Astrophysical signal consistency test adapted for gravitational-wave transient searches

V. Gayathri,[1] P. Bacon,[2] A. Pai,[1] E. Chassande-Mottin,[2] F. Salemi,[3] and G. Vedovato[4]

[1]*Department of Physics, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India*
[2]*APC, Univ Paris Diderot, CNRS/IN2P3, CEA/Irfu,*
Obs. de Paris, Sorbonne Paris Cité, F-75013 Paris, France
[3]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-30167 Hannover, Germany*
[4]*INFN, Sezione di Padova, I-35131 Padova, Italy*
(Dated: July 26, 2019)

Gravitational wave astronomy is established with direct observation of gravitational wave from merging binary black holes and binary neutron stars during the first and second observing run of LIGO and Virgo detectors. The gravitational-wave transient searches mainly categories into two families: modeled and modeled-independent searches. The modeled searches are based on matched filtering techniques and model-independent searches are based on the extraction of excess power from time-frequency representations. We have proposed a hybrid method, called wavegraph that mixes the two approaches. It uses astrophysical information at the extraction stage of model-independent search using a mathematical graph. In this work, we assess the performance of wavegraph clustering in real LIGO and Virgo noises (the sixth science run and the first observing run) and using the coherent WaveBurst transient search as a backbone. Further, we propose a new signal consistency test for this algorithm. This test uses the amplitude profile information to distinguish between the gravitational wave transients from the noisy glitches. This test is able to remove a large fraction of loud glitches, which thus results in additional overall sensitivity in the context of searches for binary black-hole mergers in the low-mass range.

PACS numbers: 04.80.Nn, 07.05.Kf, 95.55.Ym

I. INTRODUCTION

Gravitational-wave (GW) astronomy began with the first direct detection by the two Advanced LIGO detectors (Hanford and Livingston, US) [1] of the signal from binary black hole (BBH) merger on September 14th 2015 [2]. During their first and second observing runs, the LIGO detectors have further observed jointly with the Virgo detector [3] (Pisa, Italy) ten BBH merger signals [2, 4–9] and one binary neutron star merger [10]. The third observing run started on April 1st, 2019. The LIGO and Virgo collaborations have, since, reported the detection of several compact binary candidates. During the next ten years the LIGO and Virgo detectors will continue their observations while gradually approaching their design sensitivity. KAGRA (Japan) [11] and a third LIGO instrument in India [12] are expected to join the network of observing detectors end of 2019 and around 2025, respectively.

Compact binary coalescences (CBC) composed of NS and/or stellar-mass BHs are primary sources for ground based detectors. The GW signal is buried in the detector noise. To detect the GW transient signals, different search methodologies have been developed, that can be categorized into two types: model-based and model-independent searches.

The model-based search, also referred to as matched filtering search aim to in find the best fitting waveform in a discrete set of physical model waveforms, called *templates*. While this type of search is powerful, it has two limitations. It can be computationally expensive when the number of templates to compare with is large. In the second LIGO-Virgo observing run, approximately 400,000 template waveforms are obtained assuming spin-aligned, non-precessing compact coalescing binaries [13–17]. These number of templates increases under more general assumptions. For instance, it increases by a factor of 10 when considering precessing binaries with arbitrarily aligned spins [18, 19]. This type of search also relies on the availability of accurate waveform models. This is not true in all areas of the signal parameter space, in particular for precessing or eccentric binary systems [18, 20–23] where waveform development is still a field of active research.

Model-independent searches assume minimum prior information about the waveform morphology [24–26]. They are based on the extraction of excess power from time-frequency representations (TF maps) of the multi-detector observations. While computationally much cheaper, this type of search does not perform as good as the model-based search to detect CBC signals [9, 27], primarily because they get affected by non-Gaussian and non-stationary features, also called "glitches", present in the detector noise [28].

In an earlier work [29], we proposed a hybrid method, called *wavegraph* that mixes the two approaches. This method is a search for time-frequency patterns defined from a range of astrophysical waveform models. The set of time-frequency patterns are mapped to a graph data structure, which allows to perform the pattern search very efficiently using combinatorial optimization algorithm. This method was implemented into the model-independent search algorithm Coherent Wave-Burst (cWB) [30–32]. We showed using simulated Gaussian noise that wavegraph improves the sensitive range of

cWB to BBH signals by 5% to 13% depending on the total mass of the binary.

In this work, we assess the performance of wavegraph in real noise, including glitches. We propose a signal consistency test for this algorithm, similar to the glitch rejection techniques introduced in model-based searches, such as χ^2 test [33–35], bank χ^2 [36] or ξ^2 tests [15]. We demonstrate and validate this method in the context of BBH search using LIGO S6 and O1 data.

In Sec. II A, we briefly summarize the coherent waveburst algorithm as well as the wavelet based, graph theoretical based clustering scheme developed in II B. In Sec III. we propose the signal consistency test for wavegraph clustering. In Sec. IV and V, we present the results of the simulation.

II. COHERENT WAVEBURST AND WAVEGRAPH ALGORITHMS

In this section, we give a brief overview of the coherent WaveBurst and wavegraph algorithms.

A. Coherent WaveBurst algorithm

The cWB algorithm is used to search for a broad class of weakly modelled GW transients in multi-detector data with minimum prior knowledge of the targetted signal. The algorithm has been successfully applied to eyes-wide-open, all-sky, and all-time searches for short-duration GW transients in the data from Advanced LIGO and Virgo observing runs [37–42]. We now highlight the main features of the algorithm.

A time-frequency map is obtained using the Wilson-Daubechies-Meyer (WDM) transform [43] that projects the detector data onto bases of functions localized in the time-frequency domain Multiple bases are considered with functions that span a range of durations defined by the scale parameter M. We denote $w_k(t, f; M)$, the WDM transform data for the k-th detector in the multi-detector network.

The data for the whole network is collected in a vector $\mathbf{w}_{\theta,\phi}[p] \equiv \{w_k(t - \tau_k(\theta, \phi), f; M)/\sqrt{S_k(f)}\}_{k=1...K}$ where $S_k(f)$ is the noise power spectrum from the k-th detector and here we use the short-hand notation p for the time, frequency and scale coordinates. This vector depends on the sky coordinates θ, ϕ because we compensated for the propagation delay $\tau_k(\theta, \phi)$ between the detector k and a reference, assuming the source is located in (θ, ϕ). The pixel energy $E_{\theta,\phi}[p] = \|\mathbf{w}_{\theta,\phi}[p]\|_2^2$ is maximized over (θ, ϕ) and the collection of pixels with large maximum energy is retained [30, 31].

Based on the salient pixels, the cWB algorithm forms clusters of neighboring pixels. Various clustering rules can be used based on the pixel geometrical vicinity in the time-frequency domain. The wavegraph method described in the next section follows a different approach

based on the information inferred from a set of astrophysical waveform models.

The cWB algorithm then evaluates the significance of each of these clusters using a likelihood ratio under Gaussian noise assumptions. The log-likelihood ratio reads $L(\theta, \phi) = 2(\mathbf{w}|\boldsymbol{\xi}) - (\boldsymbol{\xi}|\boldsymbol{\xi})$ where $\boldsymbol{\xi}$ is the noise-scaled network response to the impinging GW signal h_+ and $h_\times$ computed in the time-frequency domain. The components of this vector are $\xi_k[p] = [F_{k,+}h_+(t, f; M) + F_{k,\times}h_\times(t, f; M)]/\sqrt{S_k(f)}$ with $F_{k,+}(\theta, \phi)$ and $F_{k,\times}(\theta, \phi)$ are the antenna pattern function for the $+$ and $\times$ polarizations of the k-th detector. By maximizing over h_+ and $h_\times$, we obtain the statistic

$$L_{max}(\theta, \phi) = \sum_{p \in C} \mathbf{w}_{\theta,\phi}[p]^T \mathbf{P}_{\theta,\phi}[p] \mathbf{w}_{\theta,\phi}[p], \qquad (1)$$

which is used to select significant cluster as an event candidate.

This statistic is divided into two parts $L_{max} = E_{in} + E_{coh}$, where E_{in} and E_{coh} are the incoherent and coherent energies respectively which capture the diagonal and off-diagonal terms of the network projection operator $\mathbf{P}_{\theta,\phi}[p]$. The null energy $E_{null} = E_{tot} - L_{max}$ with $E_{tot} = \sum_{p \in C} E[p]$ captures the energy in the plane orthogonal to the network plane.

The network correlation coefficient [30] $c_c \equiv E_{coh}/(|E_{coh}| + E_{null})$ is used to distinguish the GW signal from the spurious glitches. The transient noise events are incoherent in phase between the detectors. They are thus detected with a lower coherent energy and a higher null energy as reconstructed detector response is inconsistent between the detectors. In the case of GW signals, we instead expect higher coherent energy and lower null energy (consistent reconstructed detector response between the detectors). As a result, GW signals have $c_c \approx 1$ and spurious events have $c_c \ll 1$. Simulations show that for low-SNR ($\lesssim 7$) events, we have $c_c \lesssim 0.7$.

The main ranking statistic used by cWB is [2],

$$\eta_c = \sqrt{c_c E_{coh} K/(K-1)}. \qquad (2)$$

This statistic is approximately proportional to the overall network SNR.

B. Wavegraph clustering algorithm

In [29], we proposed a new approach to form clusters of time-frequency pixels in model-independent searches such as cWB.

The method learns from a set of representative waveforms, what are the expected salient pixels, and how they are distributed or aligned in the time-frequency plane. The set of reference waveform can be obtained from a bank of regularly or randomly-placed templates (see e.g., [44, 45]), or using an ad-hoc parameter-space discretization.

The salient pixels are extracted from the redundant time-frequency signal decompositions given by the WDM transforms using the matching-pursuit algorithm [46]. The extracted pixels for the full set of reference signals are organized in a mathematical graph where neighbours in the set of pixels extracted from a given signal are connected by edges. Pixel neighborship is defined by a rule that tidy the pixels up in an ordered series. The graph includes the position of the selected pixels, *i.e.* time, frequency, and scale and their connection with other pixels.

The graph G is used while analyzing the detector data with cWB at the stage where time-frequency clusters are formed. Instead of using generic geometrical clustering techniques, the selected cluster C^* is obtained from the following optimization problem:

$$C^\star = \arg\max_{C \in G} \sum_{p \in C} E[p] - \lambda \bar{E}(f_p, M_p), \qquad (3)$$

where $E[p] = \max_{\theta,\phi} E_{\theta,\phi}[p]$ is a proxy of the incoherent energy and $\bar{E}(f, M)$ is the median value of $E[p]$ at frequency f and scale M over all times t.

The selected cluster therefore maximizes the total incoherent energy in the cluster after removing the average contribution due to (stationary) noise. The second term in Eq. (3) can also be viewed as a penalization, whose strength is defined by λ, that promotes smaller clusters (see [29] for more information). The selected cluster is then processed following the cWB likelihood analysis.

The wavegraph algorithm combined with cWB was tested in [29] to search for BBH signals (total mass ranging from 10 $M_\odot$ to 70 $M_\odot$) in simulated Gaussian noise. Overall, this scheme shows a relative improvement of $22 - 26\%$ in the event rate recovery with respect to cWB alone.

III. SIGNAL CONSISTENCY TEST FOR WAVEGRAPH

In our previous study using simulated Gaussian noise [29], we have shown that cWB combined with wavegraph has larger noise background than the cWB. With real detector noise, the noise background is expected to increase due to non-Gaussian noise transients in the detector.

Model-based searches use "signal consistency tests" to reject the event candidates identified by the matched filtering algorithm due to the noise transients. The different versions of such test include the χ^2 test [33–35], bank χ^2 [36] or ξ^2 tests [15]..

The χ^2 signal consistency test was first introduced in [33] and is based on a χ^2 statistics that checks the consistency of the frequency-domain amplitude profile of the candidate signal with that of the best matching template. Inspired by this principle, here we develop and validate a signal consistency test adapted to wavegraph.

A. Consistency test

The GW signal is a frequency as well as amplitude modulated signal. The wavegraph clusters carry the information of the frequency modulation in terms of the location of the TF pixels and their connection. In addition to the pixel location, the pixel also stores the amplitude modulation information in terms of the associated energy. We use this information in the consistency test to distinguish between the noisy transients and the GW transients as explained below.

We consider a cluster of pixels C with amplitude for each pixel p as $A[p]$ extracted from the data by wavegraph in association to a GW signal. Let us assume that we have the pixel amplitude model $w_s[p]$ normalized to unit norm i.e., $\sum_{p \in C} w_s[p] = 1$ for the (noise-free) template signal. The pixel amplitudes as observed in the noise-scaled data differ solely by an overall factor a :

$$A_s[p] = a \, w_s[p] \,. \qquad (4)$$

We propose a test based on the mean-square deviation between the observed and model amplitudes

$$X(a) = \frac{1}{|C|} \sum_{p \in C} \left(A[p] - A_s[p] \right)^2, \qquad (5)$$

where $|C|$ is the size of C.

Minimization of X wrt a given

$$X(\hat{a}) = \frac{1}{|C|} \sum_{p \in C} \left(A[p] - \hat{a} w_s[p] \right)^2 \qquad (6)$$

where $\hat{a}$ minimizes $X(a)$, such that

$$\hat{a} = \frac{\sum_{p \in C} A[p] w_s[p]}{\sum_{p \in C} w_s^2[p]}. \qquad (7)$$

Unfortunately, a pixel in the graph is not necessarily associated with only one template signal. Pixels generally belong to multiple template clusters. As a result, a pixel amplitude can take a range of model amplitudes $w_s[p]$. The range of $w_s[p]$ can be characterized by its mean $\bar{w}_s[p]$ and variance $\sigma_s^2[p]$. The variance is likely to increase with the number of template clusters passing through the considered pixel. We thus use $\bar{w}_s[p]$ in place of $w_s[p]$ in the χ^2 test defined in Eq. (6) and down-weight pixels with large variance in the sum defined in the next subsection.

B. Bias correction

The pixel amplitude $A[p] \equiv \sqrt{E[p]}$ combines a signal term (if present) and a noise contribution. The latter creates a bias with respect to the expected value A_s. Since $E[p] = \max_{\theta,\phi} E_{\theta,\phi}[p]$, the pixel amplitude results from a maximization over sky positions selected on a sky

grid. The result of this maximization varies depending on the duration of the Wilson-Daubechies-Meyer basis functions compared to the total span for the propagation delay over the sky. This leads to an average bias μ_{grid} and variance σ^2_{grid} that varies with the scale parameter M, as shown in Figure 1. In order to produce the estimates in Figure 1, we use segments (~ 600 seconds) of LIGO O1 data with reasonably stationary noise (no glitch). This simulation provides a set of estimated bias and variance at each scale. Both those quantities decrease with the scale, since, for larger scales, the propagation delay is small compared to the typical duration of the WDM functions, resulting in a smaller change of the WDM amplitudes.
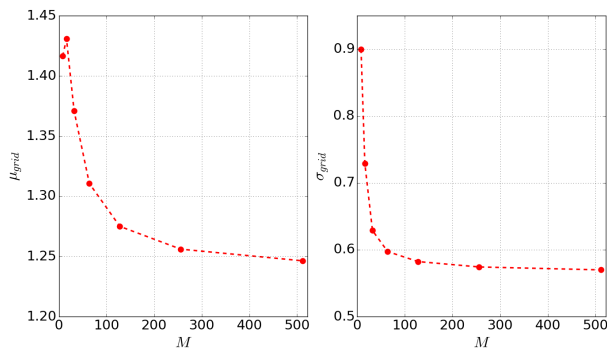
Figure. 1. Bias μ_{grid} and standard deviation σ_{grid} of the observed pixel amplitude vs. the scale parameter M, estimated using LIGO O1 data.

The observed amplitude is corrected from the bias, and the noise variability is included to the χ^2 statistics, leading to

$$\xi = \frac{1}{|C|} \sum_{p \in C} \frac{(A[p] - \mu_{grid}[p] - \hat{a}\ \bar{w}_s[p])^2}{\sigma^2_{grid}[p] + \hat{a}^2 \sigma^2_s[p]}. \tag{8}$$

For a GW signal with sufficient signal-to-noise ratio, we expect the statistic ξ to be small, while, for the noise glitches, the large amplitude discrepancy leads to large ξ. We term this as a signal consistency test which is a measure of consistency of the data to the signal model. We expect that longer is the signal duration, more effective will be the test as it combine the residual power for more number of pixels.

C.   Validation

In this subsection, we validate the proposed consistency test ξ using simulated CBC signals in real LIGO noise. We consider 16 days of data from sixth LIGO science run (S6, 2009-2010 [47]) and 48.6 days of data from the first advanced LIGO observing run (O1). S6 data is particularly suited to carry out any noise rejection technique as is include a large population of noise features.

A graph is generated using a set of CBC waveforms with the SEOBNRv2 model [48, 49] that spans a total mass range between $10 - 50\ M_\odot$ with mass ratio up to 3. We use 4 different scales ranging in $M = 2^3$ to 2^7 and a sampling frequency of 1024 Hz.

The data is analyzed with cWB with wavegraph using this graph. We estimate the analysis background using the time-slide technique[1]. We accumulate ~ 509 years and ~ 156 years of background data from O1 and S6 data, respectively. We compute the consistency statistic ξ for each cluster from this background data.

In Figure 2, we show the results for all the noise events from the background data (O1 noise in dotted orange and S6 noise in solid blue), and we compare it to events obtained from CBC injections (green dashed). The left panel shows the distribution of ξ/η_c and the right panel shows a scatter plot of ξ/η_c against η_c.

Clearly, the S6 data shows a larger background than O1 data. As expected, we show that the CBC injection events have low ξ/η_c values compared to noise events irrespective of its η_c value. At the same time, loud noise events (high η_c) have high ξ/η_c value. The large fraction of S6 noise events have ξ/η_c value higher than 2. These values are much higher than that of CBC events. We can thus reject the loud glitches by setting a threshold on ξ/η_c value. For example setting $\xi/\eta_c = 2$, we reject a large fraction of the noisy glitches and lose only few CBC signals.

IV.   ANALYSIS

In this section, we study the performance of the cWB combined with wavegraph and the signal consistency test introduced in Sec. III A. To do so, we analyse simulated GW signals from BBH mergers in the LIGO O1 noise.

A.   Parameter space

In this study, we consider three distinct regions of BBH systems, referring them to as R_1, R_2 and R_3 for compactness as tabulated in Table I which includes columns for masses, mass ratio, number of waveforms used in the template bank and number of nodes in the graph.

We consider BH with spins aligned with the orbital momentum and with $|\chi_{1,2}| = 0$ to 0.989. We compute a template bank [45] using the SEOBNRv2_ROM_DoubleSpin waveform approximant and a minimum match [2] of 0.97

———

[1] The detector data are artificially shifted by nonphysical time-delays ($\gtrsim 1\ s$) much larger than the physical wave propagation delay between detectors. This allows to estimate the chance probability of noise transients mimicking a GW signal in coincidence at the two detectors [2].

[2] The minimum match defines the minimum overlap between an arbitrary signal in the considered range and the waveforms in the
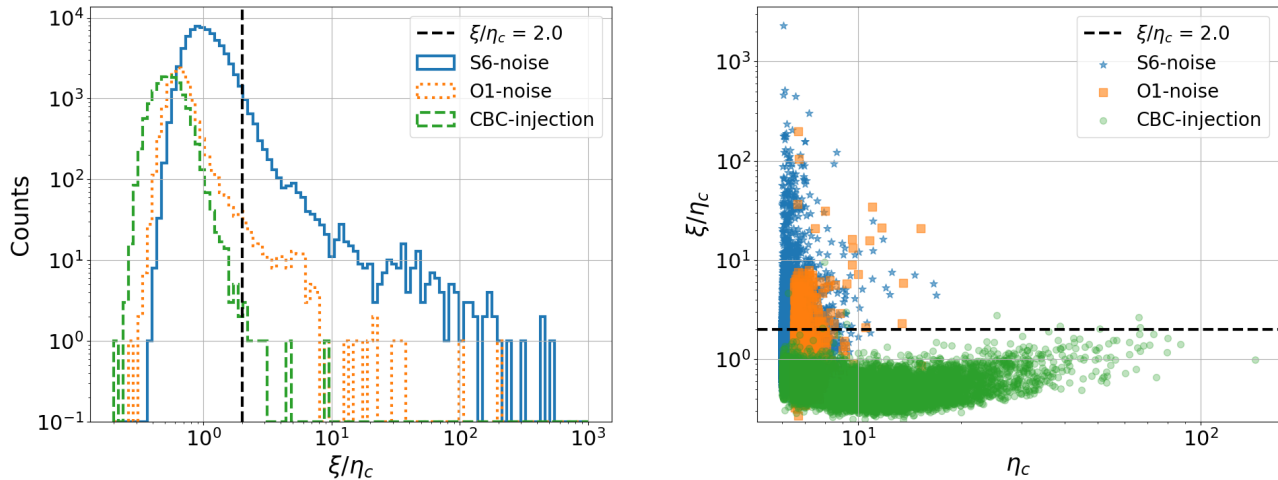
Figure. 2. The ξ/η_c distribution for selected O1 noise, S6 noise and CBC injections. The panel (a) shows the distribution of consistency statistics ξ and panel (b) shows the scatter plot of ξ/η_c vs η_c. The dashed black line in both plots shows the proposed ξ/η_c threshold.

for low-mass R_1, R_2 regions and 0.99 for R_3 region such that we get a reasonable number of templates in each region.

We generate a wavelet graph for each of the R_1, R_2, and R_3 regions from the associated template banks using the LIGO O1 power spectral density. The time-frequency pixels in the graph are selected to get a signal recovery > 80% [29]. Figure 3 shows the wavelet graphs for the R_1 (left), R_2 (middle), and R_3 (right) parameter spaces. The figure displays the location of the selected wavelets and the colors indicate the model amplitude $\bar{w}_s$. The variability and range of $\bar{w}_s$ increases with larger masses; *i.e.* moving from R_1 to R_3.

The number of time-frequency pixels in the graph decreases with the total mass; following the trend of the number of templates in the bank. Their distribution also changes based on the mass. The R_3 graph carries more time-frequency pixels at low scales than at high scales. This picture reverses for R_1 region. This is primarily because high mass BBH are short duration signals which are efficiently recovered with short duration wavelets. All in all, the connectivity of the graphs is with 80 to 150 connections between the graph nodes.

B. Estimation of the noise background

A critical part of any GW search algorithm is to measure the distribution of accidental triggers due to the noise of the detector. We estimate the false alarm rate

---

template bank. A minimum match of 0.97 decides the density of template bank and corresponds to an overall event loss of 10%.

BBH region	m_1, m_2 in $M_\odot$	q	No. of templates	No. of wavelets
R_1	5.0 to 12.5	< 3	11829	1333
R_2	12.5 to 20.0	< 2	2546	700
R_3	20.0 to 35.0	< 2	637	619

TABLE I: Details of three distinct regions in BBH parameter space. We tabulate the component masses m_1 and m_2, the mass ratio q, number of templates and the number of wavelets in the constructed graph.
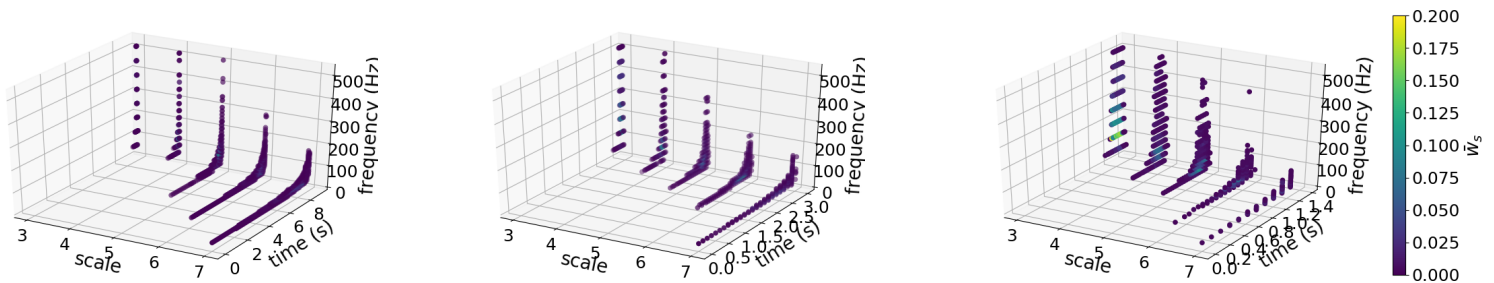
(FAR), *i.e.* the rate of accidental triggers, for the coherent WaveBurst algorithm with and without wavegraph using LIGO O1 noise.

The cWB algorithm includes a series of vetoes to reject noise transients in the data. Those vetoes are discussed in details in Appendix A. The noise background curve in terms of FAR vs η_c for cWB combined with wavegraph are shown in Figure 4 without applying the noise vetoes and in Figure 5 with the noise vetoes. We compare the noise backgrounds of cWB and cWB with wavegraph using the three graphs presented earlier (dashed, dotted and solid curves are for the R_1, R_2, and R_3 graphs) and the signal consistency test described in Sec. III accepting all the events below $\xi/\eta_c = 2$.

In Figure 4 the cyan (black) curves are for cWB with wavegraph without (with) signal consistency test (labelled CT). We observe that the long tails of the noise background curve (cyan) for cWB with wavegraph get trimmed after applying the signal consistency condition of $\xi/\eta_c = 2$ (black). In the R_1 region, we are able to

Figure. 3. Wavelet graph for three BBH parameter space generated using wavegraph algorithm. Left panel for R_1 space, middle panel for R_2 space and right panel for R_3 space. The color shows the $\bar{w}_s$ for selected wavelets.
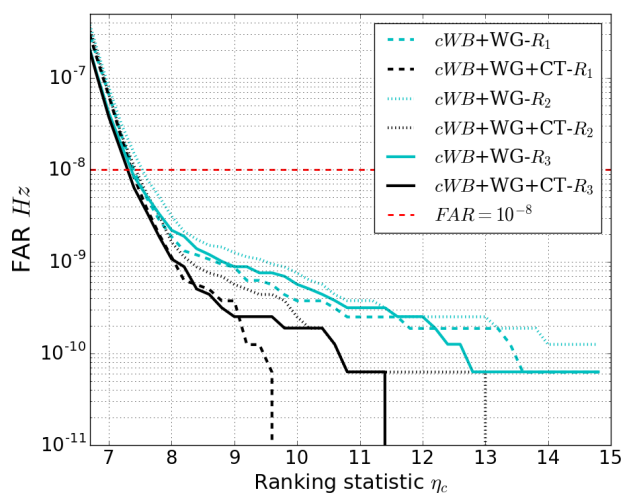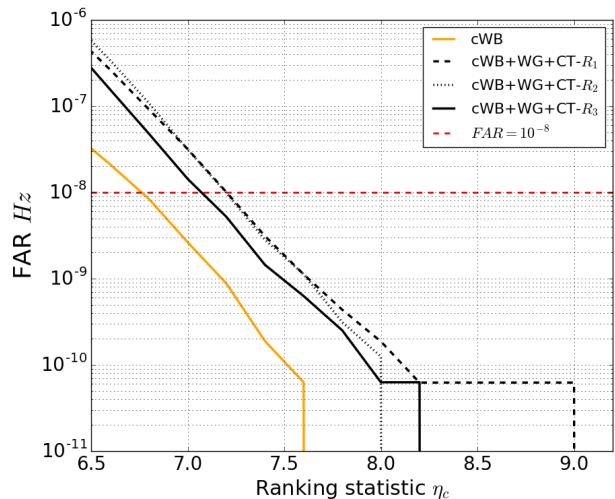

Figure. 4. Background False Alarm Rate (FAR) *vs* the statistic η_c. No noise vetoes has been applied here, besides the signal consistency test described in Sec. III. The solid, dotted and dashed curves correspond to R_3, R_2 and R_1 regions respectively. The lines in cyan correspond to *cWB* with wavegraph (WG) while the lines in black corresponds to *cWB* with WG and consistency test (CT). The horizontal dashed-red line shows the reference FAR level of 10^{-8} Hz (0.3 event per yr).

Figure. 5. Background False Alarm Rate (FAR) *vs* the statistic η_c. Noise vetoes discussed in Appendix A have been applied. The solid, dotted and dashed curves correspond to R_3, R_2 and R_1 regions. The orange (resp. black) color is for cWB (resp. cWB with wavegraph and consistency). The horizontal dashed-red line indicates the reference FAR $= 10^{-8}$ Hz (0.3 event per yr).

remove more glitches. They can be disentangled more efficiently from the signal as the latter has a longer duration.

In Figure 5 the solid orange, solid black, dotted black and dashed black lines are for cWB, cWB including wavegraph and the consistency test with R_3, R_2 and R_1 respectively. From this calculation, we compute the threshold on η_c associated to the reference FAR value of 10^{-8} Hz (0.3 events per yr). We obtain $\eta_c = 6.76, 7.07, 7.19$ and 7.2 for cWB, cWB including wavegraph and consistency test for R_3, R_2 and R_1 respectively.

It appears that cWB including wavegraph and consis-

tency test has a larger background compared to cWB. Nevertheless, as we will see in the next section, owing to the better recovery of the correlation coefficient, this algorithm shows an improved overall sensitivity.

C. Simulations

We now estimate the sensitivity of cWB with wavegraph algorithm using simulated BBH signals in the LIGO O1 noise covering the R_1, R_2 and R_3 parameters space as defined in Table I. Injections are generated using the SEOBNRv2 waveform model and are uniformly distributed over the binary parameter space (m_1, m_2, spins, sky-location, $\cos \iota$), and distributed over uniform volume up

to a fixed maximum distance: $R1$ is up to 1 Gpc, $R2$ is up to 1.5 Gpc and $R3$ is up to 3 Gpc.

Figure 6 shows the cumulative distribution of the network correlation value for recovered events at given FAR threshold of 10^{-8} Hz. The orange (resp. black) curve is for cWB (resp. cWB including wavegraph and consistency test). The solid, dotted and dashed lines are for the R_3, R_2 and R_1 regions respectively.
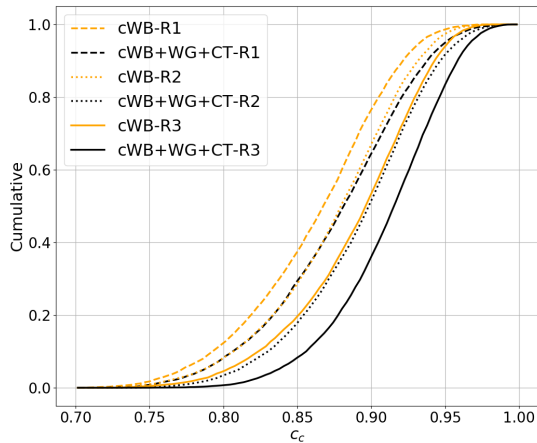

Figure. 6. Cumulative distribution of the network correlation for the recovered events. The orange (black) line is for cWB (cWB+WG+CT). The solid, dotted and dashed lines are for the R_3, R_2 and R_1 regions respectively.

We observe that cWB with wavegraph recovers injections with higher correlation coefficient c_c. This effect is more pronounced for the high-mass region. As a result, the events recovered by cWB with wavegraph have higher network correlation than the events recovered by cWB alone. That is primarily due to the wavegraph clustering method which helps to collect more coherent pixels as well as the consistency test which rejects inconsistent noisy triggers. This feature also plays an important role in removing the noisy events from the analysis background.

Figure 7 shows the detection efficiency curve for cWB as well as cWB with wavegraph algorithm in three different BBH parameter space. cWB with wavegraph and the consistency test recovers slightly more events in the low-mass range than in the high mass range compared to cWB. However, what is more important is that due to different clustering methods, amongst all the recovered events 62%, 83% and 85% of the events are common in R_1, R_2 and R_3 regions respectively. With cWB including wavegraph and consistency test, we recover additional 38%, 17% and 15% events from R_1, R_2 and R_3 regions respectively.
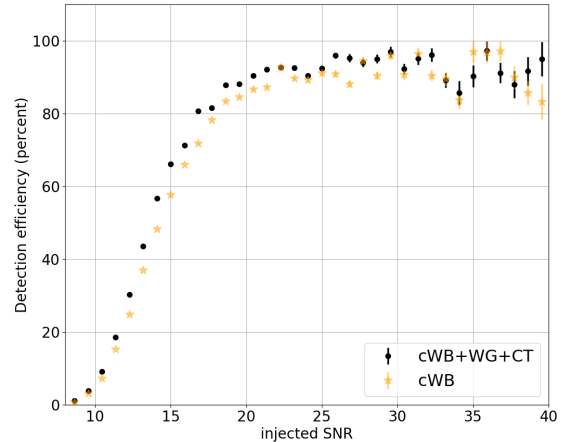
(a) R1 region
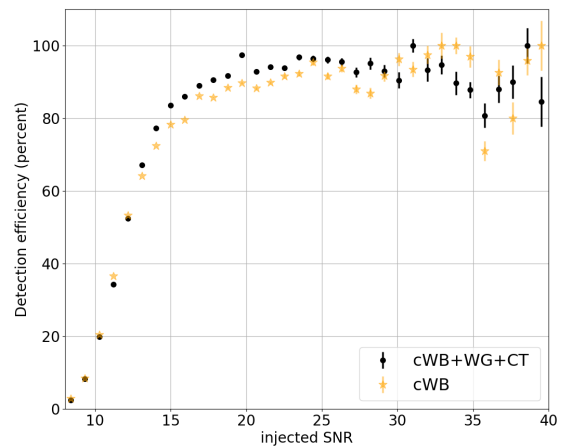
(b) R2 region
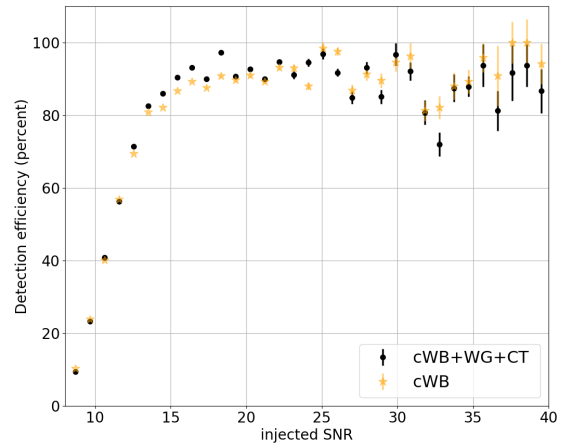
(c) R3 region

Figure. 7. Detection efficiency (in percent) vs the injected network SNR given for cWB (orange stars) and for cWB with wavegraph (black circles) with 1-sigma error bars. The (a), (b) and (c) panels correspond to the R_1, R_2 and R_3 simulations respectively.

V. CONCLUSION

We have shown in [29] that the sensitivity of cWB search can be improved by incorporating the astrophysical information at the clustering stage. The clustering uses a mathematical graph to store the astrophysical information in the form of TF pixels and their connection. We tested this method in the cWB algorithm for the GW signals emitted by BBH systems with simulated Gaussian noise using advanced two LIGO detectors and Virgo detector network.

In this work, we have proposed a new signal consistency test ξ for cWB with wavegraph algorithm. This test uses the amplitude profile information to distinguish between the GW transients from the noisy glitches. For real GW signals, the consistency test value ξ is small compared to the noisy glitches. We use this information to veto noisy glitches. We applied and tested cWB with wavegraph along with signal consistency test for simulated BBH GW signals in O1 data. We test this algorithm in three different non-precessing BBH parameter space. We observe that using cWB including the wavegraph as well as signal consistency test, we remove a large fraction of loud glitches in O1 data.

The cWB and cWB including wavegraph and consistency test recovers similar fraction of the events however, out of those 0.62, 0.83, 0.85 fraction of events are common in R_1, R_2 and R_3 regions. This clearly implies that the union of both bring in 0.38, 0.17, 0.15 fraction of additional events in R_1, R_2 and R_3 region respectively.

In summary, the wavegraph clustering approach along with the consistency test is general enough that it can be extended to a broad range of astrophysical systems which emit long duration GW signals in the ground based detectors. We are exploring this approach for detection of GW from long duration bursts signals such as eccentric binary black holes, accreting BH systems etc.

VI. ACKNOWLEDGEMENTS

Appendix A: Noise veto methods for coherent WaveBurst algorithm

For completeness, we review the glitch rejection methods included in the coherent WaveBurst algorithm and used in the context of transient searches in the LIGO/Virgo O1 and O2 data sets.

1. Norm

The frequency of CBC chirp signal evolves with time. The chirping signal energy is distributed over multiple TF maps and each part of the signal is captured by a specific wavelet. The signal is represented by the collection of wavelets from different scales/levels. The frequency change observed for noise glitches differs from the GW signal. Most of the loud glitches are localized in the narrow frequency range and thus represented by a collection of wavelets with few levels. cWB uses this information to separate glitches from astrophysical CBC signal. The E_{ft} is the time-frequency energy of the event in wavelet domain and E_t is the reconstructed event energy in time domain.

The average number of WDM resolutions used for event reconstruction is estimated by the oversampling factor $N_{norm} = E_{ft}/E_t$. The N_{norm} value strongly depends on the event SNR. The low SNR events are typically reconstructed with $N_{norm} \sim 1$ whereas very high SNR events are reconstructed with N_{norm} equal to two times of the number of WDM resolution used for the analysis.

2. Noise statistics, ζ

For real GW signal, the noise energy in the data after subtracting reconstructed signal from the strain data is low. For the noisy glitch, high residual noise energy remains in the data. The cWB algorithm uses this information to discriminate the real GW event from the noisy glitches. The residual noise energy in the data E_n is estimated as sum of residual noise energy E_r and energy contribution from Gaussian noise remaining in the data E_g. The noise statistics $\zeta^2 = E_n/N_{Dof} = (E_r + E_g)/N_{Dof}$. Here, N_{Dof} is the number of degrees of freedom give by number of detector times number of pixels. The ζ^2 is

estimated in both time and time-frequency domain. The GW signals have a ζ^2 value below 1. A significant deviation from unity is an indication of presence of a glitch. By applying a threshold on ζ^2 value, cWB separates the glitch from the real GW signal.

3. Qveto

"Blip glitches"[51] are a type of short duration noisy glitches in LIGO data of unknown origin and show raindrop like structure in TF plane. The blip glitch in the TF plane is well-localized and maximum energy of the glitch event is derived from a small number of pixels. However, in case of CBC events, the energy is distributed over all the pixels. cWB has developed Qveto for CBC events that estimates the energy distribution of the event over different time segments.

Let A_{max} be the absolute maximum amplitude of the reconstructed waveform of the event, and A_l & A_r are the left and right peak amplitudes of the peak. The total energy in these peaks is $E_1 = A_{max}^2 + A_l^2 + A_r^2$. The E_2 is the energy of the remaining peaks significant peaks (with amplitude greater than a fraction of A_{max}). The ratio E_2/E_1 gives a Q_{veto} statistic that helps to distinguish the localized glitch from the CBC signals. In the case of a localized glitch, most of the energy is accumulated around the A_{max} as opposed to the CBC signals, where the energy is distributed over longer signal duration. The Q_{veto} is smaller for glitches than the CBC signals.

4. Lveto

Detector data contains a large population of narrow band glitches (line structures in the TF map) *e.g.*, power line glitches and their harmonics. The cWB algorithm uses the statistics that computes the energy contribution to the event from the narrow frequency band.

The narrow frequency range is estimated from the loudest time-frequency pixel from the event cluster and this range is defined as $f_0 - df$ to $f_0 + df$ where f_0 is the center frequency of loud pixel and df is width of the loud pixel. The cWB algorithm stores frequency mean, RMS frequency, and ratio of energy in narrow band frequency range and full energy of the event. For the narrow glitches, the energy ratio is close to 1, that means most of energy comes from narrow frequency range. These glitches are rejected by applying thresholds on these values.

5. Chirp cut

This veto verifies whether the alignment of selected set of time-frequency pixels is roughly consistent with the expected model obtained from the zeroth- or Newtonian order approximation of the CBC signal frequency evolution, namely, $\frac{96}{5}\pi^{8/3}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/3} t + \frac{3}{8}f^{-8/3} + C = 0$, where $\mathcal{M}$ is the (unknown) chirp mass, G is the gravitational constant, c is the speed of light and C is a constant related to the merger time. The chirp mass parameter is estimated and error bars are produced using bootstrapping procedure using different subsets of selected pixels [52].

The typical conditions for accepting the event as a signal event is with N_{norm} greater that 2.5, χ^2 less than 0.2, Q_{veto} greater than 0.3 and $\mathcal{M}$ greater than 1.

[1] J. Aasi *et al.* (LIGO Scientific Collaboration), Class. Quant. Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[2] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[3] F. Acernese *et al.* (VIRGO), Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].

[4] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016), arXiv:1606.04855 [gr-qc].

[5] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **118**, 221101 (2017), arXiv:1706.01812 [gr-qc].

[6] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Astrophys. J. **851**, L35 (2017), arXiv:1711.05578 [astro-ph.HE].

[7] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].

[8] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. **X6**, 041015 (2016), [erratum: Phys. Rev.X8,no.3,039903(2018)], arXiv:1606.04856 [gr-qc].

[9] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), (2018), arXiv:1811.12907 [astro-ph.HE].

[10] B. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].

[11] Y. Aso *et al.* (KAGRA Collaboration), Phys. Rev. D **88**, 043007 (2013), arXiv:1306.6747 [gr-qc].

[12] B. Iyer *et al.*, LIGO India Technical Document (2011), https://dcc.ligo.org/LIGO-M1100296/public.

[13] K. Cannon *et al.*, Astrophys. J. **748**, 136 (2012), arXiv:1107.2665 [astro-ph.IM].

[14] S. Privitera, S. R. P. Mohapatra, P. Ajith, K. Cannon, N. Fotopoulos, M. A. Frei, C. Hanna, A. J. Weinstein, and J. T. Whelan, Phys. Rev. **D89**, 024003 (2014), arXiv:1310.5633 [gr-qc].

[15] C. Messick *et al.*, Phys. Rev. **D95**, 042001 (2017),

arXiv:1604.04324 [astro-ph.IM].

[16] T. Dal Canton *et al.*, Phys. Rev. **D90**, 082004 (2014), arXiv:1405.6731 [gr-qc].

[17] S. A. Usman *et al.*, Class. Quant. Grav. **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[18] I. Harry, S. Privitera, A. Bohé, and A. Buonanno, Phys. Rev. D **94**, 024012 (2016).

[19] N. Indik, K. Haris, T. Dal Canton, H. Fehrmann, B. Krishnan, A. Lundgren, A. B. Nielsen, and A. Pai, Phys. Rev. **D95**, 064056 (2017), arXiv:1612.05173 [gr-qc].

[20] J. Calderón Bustillo, S. Husa, A. M. Sintes, and M. Pürrer, Phys. Rev. **D93**, 084019 (2016), arXiv:1511.02060 [gr-qc].

[21] V. Varma, P. Ajith, S. Husa, J. C. Bustillo, M. Hannam, and M. Pürrer, Phys. Rev. **D90**, 124004 (2014), arXiv:1409.2349 [gr-qc].

[22] C. Capano, Y. Pan, and A. Buonanno, Phys. Rev. **D89**, 102003 (2014), arXiv:1311.1286 [gr-qc].

[23] E. A. Huerta *et al.*, Phys. Rev. **D95**, 024038 (2017), arXiv:1609.05933 [gr-qc].

[24] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, Phys. Rev. **D95**, 104046 (2017), arXiv:1511.05955 [gr-qc].

[25] N. J. Cornish and T. B. Littenberg, Class. Quant. Grav. **32**, 135012 (2015), arXiv:1410.3835 [gr-qc].

[26] E. Thrane and M. Coughlin, Phys. Rev. Lett. **115**, 181102 (2015), arXiv:1507.00537 [astro-ph.IM].

[27] B. P. Abbott *et al.* (LIGO Scientific, Virgo), (2019), arXiv:1906.08000 [gr-qc].

[28] M. Cabero *et al.*, (2019), arXiv:1901.05093 [physics.ins-det].

[29] P. Bacon, V. Gayathri, E. Chassande-Mottin, A. Pai, F. Salemi, and G. Vedovato, Phys. Rev. **D98**, 024028 (2018), arXiv:1805.04023 [astro-ph.IM].

[30] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, Class. Quant. Grav. **25**, 114029 (2008), arXiv:0802.3232 [gr-qc].

[31] S. Klimenko *et al.*, Phys. Rev. D **72**, 122002 (2005), gr-qc/0508068.

[32] S. Klimenko *et al.*, Phys. Rev. **D93**, 042004 (2016), arXiv:1511.05999 [gr-qc].

[33] B. Allen, Phys. Rev. **D71**, 062001 (2005), arXiv:gr-qc/0405045 [gr-qc].

[34] A. H. Nitz, Class. Quant. Grav. **35**, 035016 (2018),

arXiv:1709.08974 [gr-qc].

[35] T. Dal Canton, S. Bhagwat, S. Dhurandhar, and A. Lundgren, Class.Quant.Grav. **31**, 015016 (2014), arXiv:1304.0008 [gr-qc].

[36] S. Dhurandhar, A. Gupta, B. Gadre, and S. Bose, Phys. Rev. **D96**, 103018 (2017), arXiv:1708.03605 [gr-qc].

[37] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[38] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. **D93**, 122004 (2016), [Addendum: Phys. Rev.D94,no.6,069903(2016)], arXiv:1602.03843 [gr-qc].

[39] J. Aasi *et al.* (LIGO Scientific, VIRGO), Phys. Rev. **D89**, 122003 (2014), arXiv:1404.2199 [gr-qc].

[40] J. Abadie *et al.* (LIGO Scientific, VIRGO), Phys. Rev. **D85**, 122007 (2012), arXiv:1202.2788 [gr-qc].

[41] J. Abadie *et al.* (LIGO Scientific, VIRGO), Phys. Rev. **D85**, 102004 (2012), arXiv:1201.5999 [gr-qc].

[42] J. Abadie *et al.* (LIGO Scientific, VIRGO), Phys. Rev. **D81**, 102001 (2010), arXiv:1002.1036 [gr-qc].

[43] V. Necula *et al.*, J. Phys. Conf. Ser. **363**, 012032 (2012).

[44] T. Cokelaer, Phys. Rev. **D76**, 102004 (2007), arXiv:0706.4437 [gr-qc].

[45] I. W. Harry, B. Allen, and B. S. Sathyaprakash, Phys. Rev. **D80**, 104014 (2009), arXiv:0908.2090 [gr-qc].

[46] S. Mallat and Z. Zhang, IEEE Trans. Signal Proc. **41**, 3397 (1993).

[47] J. Abadie *et al.* (VIRGO, LIGO Scientific), (2012), arXiv:1203.2674 [gr-qc].

[48] A. Taracchini *et al.*, Phys. Rev. **D89**, 061502 (2014), arXiv:1311.2544 [gr-qc].

[49] M. Pürrer, Phys. Rev. **D93**, 064041 (2016), arXiv:1512.02248 [gr-qc].

[50] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, *Proceedings, 10th International LISA Symposium: Gainesville, Florida, USA, May 18-23, 2014*, J. Phys. Conf. Ser. **610**, 012021 (2015), arXiv:1410.4839 [gr-qc].

[51] B. P. Abbott *et al.* (LIGO Scientific), Phys. Rev. **D95**, 062003 (2017), arXiv:1602.03845 [gr-qc].

[52] V. Tiwari, S. Klimenko, V. Necula, and G. Mitselmakher, Class. Quant. Grav. **33**, 01LT01 (2016), arXiv:1510.02426 [astro-ph.IM].