LIGO–P070041–01–Z

Enhancing the capabilities of LIGO time-frequency plane searches through clustering

R Khan[1], **S Chatterji**[2]

[1] Columbia Astrophysics Laboratory, Columbia University, Pupin Labs Rm 1027, MC 5247, New York, NY 10027 USA

[2] LIGO Laboratory, California Institute of Technology, MS 18-34, Pasadena, CA 91125 USA

E-mail: rubab@astro.columbia.edu, shourov@ligo.caltech.edu

Abstract.

One class of gravitational wave signals LIGO is searching for consists of short duration bursts of unknown waveforms. Potential sources include core collapse supernovae, gamma ray burst progenitors, and the merger of binary black holes or neutron stars. We present a density-based clustering algorithm to improve the performance of time-frequency searches for the such gravitational-wave bursts when they are extended in time and/or frequency. We have implemented this algorithm as an extension to the QPipeline search for bursts, which currently determines the statistical significance of events based solely on the peak significance observed in minimum uncertainty regions of the time-frequency plane. Density based clustering improves the performance of such a search by considering the aggregate significance of arbitrarily shaped regions in the time-frequency plane and rejecting the isolated minimum uncertainty features expected from the background detector noise. In this paper, we present test results for simulated signals and demonstrate that density based clustering improves the performance of the QPipeline for signals that are extended in time and/or frequency.

PACS numbers: 04.80.Nn, 07.05.Kf, 95.55.Ym, 95.75.Pq

Submitted to: *Class. Quantum Grav.*

1. Introduction

The first generation of interferometric gravitational wave detectors are now collecting data at their design strain sensitivities [1, 2, 3, 4, 5, 6]. Even at this unprecedented level of sensitivity, potentially detectable signals from astrophysical sources are expected to be at or near the limits of detectability, requiring carefully designed search algorithms in order to identify and distinguish them from the background detector noise. In this study, we focus on the problem of detecting the specific class of gravitational wave signals known as gravitational-wave bursts (GWBs). These are signals lasting from a few milliseconds to a few seconds, for which we do not have sufficient theoretical understanding or reliable models to predict a waveform. This includes signals from the merger of binary compact objects, asymmetric core collapse supernovae, the progenitors of gamma ray bursts, and possibly unexpected sources.

Since accurate waveform predictions do not exist for GWBs, the typical method to identify them is to project the data under test onto a convenient basis of abstract waveforms that are chosen to cover a targeted region of the time-frequency plane, and then identify regions of this search space with statistically significant excess signal energy [7]. In this study, we focus on one such burst search algorithm, the QPipeline [8], which first projects the data under test onto an overcomplete basis of Gaussian enveloped sinusoids characterized by their center time, center frequency, and quality factor. A trigger is recorded whenever this projection exceeds a threshold value, with the magnitude of the projection indicating the significance of the trigger. Since the triggers are considered separately, the existing algorithm currently under-reports the total energy and true significance of those signals that are extended in time and/or frequency as they have a significant projection onto multiple independent basis functions. Since GWB signals with such extended features are commonly observed in simulations of core collapse supernovae, the mergers of binary compact objects, and instabilities of spinning neutron stars, there are good reasons to try to improve the sensitivity of the search algorithm to such sources.

To improve the sensitivity of the QPipeline to signals that are extended in time and/or frequency, we have investigated extensions to the QPipeline that also consider the combined statistical significance of arbitrarily shaped clusters of projections in the time-frequency plane. Although a number of clustering algorithms are commonly available [9], this work focuses on a density based clustering algorithms due to its ability to also decrease the false detection probability of GWB searches by rejecting isolated single projection events associated with noise fluctuations. In this paper we present the details of a density based clustering algorithm implementation as an extension to the Qpipeline, and demonstrate the resulting improved performance of the QPipeline for signals that are extended in time and/or frequency.

The paper is structured as follows. Section 2 briefly describes the QPipeline burst search algorithm. Section 3 considers the motivations for clustering and surveys some of the available approaches. Section 4 presents the details of the proposed density based

clustering algorithm. Section 5 demonstrates the benefit of the proposed approach for detecting simulated gravitational wave bursts of various waveforms. Finally, in section 6, we present our conclusions and discuss possible future investigations.

2. The QPipeline burst search algorithm

The QPipeline is an analysis pipeline for the detection of GWBs in data from interferometric gravitational wave detectors [8]. It is based on the Q transform [10], a multi-resolution time-frequency transform that projects the data under test onto the space of Gaussian windowed complex exponentials characterized by center time τ, center frequency ϕ, and quality factor Q.

$$X(\tau, \phi, Q) = \int_{-\infty}^{+\infty} x(t)\, e^{-4\pi^2\phi^2(t-\tau)^2/Q^2}\, e^{-i2\pi\phi t}\, dt \tag{1}$$

The space of Gaussian enveloped complex exponentials is an overcomplete basis of waveforms, whose duration σ_t and bandwidth σ_f have the minimum possible time-frequency uncertainty, $\sigma_t\sigma_f = 1/4\pi$, where $Q = \phi/\sigma_f$. As a result, they provide the tightest possible constraints on the time-frequency area of a signal, maximizing the measured signal to noise ratio (SNR) and minimizing the probability that false triggers are coincident in time and frequency between multiple detectors.

In practice, the Q transform is evaluated only for a finite number of basis functions, which are more commonly referred to as templates or tiles. These templates are selected to cover a targeted region of signal space, and are spaced such that the fractional signal energy loss $-\delta Z/Z$ due to the mismatch $\delta\tau$, $\delta\phi$, and δQ between an arbitrary basis function and the nearest measurement template,

$$\frac{-\delta Z}{Z} \simeq \frac{4\pi^2\phi^2}{Q^2}\,\delta\tau^2 + \frac{2+Q^2}{4\phi^2}\,\delta\phi^2 + \frac{1}{2Q^2}\,\delta Q^2 - \frac{1}{\phi Q}\delta\phi\,\delta Q, \tag{2}$$

is no larger than $\sim$20%. This naturally leads to a tiling of the signal space that is logarithmic in Q, logarithmic in frequency, and linear in time.

The statistical significance of Q transform projections are given by their normalized energy Z, defined as the ratio of squared projection magnitude to the mean squared projection magnitude of other templates with the same central frequency and Q. For the case of ideal white noise, Z is exponentially distributed, and related to the matched filter SNR ρ [11] by the relation

$$Z = |X|^2/\langle|X|^2\rangle_\tau = -\ln P(Z' > Z) = \rho^2/2. \tag{3}$$

The QPipeline consists of the following steps. The data is first whitened by zero-phase linear predictive filtering [12, 8]. Next, the Q transform is applied to the whitened data, and normalized energies are computed for each measurement template. Templates with statistically significant signal content are then identified by applying a threshold on the normalized energy. Finally, since a single event may potentially produce multiple overlapping triggers due to the overlap between measurement templates, only the most significant of overlapping templates are reported as triggers. As a result, the QPipeline is

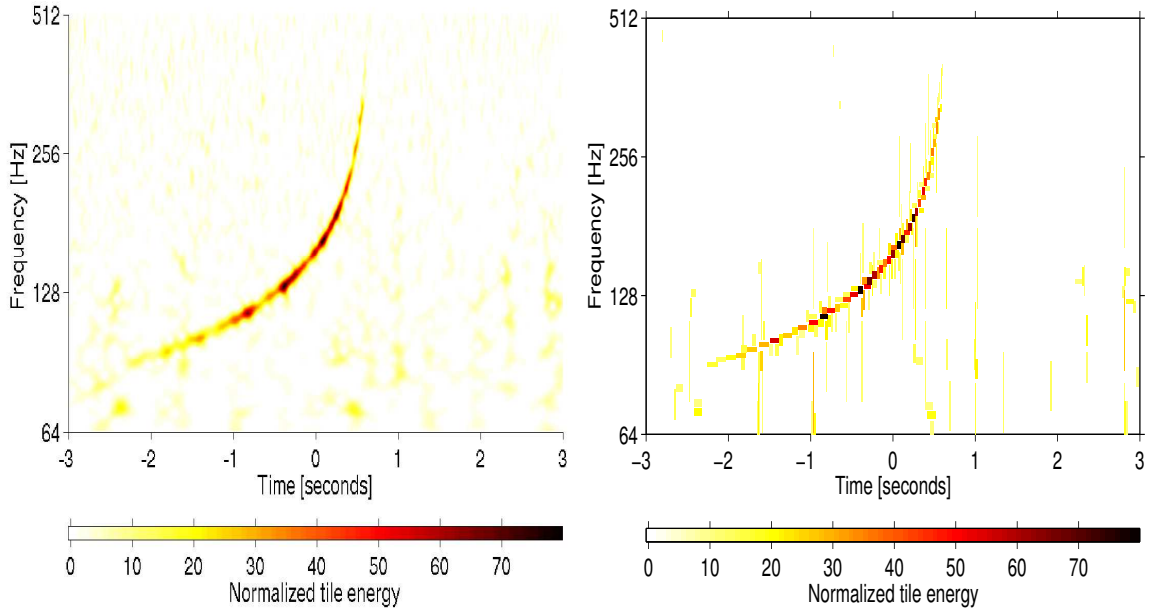

Figure 1. The QPipeline view of the inspiral phase of a simulated optimally oriented 1.4/1.4 solar mass binary neutron star merger injected into typical LIGO data with an SNR of 48.2 as measured by a matched filter search targeting inspiral signal. The QPipeline projects the whitened data onto the space of time, frequency, and Q. The upper left image shows the resulting time-frequency spectrogram of normalized signal energy for the value of Q that maximizes the measured normalized energy, while the upper right image shows the time-frequency distribution of only the most significant non-overlapping triggers regardless of Q. The authors gratefully acknowledge the LIGO Scientific Collaboration hardware injection team for providing the data used in this figure.

effectively a templated matched filter search [11] for signals that are Gaussian enveloped sinusoids in the whitened signal space.

Figure 1 shows an example of the QPipeline applied to the inspiral phase of a simulated binary neutron star coalescence signal injected into typical LIGO data.

3. Motivations and options for clustering

Currently, the QPipeline considers the significance of triggers independently. The detectability of a particular GWB signal therefore depends upon its maximum single projection onto the space of Gaussian enveloped sinusoids. As a result, the QPipeline is most sensitive to signals with near minimum time-frequency uncertainty, and less sensitive to signals that are extended in time and/or frequency such that their energy is spread across multiple non-overlapping basis functions, more commonly referred to as templates or tiles, as mentioned in Section 2. For example, the detectability of the simulated inspiral signal shown in Figure 1 is currently determined by the single most significant tile near the center of the signal, which has a single tile SNR of 12.7. This

is significantly less than the SNR of 48.2 that is recovered by a matched filter search tuned for this waveform.

While the above example focuses on binary neutron star inspiral waveforms as an astrophysically interesting example case, other potential GWB sources, such as the less well understood merger phase of coalescing binary compact objects, core collapse supernovae, and instabilities in spinning neutron stars, are all expected to produce waveforms that are extended in time and frequency[13, 14]. As a result, we seek a method to improve the sensitivity of the QPipeline to signals that are extended in time and/or frequency that is applicable to the general case of astrophysically unmodeled bursts, and is not specific to any one particular waveform.

An obvious solution is to simultaneously consider the aggregate significance of all tiles that comprise the signal. This requires an approach that identifies clusters of related tiles in the time frequency plane. In this context, we define clustering as the grouping of the set of all significant QPipeline tiles into subsets, such that all tiles within a subset are closely related by their relative distance in the time frequency plane.

There are many different clustering methods available; however, they all tend to fall into three categories [9].

Partitioning methods The classical example of a partitioning method is the K-means [15] algorithm. In this case, a fixed number of clusters is first assumed, and an initial guess to partition objects into these clusters is made. A centroid is computed for each cluster, and the total sum distance of all objects from their cluster centroid is computed as a figure of merit. K-means iteratively reassigns objects to different clusters until this figure of merit is minimized.

There are a number of drawbacks to the k-means approach. The first is that presupposes a fixed number of clusters. While there are variations that allow the number of clusters to change [16], the other difficulty with the k-means approach is the tendency to produce spherical or ellipsoidal clusters rather than arbitrary shapes. Although this is acceptable for some signal morphologies, it is not acceptable in general. For example, the signal expected from inspiralling binary compact objects, which is long and extended in time and frequency, would not be easily identified k-means clustering. A third difficulty with the k-means approach is the sensitivity to the initial guess, and the possibility of the algorithm to identify a local rather than global minimum.

Hierarchical methods This type of clustering algorithm first evaluates the pairwise difference between all N objects, then arranges them into a tree structure, where each object to be clustered is a leaf [17]. In the agglomerative hierarchical approach, the tree is constructed in N levels. At each level, the closest pair of leaves and/or branches is merged, with the Nth level representing a cluster of all objects. A clustering is formed by cutting this tree based on some criteria such as a maximum distance or the inconsistency between the distance between cluster leaves or branches and the distance to the next closest leaf of branch.

Hierarchical clustering has the flexibility of producing arbitrary numbers of clusters with arbitrary shape, and is therefore more applicable to the problem of clustering in

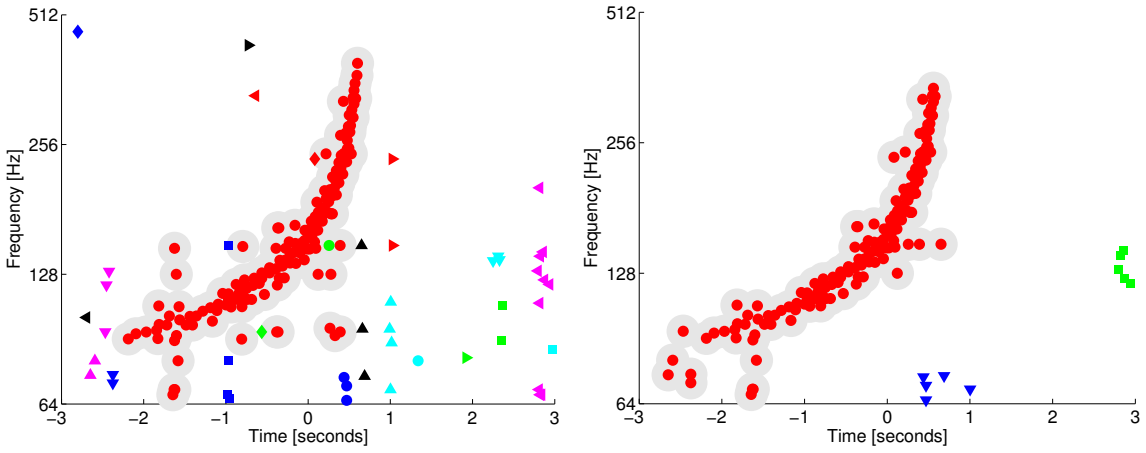

Figure 2. Two different clustering methods were applied to the QPipeline triggers from Figure 1. The lower left image shows the result of applying a hierarchical based clustering method [9], while the lower right image shows the result of applying the proposed density based clustering method[18]. Here each combination of color and shape denotes a different cluster. Although both hierarchical and density based clustering approaches succeed in isolating most of the signal energy within a single cluster, the density based approach has the additional advantage of discarding isolated triggers due to the background detector noise.

time, frequency, and Q for gravitational-wave burst detection.

The left panel from Figure 2 shows an example of hierarchical clustering applied to the detection of a simulated binary neutron star inspiral signal. The draw back to this approach is that it presupposes that every data point must be included in one cluster or another, even if a cluster is to be constituted of only one data point. As a result, it produces a large number of insignificant clusters and tends to build clusters of unrelated data points.

Density based methods Density based clustering is a variation on the hierarchical clustering approach. Instead of constructing a tree structure, density based clustering starts with single object and iteratively adds objects to that cluster using a predefined set of criteria based on the density of objects within a given neighborhood radius, until the criteria is no longer met and all objects have been tested.

Like other hierarchical methods, density based methods also allow an arbitrary number of clusters as well as arbitrary cluster shape. They also have the advantage of rejecting single isolated data points that are not potentially related with a large number of points. Thus this method only produces clusters with multiple data points and can successfully exclude unrelated points from a cluster. For this reason, we have focused on density based methods in this work.

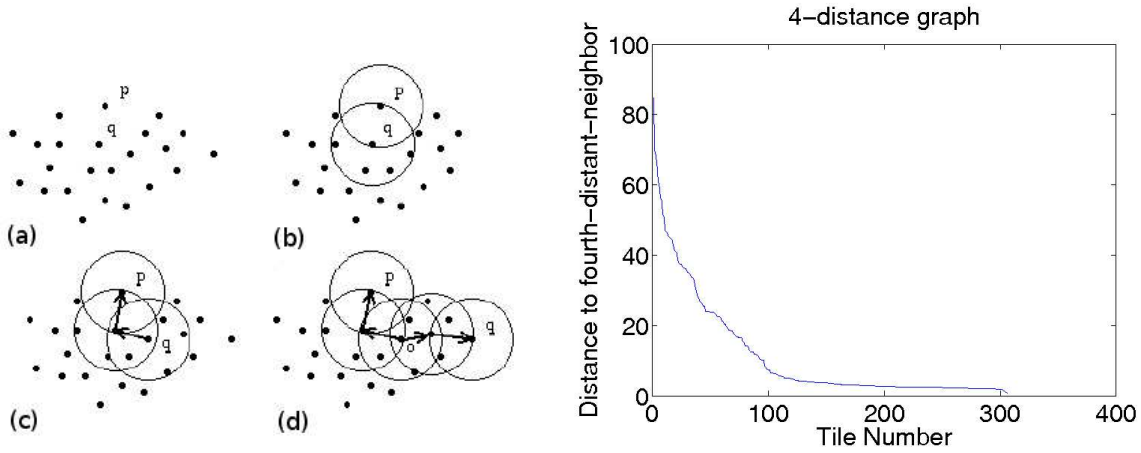

Figure 3. Building clusters from data-points using the density based clustering algorithm. (Figure from [18]).

4. Density based clustering algorithm

Density based clustering [18] facilitates searches for signals of unknown shape. It does not clutter the output with a list of numerous noise related clusters that contain just a few significant data-points. The algorithm looks for neighbors of those points that have at least a given number of neighboring points within a given distance on the time-frequency plane, and forms clusters of data-points that can be related through their common neighbors. Our implementation of density based clustering algorithm takes two parameter: minimum neighbor number and neighborhood radius, and it considers each tile produced by Qpipeline as a data-point.

Density based clustering first finds a tile's nearest neighbors, then that neighbors' neighbors, and so on. In the left panel of Fig. 3, (a) shows data points before clustering. If the density of data points within a given distance around a point is above a given threshold to form a cluster, that point becomes a cluster seed (b). Neighboring data points having a sufficient number of neighbors are then included in the cluster (c). This process repeats as long as data points with sufficient number of neighbors are found (d). The four-distance graph (Fig. 3, right panel) has the distance of the fourth closest neighbor of every point along y-axis for every corresponding point on x-axis. The turn close to distance 8 provides us with the numerical value of neighborhood radius (right).

Any clustering algorithm requires measurement of the pairwise distances between all data points and in our case, the pairwise distance between all tiles produced by Q Pipeline. However, the tiles have varied shapes which make measurement of distance between any pair of data points rather difficult. We implemented a distance metric that takes into account the issue of varied tile shapes. It also inflates the distance on frequency scale relative to the distance in time scale by a factor of β. This step has to be taken in order to compensate for the fact that most nonlocalized signals are quite limited on the time scale (seconds) while being comparatively more extended on the

frequency scale (Hz). For a pair of tiles with center times t_1 and t_2, center frequencies f_1 and f_2, Q of q_1 and q_2, and normalized energy of z_1 and z_2, the distance on the time-frequency plane D is measured from the following ralations:

$$D = \sqrt{{D_t}^2 + \beta {D_f}^2} \tag{4}$$

$$D_t = \frac{|t_2 - t_1|}{\overline{\Delta t}}, D_f = \frac{|f_2 - f_1|}{\overline{\Delta f}} \tag{5}$$

$$\overline{\Delta t} = \frac{\Delta t_1 z_1 + \Delta t_2 z_2}{z_1 + z_2}, \overline{\Delta f} = \frac{\Delta f_1 z_1 + \Delta f_2 z_2}{z_1 + z_2} \tag{6}$$

where D_t is the distance along the time scale, D_f is the distance along the frequency scale, $\overline{\Delta t}$ is the scale factor on the time scale, $\overline{\Delta f}$ is the scale factor on the frequency scale, Δt_1 and Δt_2 are durations, and Δf_1 and Δf_2 are bandwidths. It has been assumed that extended burst signals will be extended accross a very wide frequency brand while lasting for a short period. The factor of β thus reduces the possibility of such signals being seperated into multiple clusters by relatively downplaying the distance along the time scale. It was determined empirically that $\beta = 30$ gives the best performance as measured by the ROC curves in Section 5 for the typical LIGO data that was tested and the waveforms considered here.

The mismatch metric can also be used [8] for this purpose. However that is not as effective as the metric described here.

The exact numerical value of the neighborhood radius depends on the minimum neighbor number, which in this case was empirically determined to be 4. Thus the value of the neighborhood radius is determined using a 4-distance graph that has the distance of the fourth closest neighbor of every point along y-axis for every corresponding point on x-axis. The points are sorted according to descending order of their 4-distance value. Close observation of the 4-distance graph provides a cut-off distance whose numerical value depends on the distance metric used. For the specific distance metric we used, the numeric value of 8 has been chosen as the neighborhood radius from observing the 4-distance plot that we produced (Fig. 3, right) that shows a turn in the plot near that point.

The main clustering function first uses the distance function to measure pairwise distance between all tiles, and calls the expandCluster function which recursively calls itself to induct more data points into the cluster. A flowchart of the algorithm we have implemented is shown in Fig. 4. This algorithm was implemented in MATLAB along with Qpipeline.

Clustering starts at the highest energy data-point and then proceeds to the next significant data point that is not in a cluster, considering only such points as cluster seeds that have enough (4 in our case) neighbors to ensure that the least number of loops are executed. If any qualifying member of the current cluster is found to be already in a cluster, the two clusters merge. Thus, regardless of which data-point the algorithm starts clustering from, it will always find the same clusters for a given set of

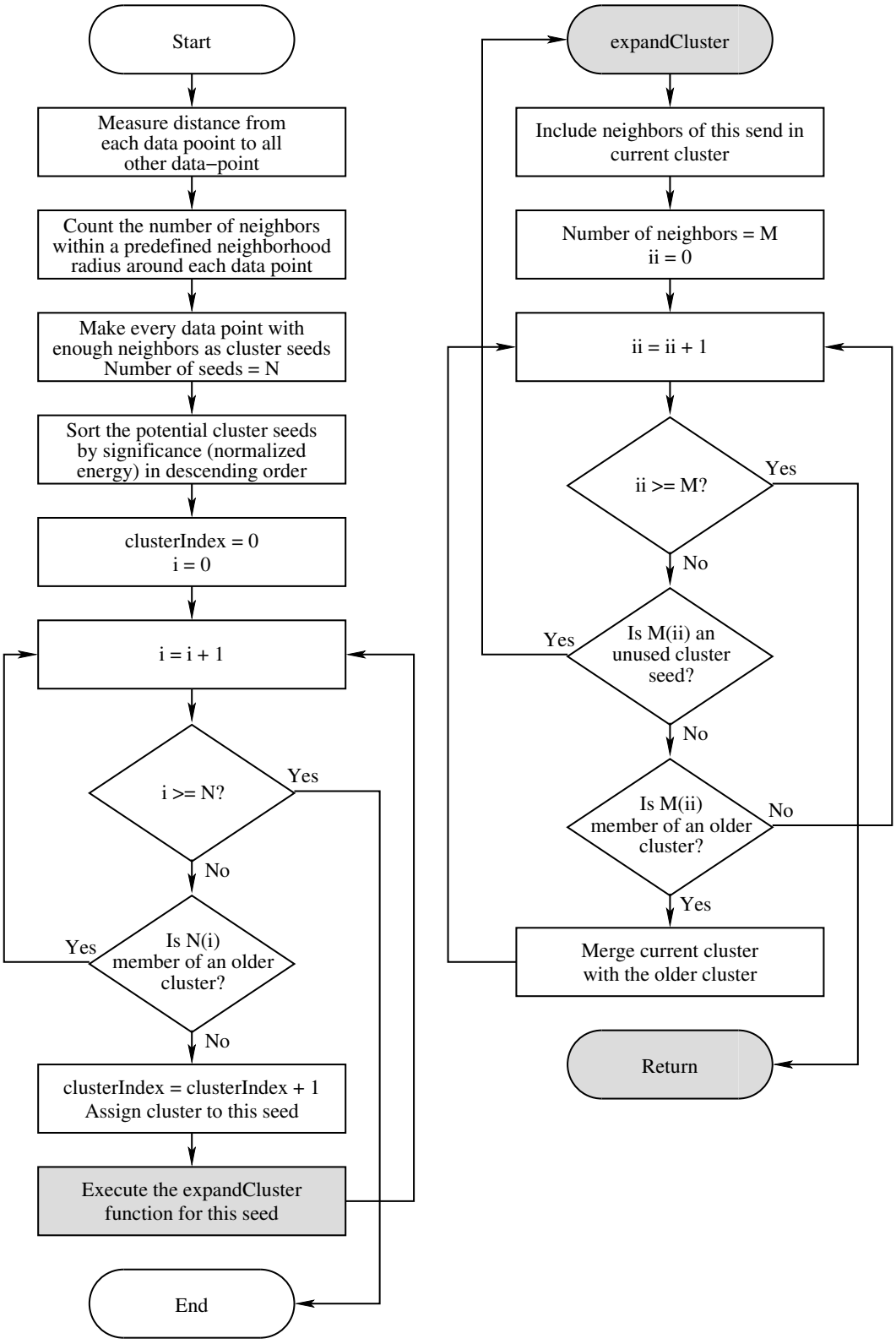

Figure 4. Flowchart of density based clustering algorithm.

data points. For speed optimization, though, our density based clustering function picks the more significant data-points first. THe right panel of Fig. 2 shows a cluster built using density based clustering algorithm. It can be seen that density based clustering has clustered together the most significant part of the previously discussed injection successfully, and almost all the noise is removed. While it loses the high-frequency end of the injection, that part contains very little energy, and does not significantly contribute to the significance of the detected trigger. The only noise cluster on the signal space is a low frequency event occuring a couple of seconds after the injected signal, and is associated with a non-stationary detector noise.

5. Evaluating performance improvements

We have evaluated our implementation of density based clustering by measuring its effect on the detection of simulated signals injected into typical LIGO data, and its effect on the rate of false detections.

In order to evaluate the false detection rate, the QPipeline was first applied to single detector data without injected signals. This was performed both with and without clustering. False events were identified as those events whose total normalized energy exceeded a specified detection threshold. Three sets of false events were identified: unclustered events, clustered events, and combined events formed by the union of unclustered and clustered events. The resulting false event rates as a function of detection threshold are shown in Figure 5.

The lower false event rate observed in Figure 5 for clustered triggers at low detection thresholds is associated with the rejection of isolated noise events as described in Section 4. At high detection thresholds, the opposite is true. The presence of transient non-stationary "glitches" in the data that are extended in time and/or frequency cause the false event rate of clustered triggers to exceed that of unclustered triggers.

To evaluate the effect of clustering on the detection of signals, we next applied the QPipeline to the recovery of simulated signals injected into the same single detector data. Again, this was performed both with and without clustering. Injections were identified as detected if a event was observed above the detection threshold within 1 second of the time of the injected signal. We define the detection efficiency as the fraction of injection signals that were correctly detected, and evaluate this efficiency as a function of detection threshold for signals injected at a constant signal to noise ratio.

In order to characterize the performance of density based clustering for a variety of signal morphologies, we have repeated this analysis for five different waveform families. Three of the waveform families were abstract and meant to characterize performance over a wide signal space. These include simple Gaussian pulses, sinusoidal Gaussian pulses, and band-limited and time-windowed noise. Two of the waveform families were astrophysically motivated. These include the inspiral phase of coalescing binary compact objects and the fundamental ring down mode of perturbed black holes. Within each waveform family, signals were injected with random parameters such as time, frequency,

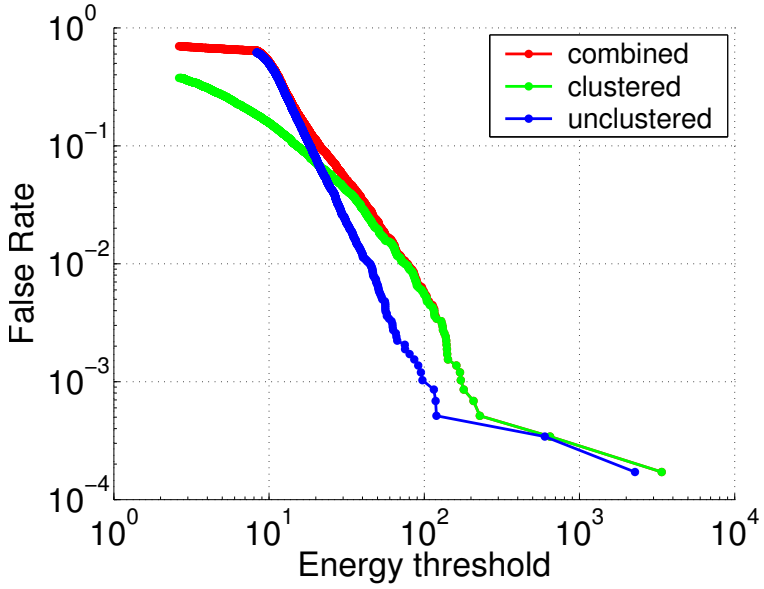

Figure 5. The false event rate of the search algorithm as a function of detection threshold when applied to typical LIGO data. The trigger rate is shown for three different trigger sets: unclustered triggers, clustered triggers, and the union of clustered and unclustered triggers.

duration, bandwidth, mass, etc.

On the left side panels of Figure 6, we present the resulting detection efficiency as a function of detection threshold for three of the waveform families that we have considered, representing both ad-hoc and astrophysics, as well as localized and non-localized. On the right side panels of Figure 6, we report the receiver operator characteristic (ROC) for each waveform, which combines the measured false rate from Figure 5 with the detection efficiencies from the left side panels.

The results indicate that for the extended waveforms, such as the inspiral and noise burst waveforms, clustering increases search efficiency and significantly improves the resulting ROC by approximately an order of magnitude in false rate. The primary reason for this improved performance is the increase in measured signal energy due to clustering, which is evident as increased detection efficiency in the left hand side of Figure 6.

Although clustering provides a marked improvement for the detection of signals that are extended in time and frequency, Figure 6 indicates that clustering also adversely impacts the performance of the search for localized waveforms. In particular, the ROC for sinusoidal Gaussians is worse by roughly a factor of 3 in false rate due to the addition of clustering. The primary cause of this decreased performance is the higher false event rate, which is due to the increased significance of detector glitches after clustering, and is evident in Figure 5. Although for signals that are extended in time and/or frequency the higher false event rate is compensated by the significant improvement in detection efficiency, for more localized signals such improvement is not as significant. However,

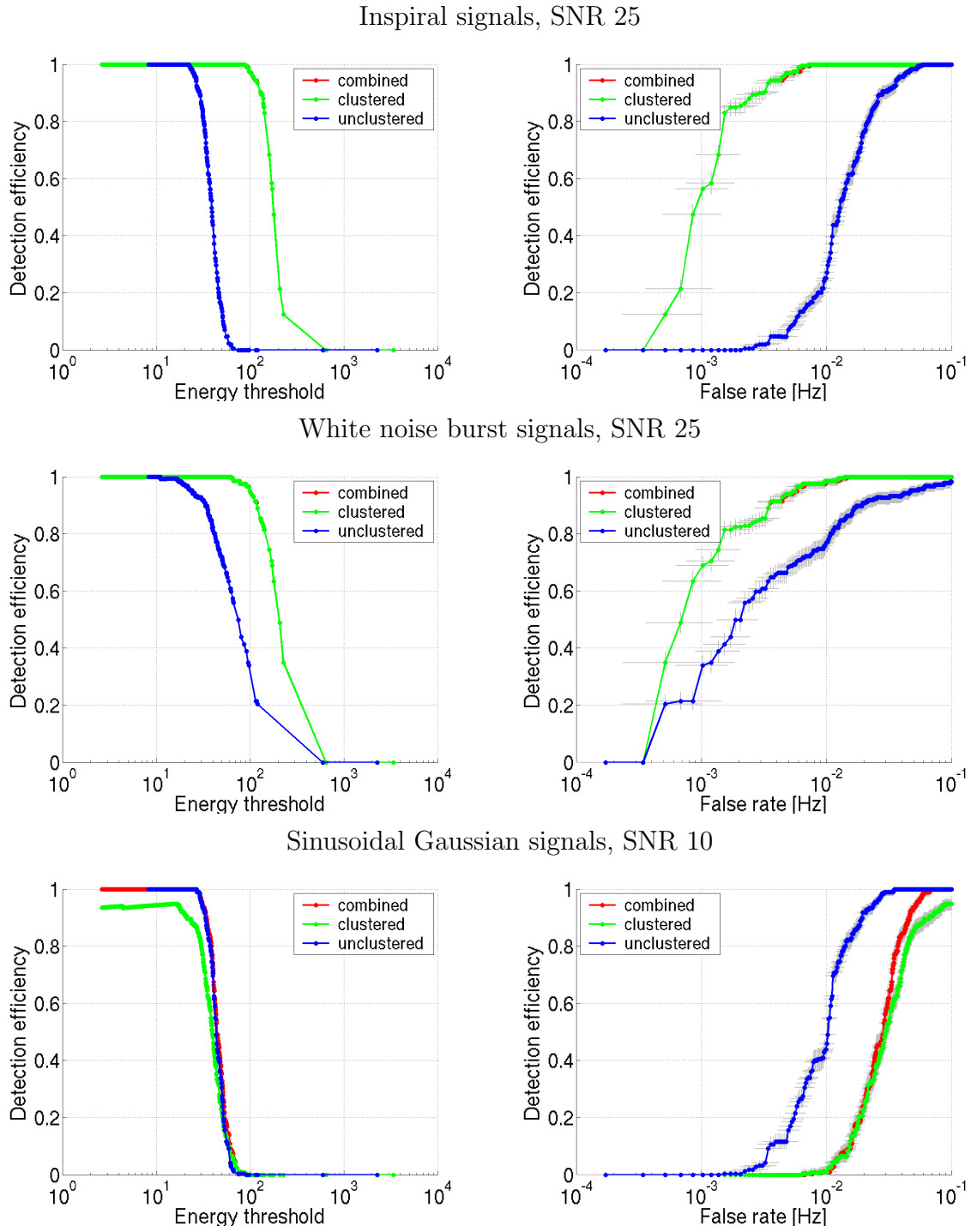

Figure 6. Comparison of the detection efficiency vs. search threshold (left) and Receiver Operator Characteristics (ROC) (right) of the search algorithm, with and without clustering, applied to the detection of simulated inspiral (top), white noise burst (middle), and sinusoidal Gaussian (bottom) waveforms injected 200 times into typical LIGO data at fixed SNR.

in practice, we expect the presence of such detector glitches to be largely mitigated by the requirement of a coincident and consistent observation of a gravitational wave in multiple detectors, as well as the absence of a signal in environmental monitors.

Another adverse effect is due to the tendency of density based clustering to exclude isolated triggers. This results in the occasional rejection of highly localized signals, regardless of the detection threshold. This is also evident in Figure 6, where detection efficiency of sinusoidal Gaussians does not converge to 100 percent for the case of clustered triggers. To overcome this, we have also considered the performance of a search consisting of the union of both clustered and unclustered triggers. The resulting combined detection efficiency is then comparable to that of the unclustered case. We note that reducing the required neighbor number within the neighborhood redius to zero would eliminate the need to combine clustered and unclustered triggers because then density based clustering methods would not discard single tile triggers anymore. Classical hierarchical clustering may also provide an interesting alternative to density based clustering in the case of a more sensitive detector [19, 20] at the relatively lower occurance of detector glitches which will eliminate the necessity to exlude isolated triggers.

6. Conclusions

Methods of clustering the measurements from neighboring or overlapping basis functions have been employed to more efficiently detect signals that are not well represented by QPipeline's particular choice of basis. Adding density based clustering algorithm to QPipeline for statistically significant events led to an improvement in the detectability of GW burst signals extended in time and/or frequency scales. Our implementation of density based clustering facilitates QPipeline to find clusters of unknown shapes and rejects noise without slowing down the search.

The proposed clustering algorithm itself is not specific to the QPipeline, and similar improvements are expected when applied to other time-frequency searches for gravitational wave bursts. In particular, the algorithm described here has already been applied to the search for bursts from the soft gamma repeaters with the flare pipeline [21]. Use of density based clustering, instead of the sum over frequency bins method which would consider total energy across all frequency bins corresponding to the targeted time bins while estimating uppper limits, improved flare pipeline's upper limit estimate for 100ms long 64Hz-1024Hz WNB injections by 42%, although no improvement is reported for 22ms long 100Hz-200Hz WNB injections which effectibely resembles the localized Sinegaussian injections. These results are consistent with our conclusion that density based clustering is useful while searching for extended signals but not so in the case of localized signals.

Since all the testing so far has been done using single detector data, the logical next step is to incorporate clustering in to Qpipeline and other potential multiple detector search algorithms, and implement coherent and co-incident search capabilities.

A reduction of the false rate is expected for coherent and co-incident searches as the presence of noise will be largely mitigated by the requirement of a coincident and consistent observation of gravitational wave in multiple detectors. Clustering can help us to extract information about signal shapes by identifying GW signal energy distributions across time and frequency bins, and investigate other signal characteristics that potential search methods can not recognize without clustering. We recommend further research to explore these promising aspects.

Acknowledgments

The authors are grateful for the support of the United States National Science Foundation under cooperative agreement PHY-04-57528, California Institute of Technology, and Columbia University in the City of New York. We are grateful to the LIGO Scientific collaboration for their support. We are indebted to many of our colleagues for frequent and fruitful discussion. In particular, we'd like to thank Albert Lazzarini for his valuable suggestions regarding this project, and Luca Matone, Zsuzsa Márka, Sharmila Kamat, Jameson Rollins, Peter Kalmus, John Dwyer, Patrick Sutton, Eirini Messeritaki, and Szabolcs Márka for their thoughtful comments on the manuscript.

The authors gratefully acknowledge the support of the United States National Science Foundation for the construction and operation of the LIGO Laboratory and the Particle Physics and Astronomy Research 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. The authors also gratefully acknowledge the support of the research by these agencies and by the Australian Research Council, the Natural Sciences and Engineering Research Council of Canada,the Council of Scientific and Industrial Research of India, the Department of Science and Technology of India, the Spanish Ministerio de Educaciony Ciencia, The National Aeronautics and Space Administration, the John Simon Guggenheim Foundation, the Alexander von Humboldt Foundation,the Leverhulme Trust, the David and Lucile Packard Foundation, the Research Corporation, and the Alfred P. Sloan Foundation. The LIGO Observatories were constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation under cooperative agreement PHY-9210038. The LIGO Laboratory operates under cooperative agreement PHY-0107417. This document has been assigned LIGO document number LIGO-P070041-01-Z.

References

[1] D. Sigg and the LIGO Science Collaboration. Status of the LIGO detectors. *Classical and Quantum Gravity*, 23:51–+, April 2006.

[2] LIGO Scientific Collaboration. LIGO detector noise curves during the S5 run. *http://www.ligo.caltech.edu/docs/G/G060009-03/*, 2007.

[3] Acernese, F. et al. *Classical and Quantum Gravity*, 23:S63, 2006.
[4] Lück, H. et al. *Classical and Quantum Gravity*, 23:S71, 2006.
[5] F. Beauville et al. Benefits of joint LIGO - Virgo coincidence searches for burst and inspiral signals. *Journal of Physics Conference Series*, 32:212–222, March 2006.
[6] Abbott B. et al. Detector description and performance for the first coincidence observations between LIGO and GEO. *Nuclear Instruments and Methods in Physics Research A*, 517:154–179, January 2004.
[7] Warren G. Anderson, Patrick R. Brady, Jolien D. E. Creighton, and Eanna E. Flanagan. An excess power statistic for detection of burst sources of gravitational radiation. *Phys. Rev.*, D63:042003, 2001.
[8] S. K. Chatterji. *The search for gravitational wave bursts in data from the second LIGO science run*. PhD thesis, Massachusetts Institute of Technology, 2005.
[9] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323, 1999.
[10] Shourov Chatterji, Lindy Blackburn, Gregory Martin, and Erik Katsavounidis. Multiresolution techniques for the detection of gravitational-wave bursts. *Class. Quantum Grav.*, 21:S1809–S1818, 2004.
[11] D. Dewey. Analysis of gravity wave antenna signals with matched filter techniques. *Molecular Genetics and Metabolism*, pages 581–590, 1986.
[12] J. Makhoul. Linear prediction: A tutorial review. *Proc. IEEE*, 63, 1975.
[13] T. Z. Summerscales et al. Maximum Entropy for Gravitational Wave Data Analysis: Inferring the Physical Parameters of Core-Collapse Supernovae. *ArXiv e-prints*, 704, April 2007.
[14] T. Baumgarte et al. Learning about compact binary merger: the interplay between numerical relativity and gravitational-wave astronomy. *ArXiv General Relativity and Quantum Cosmology e-prints*, December 2006.
[15] Tapas Kanungo et al. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):881–892, 2002.
[16] H. Lei et al. Cluster Analysis of Simulated Gravitational Wave Triggers Using S-means and Constrained Validation Clustering.
http://www.ligo.caltech.edu/docs/P/P080035-00/, March 2008. LIGO P document. Not yet uploaded. Authors include Soumya Mohanty and Soma Mukherjee.
[17] S. Krishnamachari and M. Abdel-Mottaleb. Hierarchical clustering algorithm for fast image retrieval. In M. M. Yeung, B.-L. Yeo, and C. A. Bouman, editors, *Proc. SPIE Vol. 3656, p. 427-435, Storage and Retrieval for Image and Video Databases VII, Minerva M. Yeung; Boon-Lock Yeo; Charles A. Bouman; Eds.*, pages 427–435, December 1998.
[18] Martin Ester et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231, 1996.
[19] S. Waldman R. Adhikari, P. Fritschel. Enhanced LIGO. 2006. LIGO technical document T060156.
[20] http://www.astro.cf.ac.uk/geo/advligo/.
[21] P. Kalmus et al. Search method for unmodeled transient gravitational waves associated with SGR flares. *Class. Quant. Grav.*, 24:S659–S669, 2007.