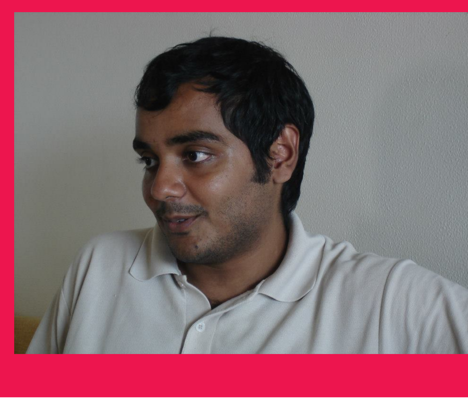
# Event classification for a gravitational-wave inspiral search with a sine-Gaussian glitch veto



LIGO Document G1500537

**Shasvath J. Kapadia** *†, Thomas Dent†, Tito Dal Canton†

* *Department of Physics, University of Arkansas, Fayetteville, USA*

† *Albert-Einstein-Institut (Max-Planck-Institut für Gravitationsphysik), Hannover, Germany*

## Motivation

- Gravitational waves (GWs) from inspiraling compact binaries – neutron stars and black holes – are likely first detections with advanced LIGO-Virgo.
- Sensitivity of templated searches is limited by high signal-to-noise-ratio (SNR) events caused by non-Gaussian noise or "glitches".
- To distinguish between glitches and real signals, recent searches used a $\chi^2$ test [**?**] and ranked candidate events by "re-weighted SNR" $\hat{\rho}$ [**?**], a fixed function $\chi^2$ and SNR, both of which are properties of **individual** events.
- However some events with high $\hat{\rho}$ are still easily identified as glitches by examining surrounding data.
- We propose using information from **multiple** inspiral and sine-Gaussian triggers [**?**] **around the time** of each candidate event to improve classification.

## Inspiral Triggers

- Triggers generated using non-spinning matched filter templates applied to simulated (glitchy) early aLIGO data injected with spinning inspiral signals. Implementation and details of the simulations are described in [**?**].

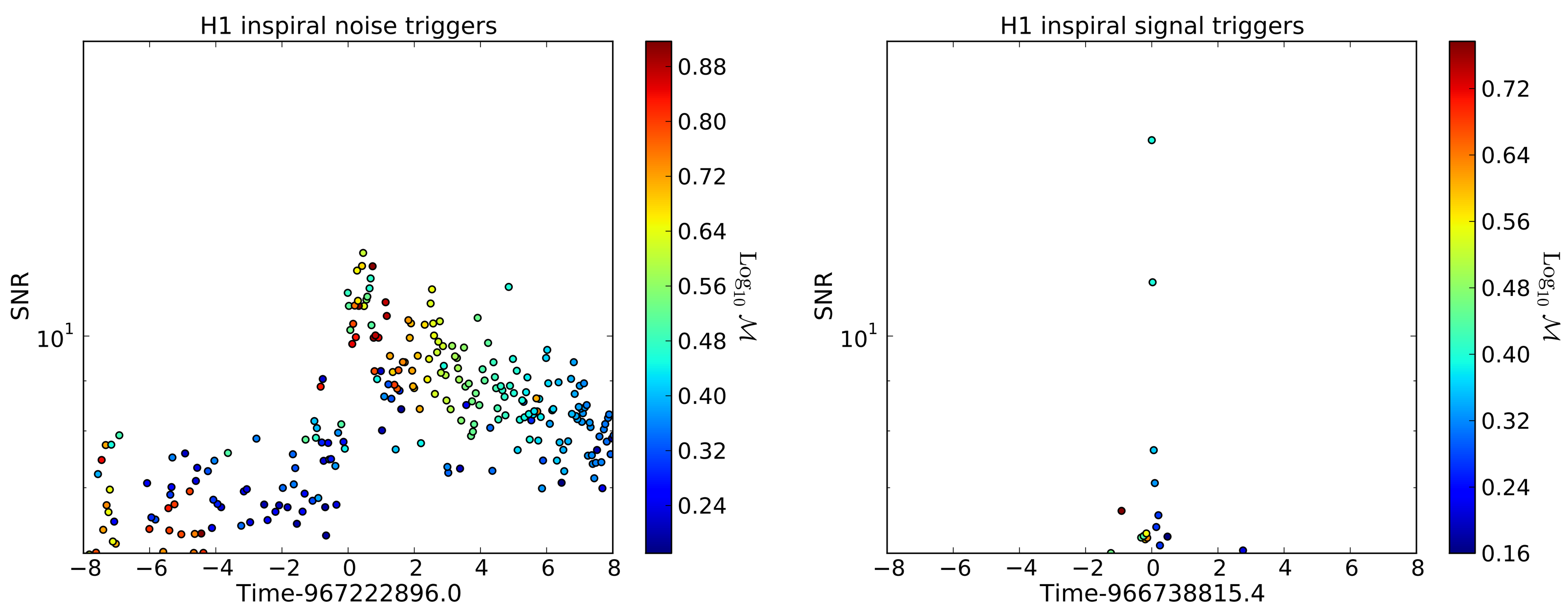



Figure 1: Inspiral triggers around a high-$\hat{\rho}$ noise event (left) and simulated signal event (right). Excess triggers produced over a time window in glitchy data are clearly distinguishable from the 'clean' signal.

## Omicron Triggers

- Triggers generated using sine-Gaussian templates and "clustered" together into tiles if occuring within a short enough time interval.
- For simulated inspiral signals, high-SNR tiles fall along the Newtonian inspiral time-frequency track; for noise events, the tiles' time and frequency have more random scatter – see Figure **??**.
- Information provided by Omicron triggers can contribute to more accurate classification of candidate inspiral events.

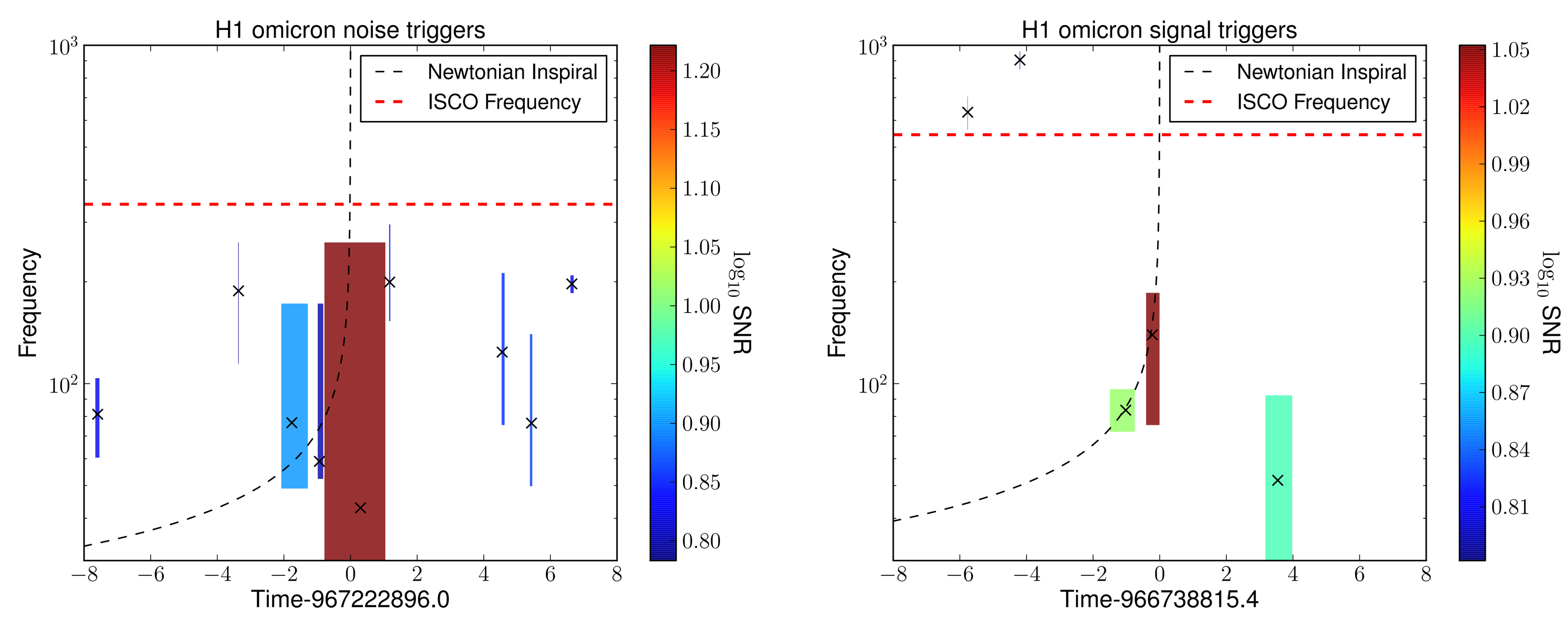



Figure 2: Clustered Omicron triggers (tiles) around the same noise (left) and signal (right) events as in Figure **??**. The dimensions of each rectangle give the duration and bandwidth of the clustered tile. The x-mark shows the frequency and time of the loudest trigger in each cluster.

## Classification Features

Features are constructed from individual events and their neighbouring triggers, that could help discriminate between signals and glitches:

- Inspiral SNR and chirp mass of candidate event
- Number of nearby inspiral and Omicron triggers
- Individual properties of loudest nearby inspiral and Omicron triggers
- Time lag between loudest nearby Omicron trigger and Newtonian inspiral track

## The Random Forest (RF) Classifier

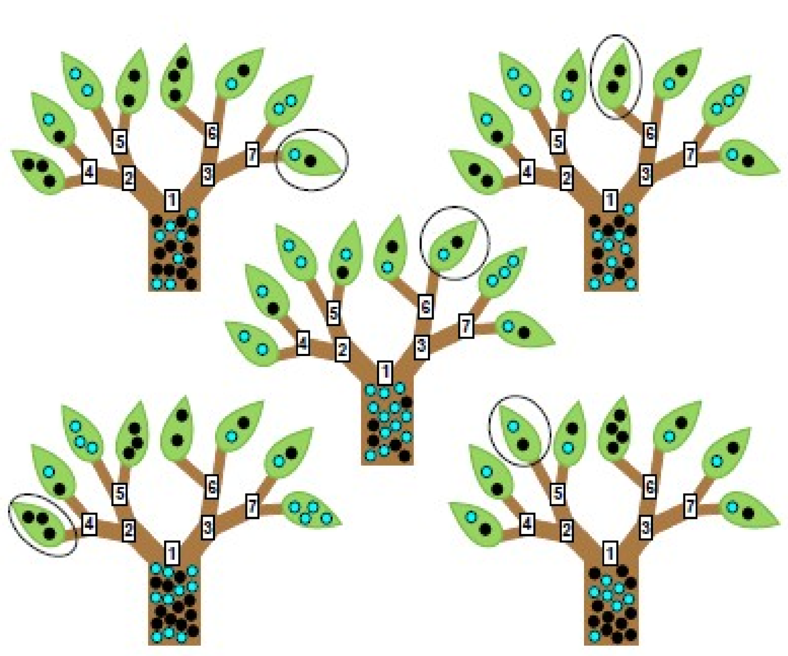


Figure 3: The RF contains many decision trees, each trained on a randomized subset of noise and signal events. An event given to the RF for classification goes from root node to leaf node on each tree, following a path determined by the splitting condition imposed at each node. The RF assigns a ranking score to the event by computing the distribution of training events, in the leaves where the event is placed, over all trees. Picture taken from [**?**].

## Results: Varying RF parameters

- We use a 2-fold 'round robin': splitting the data into two parts, training the RF on one part and evaluating scores for the other.
- Assign false alarm probability (FAP) values to simulated signals by comparing their scores to noise events and compute a weighted sum over simulated signals as a measure of the expected number of detections.
- Errors estimated over 5 realizations of random 2-fold splitting.

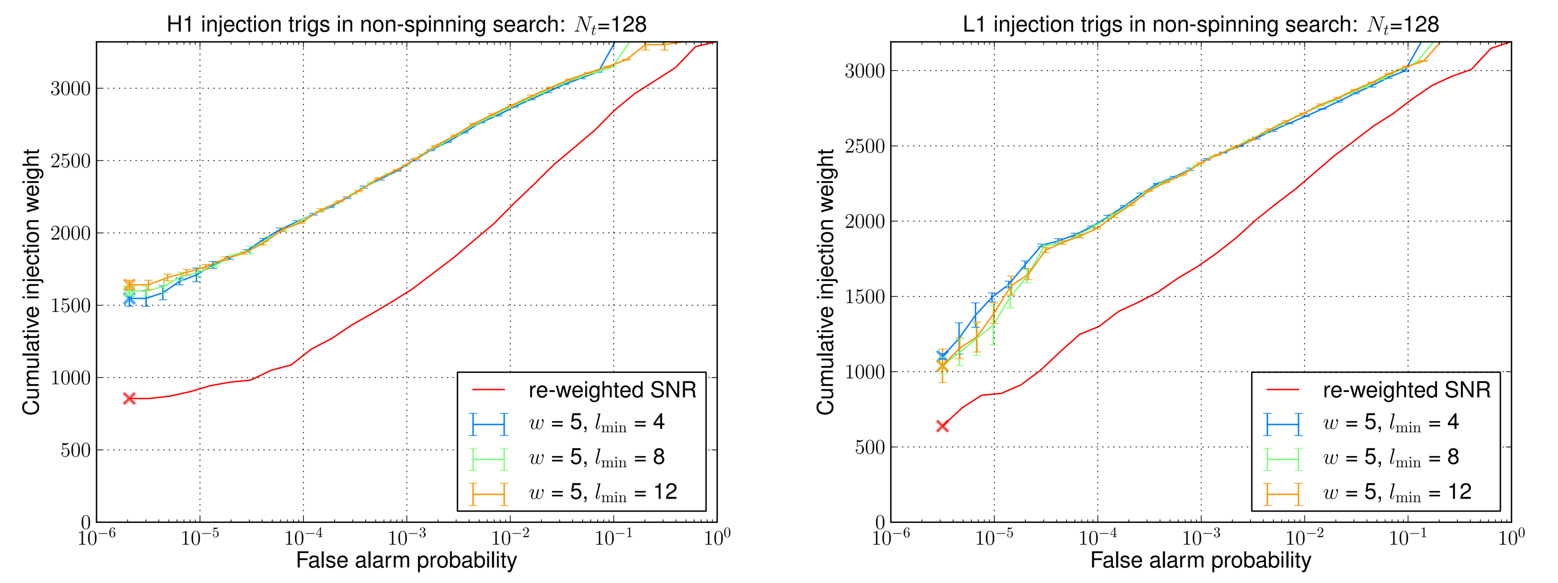



Figure 4: Performance of the classifier for different RF parameters and for data sets from different detectors. The classifier is able to detect a consistently higher number of simulated signals above noise, as compared to ranking by re-weighted SNR.

## Results: Adding $\chi^2$ as a feature

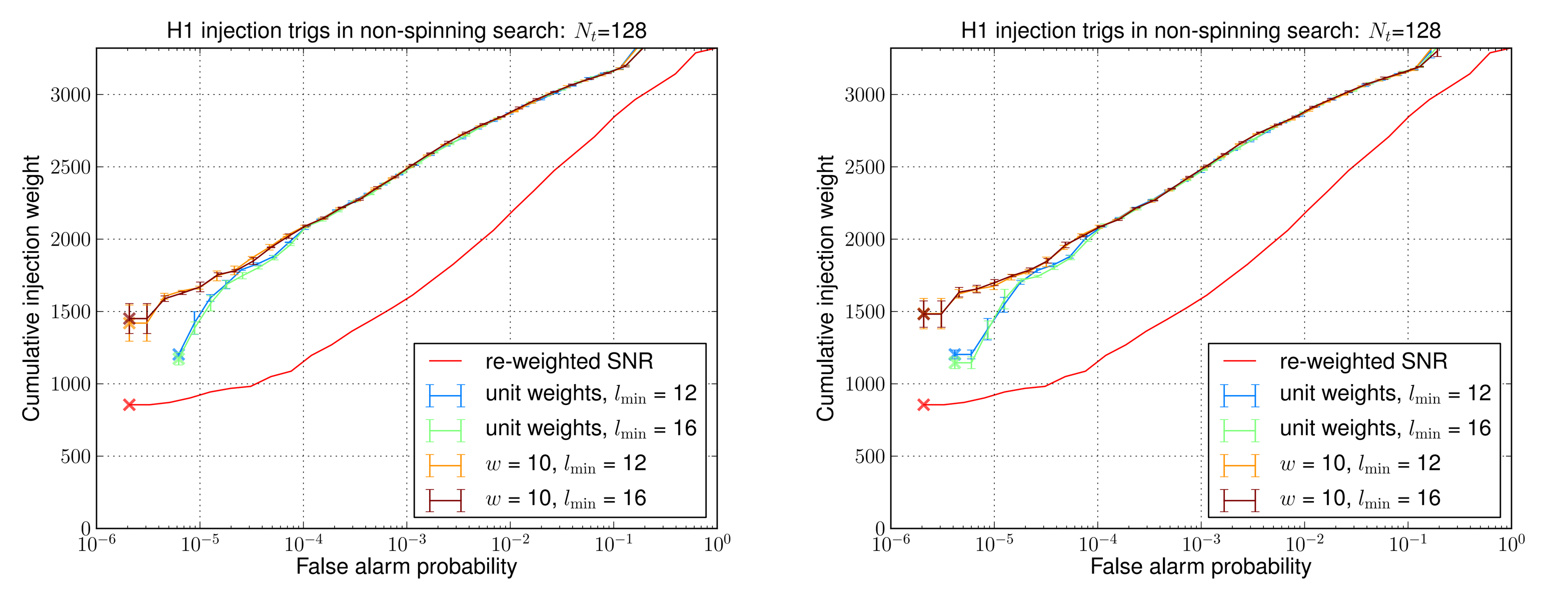



Figure 5: With (left) and without (right) $\chi^2$: Adding $\chi^2$ values to the classifier does not yield further increases in the expected number of detections.

## Summary and Outlook

- Using a multivariate classifier based on features from multiple inspiral and sine-Gaussian triggers around candidate events, we find an almost 2-fold increase in the expected number of inspiral signals recovered at low FAP, in glitchy data, as compared to ranking via re-weighted SNR.
- Features we have identified are able to substitute for and improve on the $\chi^2$ test, for spinning signals recovered with non-spinning templates.
- Currently the classifier is implemented for single-detector clustered inspiral triggers; more development is required to extend it to a multi-detector analysis.

## References

[1] B. Allen, Phys. Rev. D **71**, 062001 (2005)

[2] S. Babak *et al*, Phys. Rev. D **87**, 024033, (2012)

[3] S. Chatterji, Ph.D. thesis, Massachussetts Institute of Technology, (1995)

[4] T. Dal Canton *et al*, Phys. Rev. D **90**, 082004, (2014)

[5] P. T. Baker *et al*, Phys. Rev. D **91**, 062004, (2015)