF Type 2	0.2%	51.5%

calculated from the waveforms of known transient types [11, 12]. These known transients may have been previously classified by examining spectrograms of the transients or by one of the other methods. Thus PC-LIB can only classify transients that have occurred in the data many times before. When transients of a new type start to appear in the data new signal models must be created.

2˙3. *WDF-ML*. – Wavelet detection filter (WDF) is the ETG [2] used by WDF-ML method. The data are firstly down-sampled and then whitened using parameters estimated at the beginning of each locked segment. After whitening, a wavelet-transform is applied, using a bank of wavelets, as described in [17].

The wavelet coefficients identified by the WDF-ML ETG are further cleaned using a wavelet de-noising procedure where only wavelet coefficients above the noise level are retained [17]. WDF-ML produces a list of wavelet coefficients, frequency, duration and SNR for each transient. The dimensions of the wavelet coefficients are then reduced by applying PCA and Spectral Embedding [15, 7]. The transient classification is then performed by applying a machine learning (ML) unsupervised algorithm, the GMM classifier, to the reduced wavelet coefficients [16].

3. – Results

3˙1. *Tests on simulated data*. – To test and compare these methods we create a simulated data set in aLIGO Gaussian noise [17]. As example, we report the results on data set containing 1000 sine Gaussian waveforms and 1000 Gaussian waveforms in simulated Gaussian noise. The sine Gaussian waveforms have a frequency = 400Hz and an SNR between 5 and 30. The Gaussian waveforms are centred at f = 0Hz and have an SNR between 20 and 250. Table I shows the % of detected transients that were classified in each type. A few low frequency SG, and low SNR G were in the incorrect classes, but the overall classification efficiency was very good.

3˙2. *Tests on real LIGO data*. – In [18] the above described methods have been applied on LIGO ER7 real data. Some examples of real glitches for ER7 LIGO Livingston have

4 ELENA CUOCO [1][2], JADE POWELL[3], ALEJANDRO TORRES-FORNÉ[4], RYAN LYNCH[5], DANIELE TRIFIRÒ[6][7], MARCO CAVAGLIÀ[7]
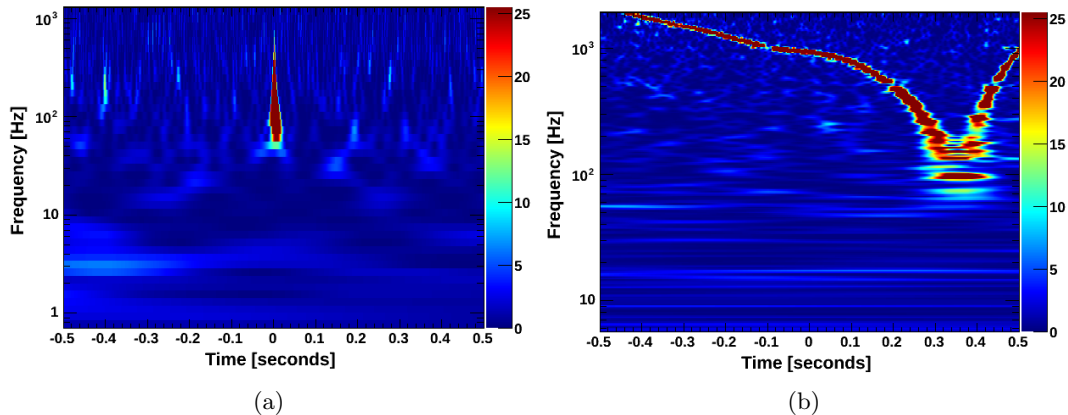
(a)

(b)

Fig. 1. – Spectrograms, generated using [6], of typical transient types found in the aLIGO Livingston(L1) ER7 data. (a) L1: A transient characterized by a tear drop shape in the spectrogram. (b)L1: A "whistle" glitch that often has a long duration and occurs at high frequencies.

been reported in figure **3**˙2.

Glitches of different types are often recognised by their shape in a spectrogram such as those shown in Figure 1. Figure 1(a) shows glitches characterized by a tear drop shape. Figure 1(b) shows longer duration transients known as "whistles", which are caused by radio frequency beats. In this studies all the glitches were labelled and classified "by hand". This classification is used as reference to check the efficiency of the automatic classification pipelines. In the ER7 data from aLIGO Livingston PCAT missed 90 transients and classified 95% of the remaining transients correctly. PC-LIB missed 33 transients and classified 98% of the remaining transients correctly. WDF-ML classified all transients and 97% of them were correct. In aLIGO Hanford PCAT missed 120 transients and classified 99% of the remaining transients correctly. PC-LIB missed 6 transients and classified 95% of the remaining transients correctly. WDF-ML classified all transients and 92% of them were correct. We conclude that our methods have a high efficiency also in real non-stationary and non-Gaussian detector noise.

4. – Conclusion

Three different methods have been developed for the fast classification of noise transients. Transients are split in to types by waveform morphology first, and then can be split up in to further types by frequency and SNR. These pipelines performed very well either on simulated realistic data on real data. Results are similar for all methods of classification. The WDF-ML acts also efficiently as Event Trigger Generator. We plan to introduce in the WDF-ML pipeline different machine learning methods than the unsupervised one used in these works. Further development and tests on real data have been done by all of the 3 methods.

REFERENCES

[1] J. Aasi et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, 2015.

[2] F. Acernese et al. Gravitational waves by gamma-ray bursts and the virgo detector: the case of grb 050915a. *Classical and Quantum Gravity*, 24(19):S671, 2007.

[3] F. Acernese et al. Status of Virgo. *Classical and Quantum Gravity*, 25(11):114045, June 2008.

[4] P. Ajith, T. Isogai, N. Christensen, R. X. Adhikari, A. B. Pearlman, A. Wein, A. J. Weinstein, and B. Yuan. Instrumental vetoes for transient gravitational-wave triggers using noise-coupling models: The bilinear-coupling veto. *PRD*, 89(12):122001, June 2014.

[5] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton. FINDCHIRP: an algorithm for detection of gravitational waves from inspiraling compact binaries. *arXiv.org*, (12):122006, September 2005.

[6] J. Areeda and J. Smith. LIGOdv-web: Providing easy, secure and universal access to a large distributed scientific data store for the LIGO Scientific Collaboration. 2016.

[7] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[8] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis. Multiresolution techniques for the detection of gravitational-wave bursts. *Classical and Quantum Gravity*, 21:S1809–S1818, October 2004.

[9] R. Essick, S. Vitale, E. Katsavounidis, G. Vedovato, and S. Klimenko. Localization of short duration gravitational-wave transients with the early advanced LIGO and Virgo detectors. *ArXiv e-prints*, September 2014.

[10] G. M. Harry and LIGO Scientific Collaboration. Advanced LIGO: the next generation of gravitational wave detectors. *Classical and Quantum Gravity*, 27(8):084006, April 2010.

[11] I. S. Heng. Rotating stellar core-collapse waveform decomposition: a principal component analysis approach. *Classical and Quantum Gravity*, 26(10):105005, May 2009.

[12] J. Logue, C. D. Ott, I. S. Heng, P. Kalmus, and J. H. C. Scargill. Inferring core-collapse supernova physics with gravitational waves. *PRD*, 86(4):044023, August 2012.

[13] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet. An information-theoretic approach to the gravitational-wave burst detection problem. *ArXiv e-prints*, November 2015.

[14] S Mukherjee, R Obaid, and B Matkarimov. Classification of glitch waveforms in gravitational wave detector characterization. *Journal of Physics: Conference Series*, 243(1):012006, 2010.

[15] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856. MIT Press, 2001.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglià. Classification methods for noise transients in advanced gravitational-wave detectors. *Classical and Quantum Gravity*, 32(21):215012, November 2015.

[18] Jade Powell, Alejandro Torres-Forn, Ryan Lynch, Daniele Trifir, Elena Cuoco, Marco Cavagli, Ik Siong Heng, and Jos A Font. Classification methods for noise transients in advanced gravitational-wave detectors ii: performance tests on advanced ligo data. *Classical and Quantum Gravity*, 34(3):034002, 2017.

[19] F. Robinet. Omicron: an algorithm to detect and characterize transient events in gravitational-wave detectors. 2015.

[20] J. R. Smith, T. Abbott, E. Hirose, N. Leroy, D. MacLeod, J. McIver, P. Saulson, and P. Shawhan. A hierarchical method for vetoing noise transients in gravitational-wave detectors. *Classical and Quantum Gravity*, 28(23):235005, December 2011.

[21] J. Friedm T. Hastie, R. Tibshirani. *The Elements of Statistical Learning*. Springer, 2001.

[22] The LIGO Scientific Collaboration, The Virgo Collaboration, J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, et al. Characterization of the LIGO detectors during their sixth science run. *ArXiv e-prints*, October 2014.