PAPER

QoQ: a Q-transform based test for gravitational wave transient events

To cite this article: Siddharth Soni *et al* 2024 *Class. Quantum Grav.* **41** 015012

View the article online for updates and enhancements.

QoQ: a Q-transform based test for gravitational wave transient events

Siddharth Soni[1,][*] , **Ethan Marx**[1]**, Erik Katsavounidis**[1]**,**
Reed Essick[2,3,4] **, G S Cabourn Davies**[5] **, Patrick Brockill**[6]**,**
Michael W Coughlin[7] **, Shaon Ghosh**[8] **and Patrick Godwin**[9]

[1] LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139,
United States of America
[2] Canadian Institute for Theoretical Astrophysics, Toronto, ON M5S 3H8, Canada
[3] Department of Physics, University of Toronto, Toronto, ON M5S 1A7, Canada
[4] David A. Dunlap Department of Astronomy, University of Toronto, ON M5S 3H4,
Canada
[5] University of Portsmouth, Portsmouth PO1 3FX, United Kingdom
[6] University of Wisconcin–Milwaukee, Milwaukee, WI 53201, United States of
America
[7] School of Physics and Astronomy, University of Minnesota, Minneapolis, MN
55455, United States of America
[8] Montclair State University, 1 Normal Ave, Montclair, NJ 07043, United States of
America
[9] LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125,
United States of America

E-mail: siddsoni@mit.edu

Abstract
The observation of transient gravitational waves (GWs) is hindered by the presence of transient noise, colloquially referred to as glitches. These glitches can often be misidentified as GWs by searches for unmodeled transients using the excess-power type of methods and sometimes even excite template waveforms for compact binary coalescences while using matched filter techniques. They thus create a significant background in the searches. This background is more critical in getting identified promptly and efficiently within the context of real-time searches for GW transients. Such searches are the ones that have enabled multi-messenger astrophysics with the start of the Advanced LIGO and Advanced Virgo data taking in 2015 and they will continue to enable the field for further discoveries. With this work we propose and demonstrate the use of a signal-based test that quantifies the fidelity of the

---

[*] Author to whom any correspondence should be addressed.

time-frequency decomposition of the putative signal based on first principles on how astrophysical transients are expected to be registered in the detectors and empirically measuring the instrumental noise. It is based on the Q-transform and a measure of the occupancy of the corresponding time-frequency pixels over select time-frequency volumes; we call it 'QoQ'. Our method shows a 40% reduction in the number of retraction of public alerts that were issued by the LIGO-Virgo-KAGRA collaborations during the third observing run with negligible loss in sensitivity. Receiver Operator Characteristic measurements suggest the method can be used in online and offline searches for transients, reducing their background significantly.

Keywords: transient noise, gravitational wave, data quality tool, signal vs noise classifier

1. Introduction and motivation

The direct detection of gravitational waves (GWs) by ground-based interferometric detectors, Advanced LIGO (aLIGO) [1] and Advanced Virgo (AdV) [2] has burgeoned the field of GW physics [3]. Together with that, the first detection of a binary neutron star coalescence by LIGO-Virgo in association with a Gamma-Ray Burst [4] has also inaugurated a new era in multi-messenger observations enabled by GW observations. Three observing runs of the international network of GW detectors have already taken place starting in 2015 and until 2020. They are referred to as 'O1', 'O2' and 'O3'; they resulted in about 100 GW observations following thorough analyses of data conducted over these years [5–13]. These GW signals are the result of collisions between compact binary objects such as black holes and neutron stars. Additionally, during these observing runs, real-time searches were also conducted by the collaborations to disseminate transient event detection candidates to the broader astronomy community in order to enable their multi-messenger observations [14]. During the O3 run, 80 public alerts for GW candidates have been distributed via the general coordinates network (GCN)[10] Out of these, 24 were retracted on timescales of hours to days following their initial release, primarily reflecting manual examination of the events by the LIGO-Virgo Collaboration and deeming them to be of non-astrophysical origin [7, 8, 15]. Quantifying and automating the retraction process for this kind of event in upcoming observing runs of the LIGO-Virgo-KAGRA (LVK) network of detectors has been the primary motivation and goal of this work [16].

The output of ground-based GW detectors is strain amplitude, also referred to as $h(t)$. This captures with the highest sensitivity the differential change in length of the kilometer-long arms of the interferometer. The $h(t)$ is impacted by multiple sources of noise which can be broadly classified into two categories, short duration non-Gaussian transients, also known as *glitches*, and noise that behaves in a Gaussian and stationary fashion over a given time and frequency range [17–20]. These transients can adversely impact GW searches both at the detection confidence level as well as in their source parameter estimation, including localization. [21–28]. Aside from their impact to searches, identification of transient noise may lead to studies on their coupling mechanism in the detector and subsequent hardware changes that can help remove or at least reduce such noise sources [29].

---

[10] https://gcn.nasa.gov.

Identifying transient GW candidate events utilizes the $h(t)$ time-series. It is performed by a variety of transient-finding search pipelines developed by the LVK collaboration, as well as the broader community. In this paper, we focus on low-latency searches [30–34]. These searches are primarily responsible for producing public GCN alerts in close to real time in order to allow for their multi-messenger follow-up. LVK collaboration also performs searches offline [12, 32, 35, 36] with refined versions of their online pipelines. Such refinement allows for, among other things, improved sensitivity, better noise rejection, although it often comes with higher computational complexity and slower turn-around times. The broader GW community outside the LVK also analyses the interferometric data once they become publicly available, with redundant and complementary results obtained. In the typical searches for compact binary coalescences (CBCs), the waveform of such signals can be modelled using post-Newtonian and numerical relativity methods, thus resulting in a family of signals spanning a space parameterized by masses and spins of the binaries; this is referred to as *template banks* [37]. The various search pipelines employ what is known as *a matched filter* to look for GW signals in the data, by matching the incoming strain with waveforms in such template banks [32, 36, 38]. However, the presence of noise transients in the data complicates this process as they may mask and sometimes mimic a true GW signal [4]. To remedy this, the search pipelines use a χ^2 test to differentiate the time-frequency distribution of power in real GW signals and noise transients [39, 40]. The candidates are assigned a false alarm rate (FAR) and GW Candidate Event Database (GraceDb[11]) alerts are sent for events below a certain FAR threshold[12]. Even with all these precautions, transient noise can trigger a pipeline alert for a GW candidate event, which consequently has to be retracted. The decision to retract an event is carried out via human intervention with minimal quantitative analysis.

In this article, we present a tool to distinguish binary black hole (BBH) signals from transient noise based on the energy distribution in the signal across the time-frequency plane, as obtained via the Q-transform [41]. In section 2, we discuss events retracted during O3. In section 3, we look at the energy distribution of Gaussian and transient noise and describe in detail the Q-occupancy (QoQ) test. In section 4, we discuss its application on O3 low latency astrophysical candidates and retracted events. In section 5, we discuss our QoQ test analysis on the mock data challenge (MDC) Injection Data Set and PyCBC background events. In section 6, we extend the analysis of section 4 to events found in offline analysis in O3. Finally we conclude and summarize in section 7.

2. O3 retracted events

During the third Observing run (O3), a total of 80 events were identified in low latency, of which 23 were later retracted as their origin was found to be environmental or instrumental rather than astrophysical [7, 8, 15]. An additional event was retracted as the alert was sent out due to an error[13] [7, 8, 42–44]. Offline re-analysis of the O3 data refined the admission criteria of astrophysical events, including the requirement of the probability of an event being astrophysical to be greater than 0.5 and having a FAR of less than 2 per day; this led to the GWTC-3 catalog [8]. Of the non-retracted 56 events identified in low latency, 44 events were also identified by the offline re-analysis and thus included in the GWTC-3 catalog [8]. A long-term solution to reduce the number of retractions is to identify the transient noise coupling

---

[11] https://gracedb.ligo.org.

[12] https://emfollow.docs.ligo.org/userguide/content.html.

[13] https://gracedb.ligo.org/superevents/S190405ar/view/.

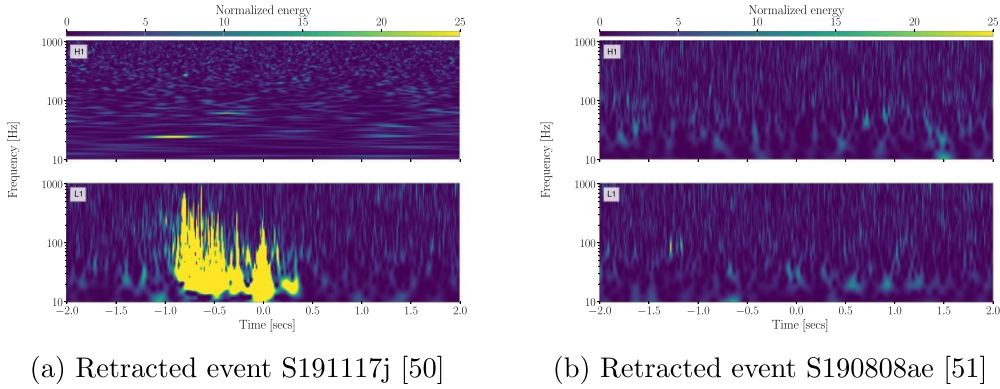(a) Retracted event S191117j [50]   (b) Retracted event S190808ae [51]

Figure 1. Q-transform visualizations of example retracted events from O3. Left: L1 data has a loud glitch at the time of event S191117j [49]. Right: The retracted event S190808ae do not contain any visible presence of transient noise.

in the detector and remedy it by making hardware changes [29]. During O3b, after the reaction chain tracking was employed at LIGO Livingston Observatory (LLO) and LIGO Hanford Observatory (LHO) to diminish the slow scattering transient noise, the average fraction of retracted O3b public alerts dropped from 0.55 to 0.21 [8]. However, reduction of transient noise is only sometimes possible as the noise originates in the complex instrumentation and environment of the detector, and new categories of transient noise may always show up as the sensitivity of the detector improves and new subsystems get added [45–48]. So, a need for prompt, quantifiable, and automated ability to address noise transients and their impact on public alerts from the LIGO-Virgo-KAGRA network remains.

Retracted events during O3 can be broadly classified into two categories: events with clear evidence of loud transient noise in one or more detectors and low SNR events with almost no visible presence of transient noise. Figure 1 shows an example from each category. About half of the O3 retractions belong to the first category. We aim to separate these false positives from astrophysical transient candidate events. The QoQ test explained in section 4 later on in the paper is a method for distinguishing transient noise from astrophysical events using the Q-transform and the time-frequency pixel-energy distribution measure that it provides. There has been a growing interest among researchers in solving such problems [51–53].

3. The Q-transform in GW data analysis

A commonly used analysis tool for identifying and characterizing transients both in the GW strain channel and in the wealth of auxiliary channels recorded at the observatory is the Q-transform [54, 55]. The auxiliary channels are data streams that record various environmental and instrumental variations. It is a modification of the short Fourier transform that covers the time-frequency plane with pixels of constant Q (quality factor) by invoking analysis windows with durations that vary inversely proportional to the frequency. It can be shown that it is equivalent to a matched filter search for minimum uncertainty waveforms [41, 56]. A search for GW bursts using the Q-transform was first implemented in the early science runs of initial LIGO [41, 56]. The implementation of the Q-transform as an unknown-morphology transient-finding method generally tiles the time-frequency plane using a range of quality factors Q. We will
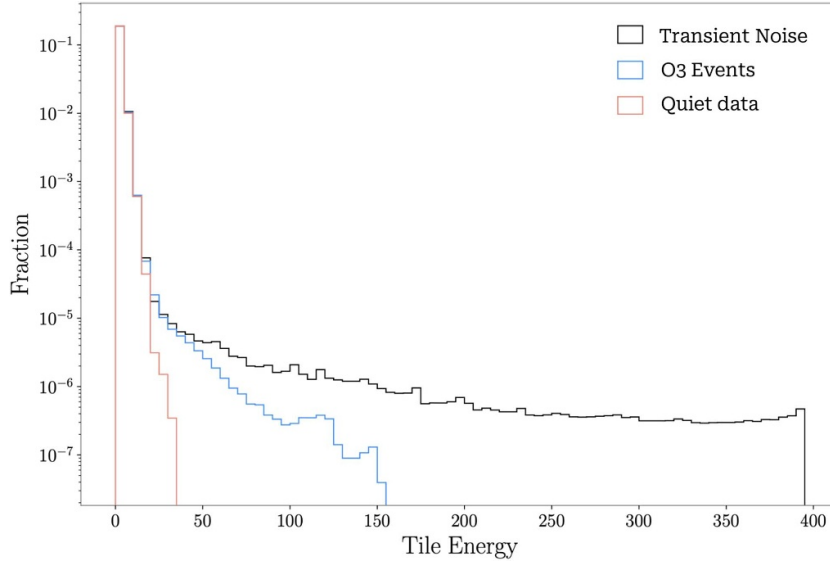
Figure 2. Q-transform pixel-energy distribution comparison of detector data around noisy, O3 catalog, and quiet times in the LIGO detectors. 'Noisy' times are identified by *Omicron* as being subject to a glitch while 'quiet' times are selected via randomly sampling times of the LIGO detectors data taking over which no glitch by *Omicron* was identified. As compared to quiet times, noise transients or short duration glitches appear in the tail of the pixel-energy distribution.

refer to these tiles as *pixels*. The squared magnitude of their (discrete) Q-transform coefficient is a measure of the signal energy, and we will refer to it as *pixel energy*. For stationary noise, such pixel energy at a given frequency is exponentially distributed. Transients are generally identifiable as outliers in the pixel energy distribution via invoking a threshold and some clustering algorithm in order to achieve a desired FAR on the assumption of Gaussian (or otherwise directly measured from the data) noise.

This can be seen in figure 2 where we histogram the energy of all the pixels generated from a 4 s long Q-transform of 16 'noisy' and 16 'quiet' times of the LIGO detectors. The noisy times refer to ones identified by the *Omicron* [57] method as having a glitch of signal-to-noise (SNR) ratio equal to or greater than 6 within such time-window, while 'quiet' times have been randomly sampled from the LIGO detectors data taking corresponding to livetime over which no glitch has been identified by *Omicron*.

3.1. Pixel occupancy

Our proposed test for GW transient events uses as a discriminant the percentage of pixels within a few time-frequency volumes that exceed pre-set values of pixel energy. We refer to this quantity as 'occupancy' or 'density' of pixels, thus the name of our method as 'Q-occupancy', or *QoQ*. Astrophysical transients of known morphology, including CBCs, have numerically calculable occupancy values (given the astrophysical source parameters). For those of unknown morphology, it is *a priori* impossible to discriminate them from non-Gaussian noise

artifacts via a single *QoQ* measurement. In this case, in order to utilize *QoQ*, some basic (or non) assumptions need to be made on the physics and the detection of the (unknown morphology) astrophysical signal one is after. Astrophysical signals are generally louder than Gaussian noise but usually not as loud as the glitches responsible for the retracted events. As shown in figure 2, the pixel-energy distribution of the true events extends beyond the Gaussian noise but not as far as the transient noise. And the kind of noise artifacts in GW interferometers we are primarily gearing our method to identify has an extended structure in the time-frequency plane and often exceeds pixel energies expected from astrophysical populations we currently know [17]. This is learned by the noise data the instruments record and its implications to searches is assessed via standard confusion matrix and receiver operating characteristic (ROC) analysis under some assumption on the astrophysical signal in consideration. This analysis is carried out in section 5.

In this work we utilized the *GWpy* [58] and *Omicron* [57, 59] implementation of the Q-transform. We use the quality factor (*Q*) values between 4 and 64 to logarithmically tile the time-frequency plane. Analysis of the *h(t)* time series via the Q-transform results in 'triggers' corresponding to transients present in the data. Such triggers are described with a few parameters such as the central event time, central frequency, quality factor *Q*, and SNR; we will use these for the main feature description of the transients we will be analyzing. Additionally, the full time-frequency decomposition with all pixel energy values across resolutions is made available and is what we use for deriving the *QoQ* quantities.

Several options exist on how to define the time-frequency volume over which to measure *QoQ*. We first need to pick a threshold on the pixel-energy at or above which we will consider the pixels as 'loud' and thus counted toward the *QoQ*. We generally need to pick values above the Gaussian noise level, although the exact threshold has to be informed by the character of the noise and the corresponding impact on the misclassification of signals from astrophysical populations. The overall time-frequency volume over which *h(t)* is analyzed can vary significantly depending on the astrophysical search within which noise rejection via this method is being implemented. In this analysis, we will focus on transient GWs from BBH mergers typically lasting from tens of milliseconds to a few seconds. Extension of the method to other astrophysical populations is straightforward.

We use a duration of 4 s set by the scale of BBH events for the Q-transform and restrict the analysis to the frequency band of 10–1024 Hz. The frequency span of the search is motivated by the overall frequency content of noise artifacts in aLIGO and AdV as well as that of the BBH and CBC—in general—signals. This Q-transform is then sub-divided into *n* (disconnected, i.e. mutually exclusive) time-frequency regions spanning the original time-frequency volume. For each of these *n* regions, the percentage of pixels above a given pixel-energy threshold is computed. We will refer to this percentage as the *QoQ* value. If the *QoQ* exceeds a certain threshold in a given region, we use this to infer that the corresponding region is contaminated with noise. This procedure is shown as a flow chart in figure 3.

The method allows for a different *QoQ* threshold to be used in each of the *n* sub-divisions of the time-frequency volume one starts with. In our approach, currently, any one of such *n* sub-divisions exceeding the *QoQ* threshold flags the event for further inspection. Clearly, more complicated logic can be implemented, making certain assumptions for the hypothetical astrophysical signal as well as for the noise behavior that may be targeted via such an approach. There is an additional conceptually different approach in the sub-division of the original time-frequency volume, namely in *m* overlapping rather than disconnected regions. We invoke this approach as well since, as we will show below, it offers a unique handle in identifying astrophysical events that may be overlapping with noise artifacts.
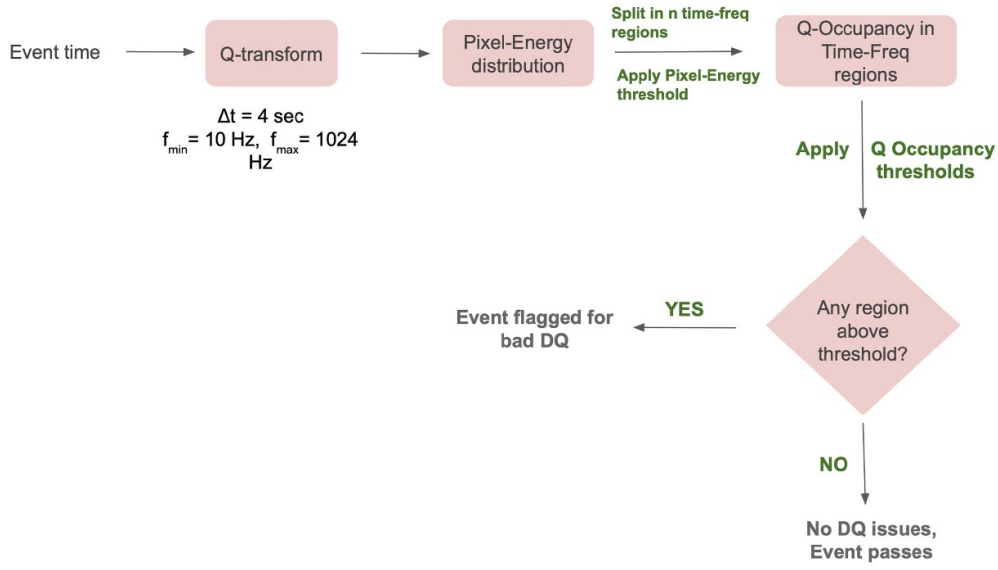
Figure 3. Flowchart of the QoQ test. Given an event time, the tool makes a 4 s long Q-transform. We then split the pixel-energy of this Q-transform into multiple time-frequency regions and apply a pixel-energy threshold. This gives us, for each time-freq region, the percentage of pixels above the pixel-energy threshold i.e. its Q-occupancy. The next step is to apply a QoQ threshold on each of these regions. If any of these time-freq regions are above its QoQ threshold, the event is flagged for bad data quality, otherwise it passes the QoQ test.

To decide the values of pixel-energy and QoQ thresholds to be used in the QoQ test, we plot the cumulative distribution of QoQ values in the 1 s window surrounding the event time of O3 low latency candidates and O3 retracted events. Figure 4 shows this QoQ distribution of O3 low latency candidates for different pixel-energy thresholds and of O3 retracted events for a pixel-energy threshold of 60. As we can see from this figure, in the case of O3 online candidates, for a pixel-energy value of 60, the maximum QoQ value for the central 1 s window is 2.0, we use these values as our thresholds for the pixel-energy and QoQ, respectively for the QoQ analysis on O3 online candidates next.

4. O3 online candidates analysis

We apply the QoQ test on the O3 online candidates. This set includes 56 non retracted and 23 retracted events. We first perform the overlapping and non overlapping analysis individually in sections 4.1 and 4.2 respectively and then combine the two analyses to obtain the results shown in table 1.

4.1. Overlapping time windows and O3 online candidates analysis

The frequency axis in the Q-transform shown on the left in figure 5 is divided into three bands, $(10 - 100 \text{ Hz})$, $(100 - 512 \text{ Hz})$, and $(512 - 1024 \text{ Hz})$ and the time axis into three windows $(-0.5 \text{ s}, 0.5 \text{ s})$, $(-1.0 \text{ s}, 1.0 \text{ s})$ and $(-2.0 \text{ s}, 2.0 \text{ s})$. The intersection of these time and frequency windows gives us a total of 9 regions. For each region, we calculate the percentage of pixels in
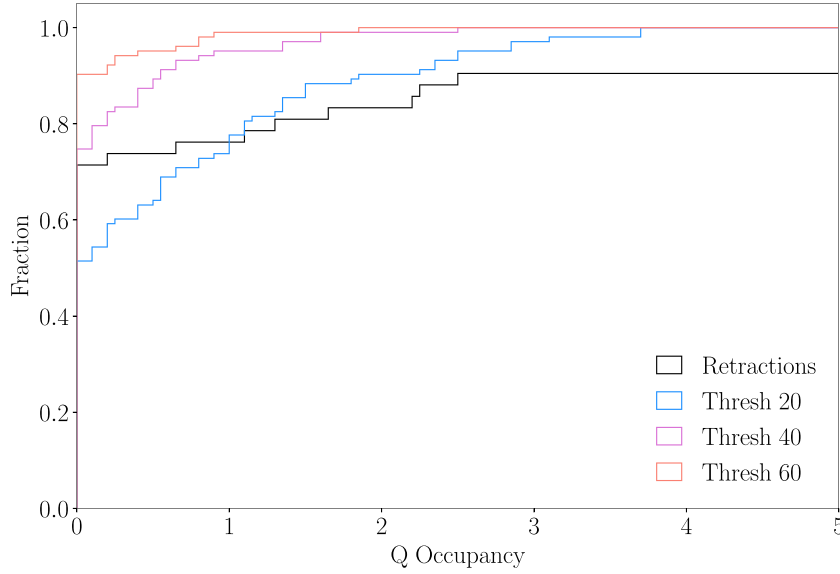
Figure 4. Q occupancy distribution of 1 s window surrounding the event time of all the O3 online public alerts and O3 retracted events, for different values of pixel-energy thresholds. The retracted events with clear presence of transient noise have much higher Q occupancy values compared to the astrophysical candidates. For the O3 online candidates, the max Q occupancy is below 4, for the retractions the distribution extends to Q occupancy of much higher than 5 shown here.
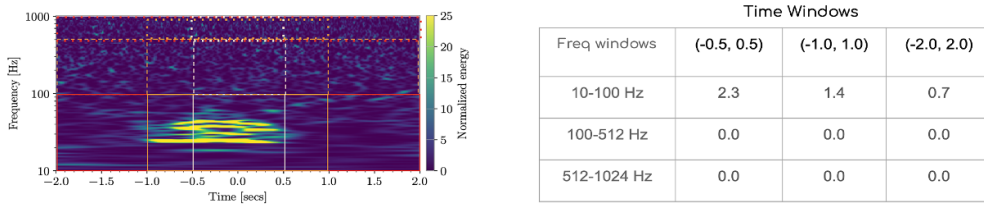

Time Windows			
Freq windows	(-0.5, 0.5)	(-1.0, 1.0)	(-2.0, 2.0)
10-100 Hz	2.3	1.4	0.7
100-512 Hz	0.0	0.0	0.0
512-1024 Hz	0.0	0.0	0.0

Figure 5. Left: A Q-transform of H1 data around gpstime 12 601 74 466 and divided into 9 time frequency blocks. These blocks which are 1, 2 and 4 s long are centered at $t = 0$ and overlap in time. They cover three frequency bands from 10 to 100 Hz, 100 to 512 Hz and 512 to 1024 Hz. Right: This table shows the percentage of pixels with energy above the pixel-energy threshold of 60 in each of the 9 time-frequency regions.

the given region with a pixel-energy value above the pixel-energy threshold. The percentage for each region are shown on the right in figure 5 for a pixel-energy threshold of 60. For example, the value 2.3 in the first row and first column indicates that for the 1 s long window in the frequency band, 10–100 Hz, 2.3 percent of the pixels have pixel-energy above 60. As shown in figure 4, the QoQ values for GW signals are smaller in comparison, and a suitable threshold can be used to separate the transient noise-generated false alerts from true GW candidate events.

We use the **QoQ thresholds of 2, 1, and 1 for the three time windows** $t_1 = 1$ **s** $t_2 = 2$ **s, and** $t_3 = 4$ **s respectively**. To reduce the chances of flagging a true GW signal, the threshold on the central 1 s window is higher as the majority of BBH signals fall within this. With these QoQ thresholds and a pixel-energy threshold of 60, 9 out of the 23 retractions and 1 of the 56
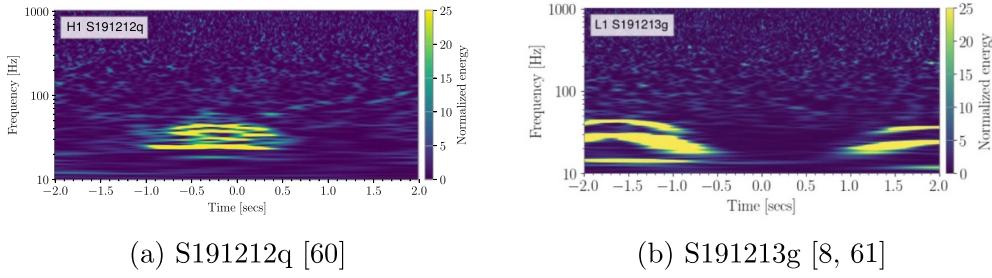
(a) S191212q [60]

(b) S191213g [8, 61]

Figure 6. Left: The H1 data quality around the time of retracted event S191212q [60] is impaired by the presence of Scattering glitch. The transient noise is clearly visible in the Q-transform of H1. Right: The O3 online event S191213g was found with FAR of 1.1 per year in low latency and a public alert was issued. The surrounding data quality is heavily contaminated with scattered light glitches. The O3 offline analysis did not identify this event as a significant candidate [8].
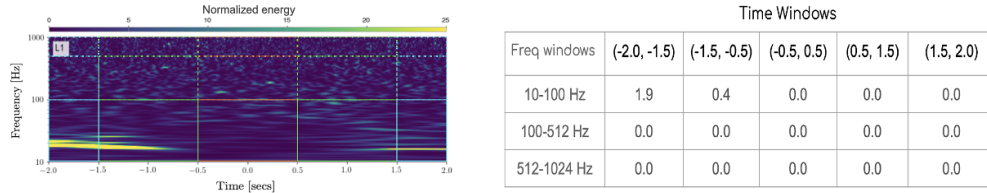

Freq windows	(-2.0, -1.5)	(-1.5, -0.5)	(-0.5, 0.5)	(0.5, 1.5)	(1.5, 2.0)
10–100 Hz	1.9	0.4	0.0	0.0	0.0
100–512 Hz	0.0	0.0	0.0	0.0	0.0
512–1024 Hz	0.0	0.0	0.0	0.0	0.0

Figure 7. Left: A Q-transform of LLO data around the gpstime 12 418 52 074 and divided into 15 time frequency regions. There is no overlap in time among these blocks. The first time window is half a second long and extends from −2.0 to −1.5 s, the next three windows are each 1 s long and the final region is again 0.5 s long and extends from 1.5 s to 2.0 s section. The division along the frequency axes is similar to the overlapping case. Right: The table shows the percentage of pixels with energy above the pixel-energy threshold of 60, for each of these regions. There is some transient activity in the first one and a half second as seen in the Q-transform on left, this is reflected in the table on right.

O3 online transient astrophysical candidates are flagged. Appendix contains more details for each of the flagged events by the QoQ analysis at this stage. Figure 6 shows two O3 online events that our method classifies as noise. More than half of the retracted events do not contain visible glitch activity, and the QoQ test does not flag these events. Out of 23 retractions, 14 were not flagged as noise by the QoQ test as they are low SNR events at one or both detectors.

4.2. Non overlapping time windows and O3 online candidates analysis

Another way to create the time-frequency regions in the Q-transform is with windows that do not overlap in time. Compared to the overlapping analysis, there is no change in frequency delineation, but the time axis is now split into five windows: $(-2.0\,\text{s}, -1.5\,\text{s})$, $(-1.5\,\text{s}, -0.5\,\text{s})$, $(-0.5\,\text{s}, 0.5\,\text{s})$, $(0.5\,\text{s}, 1.5\,\text{s})$ and $(1.5\,\text{s}, 2.0\,\text{s})$. In this case, we get a total of 15 time-frequency regions. And again, we calculate the percentage of pixels with a pixel-energy value above a certain pixel-energy threshold for each region. For example, figure 7 shows the 15 time-frequency regions with no overlap in time and the QoQ of pixels for each region using a pixel-energy threshold of 60.
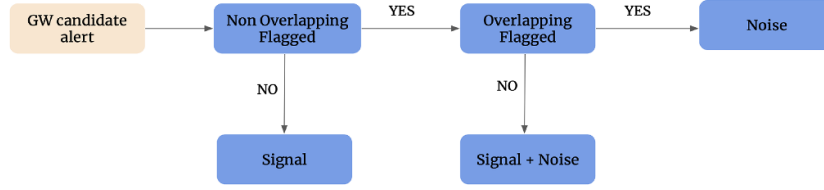
Figure 8. To evaluate the data quality of a GW candidate event, we first feed the event to Non Overlap QoQ test, if the event is not flagged at this stage, it is considered a 'Signal'. However if the event is flagged, it is then analyzed by the Overlap QoQ test. If the event is flagged at this stage, it is considered a 'Noise', otherwise it is classified as a 'Signal with presence of transient noise'.

In many cases, environmental or instrumental noise artifacts adversely impact the data quality a few seconds before or after a GW chirp signal. These events may not be retracted, but they require closer inspection for any noise coupling with the auxiliary channels. The GWs resulting from the merger of BBHs are of the order of a few tens of milliseconds to a few seconds in the sensitive frequency band of ground-based detectors. This means any signal energy outside this time window can be attributed to a non-astrophysical phenomenon. The non-overlapping time windows method can be conveniently used to catch such events. Any events for which the QoQ in the central time window $(-0.5, 0.5)$ is above 2 (same as the overlapping case) or in any of the non-central windows is above 0 can be flagged for stricter event validation. In this qualitative way, the non-overlapping time windows method differs from the overlapping time windows.

The QoQ thresholds used in this case are $[0, 0, 2, 0, 0]$, i.e. events with QoQ **above** 2% **for the central window** ($t3$) **or any non-zero value in any of the non-central windows will be flagged as transient noise**.

The O3 retracted events flagged by this method are the same as those flagged by the overlapping time windows method shown in section 4.1. However, unlike the overlapping case, this method finds the presence of transient noise in three GW candidates found in low latency in O3. Unsurprisingly, none of the central time window (t3) values are flagged since this window is equivalent to the central window (t1) used in the overlapping method. However, the non-central windows flag a total of three O3 online astrophysical candidates. All of these three events are real GW signals polluted by the presence of noise artifacts in their vicinity at one or both detectors. Appendix contains more details on these events. During O3, these events were examined for any potential correlation with the auxiliary channels and underwent a careful process of offline event validation [62, 63]. To summarize, the non-overlapping time windows method effectively catches both false pipeline alerts due to transient noise as well as real astrophysical signals polluted by nearby transient noise.

Based on this analysis, we can use binary decision-making as shown in figure 8 to predict the nature of a GW candidate alert. Suppose an event is flagged by both the overlapping and the non-overlapping time windows analyses. In that case, such an event is likely a noise transient, and caution should be exercised before disseminating the information to the public. If the event is only flagged by the non-overlapping time windows method, then it is likely an astrophysical signal along with some transient noise nearby. Conversely, if an event is not flagged by the

Table 1. Of the 23 O3 retracted public alerts, the QoQ test classified 9 as 'Noise' and 1 as 'Signal + Noise'. Out of the 56 O3 low latency astrophysical events, 3 were classified as 'Signal + Noise', and 1 event was flagged as 'Noise'. A visual inspection supports this classification. Appendix contains more details on these flagged retracted and O3 low latency events.

	O3 online retractions	O3 online nonretracted alerts
Signal	13/23	52/56
Signal + Noise	1/23	3/56
Noise	9/23	1/56

non-overlapping method, then it is likely an astrophysical event. Table 1 shows the results of QoQ test on O3 events found in low latency and O3 retractions at different stages of the flowchart in figure 8.

5. MDC Data and PyCBC background analysis

5.1. MDC data

To more robustly assess the QoQ test response to GW events, we analyze a set of BBH injections. We use the injection set from the O3 MDC. This MDC set contains signals injected in the O3 LIGO Livingston (LLO) and LIGO Hanford (LHO) data between 1 January 2020 and 14 February 2020. The dataset contains injections with source frame masses between $1\,M_\odot$ and $100\,M_\odot$, injected isotropically in the sky. The purpose of these injections is to evaluate the detection efficiency of the search pipelines, perform latency and alert generation checks as well as test source parameter estimation. The MDC injections will thus allow us to perform end-to-end testing of the low latency search pipelines in preparation for O4 [64, 65]. Since the bigger goal of these MDC injections is to check pipeline consistency, the distribution of their source properties is not astrophysical. Compared to the catalog events, there is an excess of low luminosity distance, and high mass events, as shown in figure 9. We will correct for this when we draw conclusions for our QoQ method, as ultimately, we need to assess the performance on the astrophysical population.

To isolate the injection and background data for BBH coalescences, we use the thresholds: $m_1 > 5\,M_\odot$ and $m_2 > 5\,M_\odot$ where m_1 and m_2 refers to primary and secondary mass in the source frame with $m_1 > m_2$. To ensure the high significance of events, we also use injection events with $\rho > 5$ where ρ represents the detector SNR. With these cuts, we get a total of 1354 MDC Injections. Since the MDC waveforms were injected into the O3 data, some of them are polluted by the presence of very high SNR glitches near them. To enhance the purity of this dataset, we removed those injections that contained omicron transients with SNR above 50 within a 3 s window of the injection time. This reduced the size of the injection set marginally from 1354 to 1331 injections. Finally, we apply a threshold on network SNR of $\rho_{\rm net} > 10$ to characterize found injections [66]. This brings the total number of mock injections to 1053. Next, as shown in figure 9 the MDC injection set consists of high mass and low luminosity distance events and does not match the expected astrophysical distribution. To obtain a set that more closely resembles an astrophysical population, we use Importance Sampling described next.
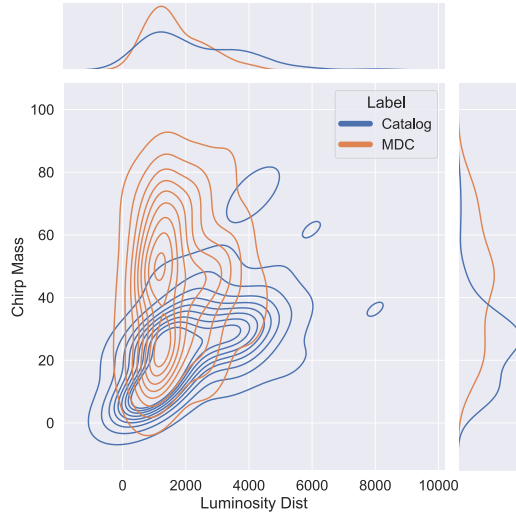
Figure 9. Joint plot distribution of the source chirp mass and source luminosity distance of the Catalog events and the MDC injections. The MDC set contains multiple injections at much lower luminosity distance and much higher chirp mass compared to expected values from an astrophysical population.

5.2. Importance sampling

Say we would like to calculate the expectation value of a random variable $f(x)$, where x is a random vector. Assuming continuous distribution and $p(x)$ the probability density, we can write the expectation value μ as:

$$\mu_f = \int f(x) p(x) \, \mathrm{d}x. \tag{1}$$

Now, let us say we have another distribution defined by the probability density $q(x)$, and we would like to calculate μ_f by drawing samples from this new distribution $q(x)$. Given that $q(x) \neq 0$ whenever $f(x)p(x) \neq 0$, we can take the above equation and modify it as:

$$\mu_f = \int \frac{f(x) p(x)}{q(x)} q(x) \, \mathrm{d}x. \tag{2}$$

The ratio $\frac{p(x)}{q(x)}$ weights the samples from this new distribution. Certain elements from $q(x)$ have a higher impact or they are more 'important' and the weights emphasize these elements while sampling. Essentially, this allows to calculate the expectation value of samples drawn from $q(x)$ by adjusting their importance with respect to the original distribution $p(x)$ [67]. So for samples $X_1 \ldots Xn$ the importance sampling algorithm can be written as:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i) p(X_i)}{q(X_i)}. \tag{3}$$

In our case, the $p(x)$ and $q(x)$ refers to the Catalog and MDC Injections respectively and $f(x)$ is the QoQ test function as shown in figure 8. We use luminosity distance and source

Table 2. This table shows the results on MDC Injections and PyCBC background data for pixel-energy threshold of 60 and QoQ threshold values of 2, 3 and 4 (Thresh A, Thresh B and Thresh C respectively). For QoQ threshold of 2, the test classifies 95% of the Injection data as 'Signal', 2.9 % as 'Signal + Noise' and 2.1 % as just 'Signal'. The lower and upper limits cover the 95 % confidence interval for these values. Figure 10 shows these results for other values of pixel-energy thresholds.

	Injections			Background		
	Thresh A	Thresh B	Thresh C	Thresh A	Thresh B	Thresh C
Signal	$95_{88.9}^{100.0}\%$	$95.6_{89.5}^{100.0}\%$	$96.2_{90.2}^{100.0}\%$	$52.7_{50.2}^{55.1}\%$	$55.1_{52.6}^{57.5}\%$	$57.2_{54.7}^{59.6}\%$
Signal + Noise	$2.9_{2.0}^{3.7}\%$	$2.3_{1.4}^{3.2}\%$	$1.7_{0.8}^{2.6}\%$	$4.2_{3.2}^{5.2}\%$	$4.1_{3.1}^{5.1}\%$	$4.1_{3.1}^{5.1}\%$
Noise	$2.1_{1.5}^{2.7}\%$	$2.1_{1.5}^{2.7}\%$	$2.1_{1.5}^{2.7}\%$	$43.1_{40.6}^{45.5}\%$	$40.8_{38.4}^{43.2}\%$	$38.7_{36.3}^{41.0}\%$

chirp mass to generate the kernel density estimate, and so we can rewrite the above equation as:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{f_{\text{QoQ}}(X_i, Y_i) p_{\text{cat}}(X_i, Y_i)}{q_{\text{mdc}}(X_i, Y_i)}. \tag{4}$$

Here, X_i and Y_i represent the luminosity distance and the source chirp mass respectively of the ith sample. An MDC injection sample whose luminosity distance and source chirp mass closely resembles the respective distributions of Catalog events is assigned a higher weight as opposed to injections that do not. This way, we can sample all the MDC injections and calculate the weight-adjusted expectation value.

To calculate the density estimate, we use a Gaussian kernel:

$$K(x, y) = \frac{1}{2\pi w_x w_y} \frac{1}{n} \sum_{i=1}^{n} exp \left[-\frac{(x - x_i)^2}{2w_x^2} - \frac{(y - y_i)^2}{2w_y^2} \right] \tag{5}$$

where w_x and w_y are the bandwidth values.

5.3. Q-occupancy analysis of BBH data

Each event follows the steps outlined in the flowchart shown in figure 8. For each MDC injection, we first calculate its weight $\frac{p_{\text{cat}}(X_i, Y_i)}{q_{\text{mdc}}(X_i, Y_i)}$. We then apply the QoQ test on this event as outlined in figure 8. Any event that is not flagged by the nonoverlap method is classified as a 'Signal'. Any event that fails to pass the nonoverlap method then goes through the overlap method. If the event is flagged at this stage, it is then classified as 'Noise', otherwise, it is classified as 'Signal + Noise'. We use a total of three sets of QoQ thresholds, Thresh A, Thresh B and Thresh C. The non overlap and overlap Q occupancy threshold values for Thresh A, Thresh B and Thresh C are $([0, 0, 2, 0, 0], [2, 1, 1])$, $([0, 0, 3, 0, 0], [3, 1, 1])$ and $([0, 0, 4, 0, 0], [4, 1, 1])$ respectively. The percentage of events flagged at each stage of figure 8, for each set of QoQ thresholds is shown in table 2.

Ideally, we would not misclassify a true GW signal as noise. There are multiple reasons why a small percentage of MDC injections are classified as noise. These chirp signals were injected into the actual O3 data from LLO and LHO. And so, a number of these signals are polluted by the presence of noise transients near them. The Q-transforms of the injection events that are flagged by the non-overlap method but not by the overlap method reveal the presence of nearby transient noise along with the chirp morphology.
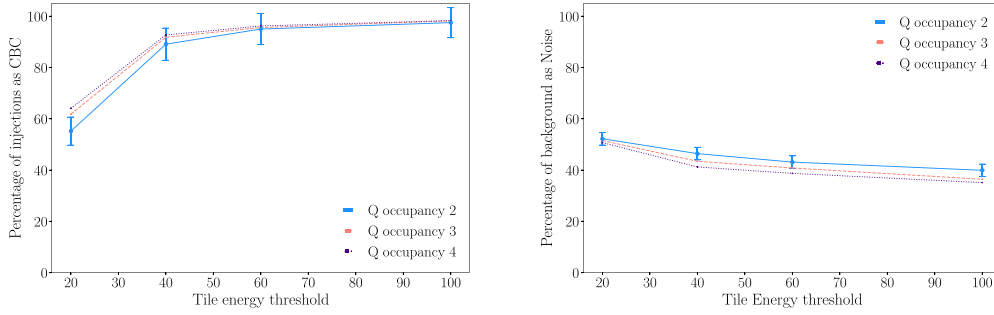
Figure 10. Here we show the variation in MDC Injections classified as 'CBC' and PyCBC background classified as 'Noise' as we change the pixel-energy threshold and Q-occupancy threshold values. The 95 % confidence interval error bars are shown for Q-occupancy threshold of 2. Left: As we increase the pixel-energy threshold, a larger fraction of injections are classified as 'CBC'. However, after pixel-energy threshold of 60, this change is rather small. Right: As expected, increasing pixel-energy threshold results in smaller percentage of background events classified as 'Noise'.
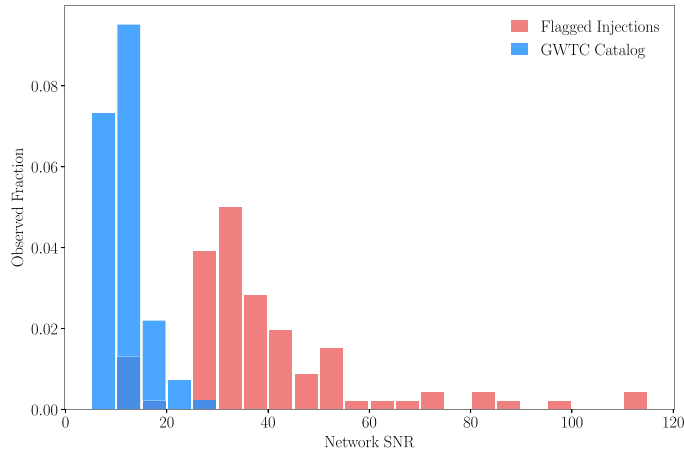

Figure 11. About 2% of the MDC injections were classified as 'Noise' by the QoQ test. This figure shows the network SNR distribution comparison between these injections flagged as 'Noise' and Catalog events. The histogram is normalized so that the area under each distribution sums to 1. These flagged injections have a much high network SNR, not expected from astrophysical signals.

For the Thresh A case, about 2.0% of the injection events are classified as noise. Since these events come from the injection set, this value represents the false positive in our analysis. Figure 11 compares the Network SNR of these flagged injections with the GWTC Catalog. As we can see from this comparison, majority of these flagged MDC injections are very loud, and their SNR distribution does not resemble the distribution expected from a set of astrophysical events [68]. The importance sampling weight of all these events is below 1. Furthermore, figure 12 shows the Q-transforms of the three categories of injections flagged at different stages of the QoQ test ('Signal', 'Signal + Noise', and 'Noise').
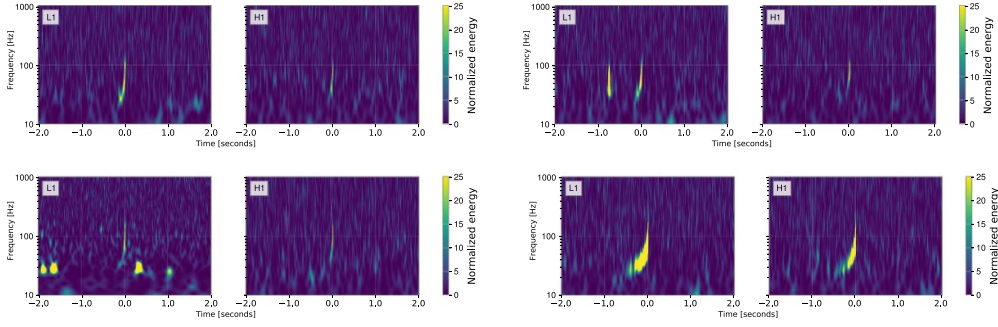
Figure 12. Q-transforms of BBH injections analyzed at different stages in the figure 8. Top left: An MDC injection classified as 'Signal' by the QoQ test. This event is not flagged by the non overlapping analysis. About 95% of the MDC injections are classified as 'Signal' by the QoQ test. Top right and Bottom left: These two injections are flagged by non overlapping analysis but not by overlapping, and are classified as 'Signal + Noise'. For both of these injections, we can observe the presence of transients in the L1 data. About 3% of the MDC injections are classified in this category by the QoQ test. Bottom right: This injection is classified as noise due to its very high energy content. About 2% of the MDC injections were classified as 'Noise' by the QoQ test.

5.4. PyCBC background

In order to assess the QoQ's test ability to reject false pipeline alerts, we analyze a set of background events generated by PyCBC [35, 38]. PyCBC is a matched filtering pipeline that identifies triggers by finding peaks in the matched filter SNR between the data and a template bank of waveforms. Events are then constructed by matching coincident sets of these triggers in the detector network. Each event is compared to a background to assess the FAR.

Background events are constructed through time shifts, where the triggers from one detector are shifted in time by more than the GW travel time compared to other detectors, breaking the astrophysical coincidence. The background events we use come from the PyCBC-broad analysis of the first half of the third LVK observing run, O3a. These analyses were used to produce results in GWTC-2.1 [7]. We use the PyCBC 'exclusive' background only, meaning any triggers which appear in the un-shifted coincident sets of triggers are not used, ensuring the background is not contaminated by potential astrophysical signals. These background events mimic possible false coincidences that could appear during an online observing run, providing a useful data set to evaluate the QoQ test.

To obtain the background data for BBH coalescences, we use similar thresholds on primary and secondary masses as for MDC injections. And to ensure the high significance of events, we use background events with *ifar* > 50 years where ifar is the inverse FAR. With these cuts, we get a total of 1591 PyCBC background coincidences. The importance sampling is not required for the background data since they are not astrophysical events and are not expected to match the astrophysical distribution of source properties.

5.4.1. Analysis of BBH Background flagged as noise. Depending on the Q-occupancy threshold, table 2 shows that the test flags between 39% and 43% of the background events as noise, while between 53% and 57% are classified as 'Signals'. As mentioned earlier in section 2, O3 retractions events fall into two main categories: events with a clear presence of transient noise and low SNR events with no visible transient noise present. The majority of the
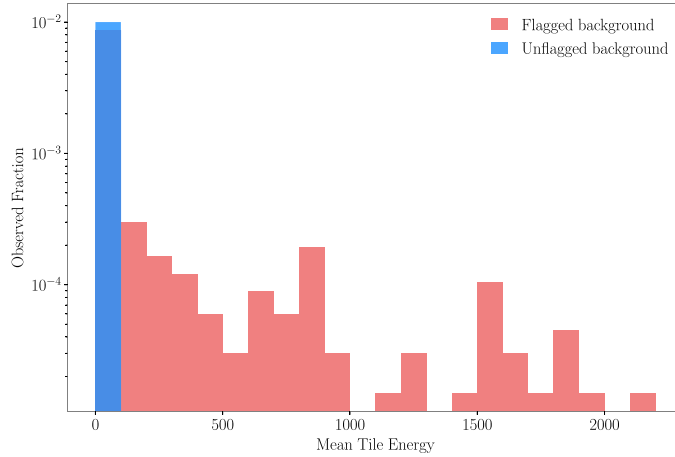
Figure 13. A comparison of distribution of mean pixel energies between unflagged background events and background events flagged as noise by the QoQ analysis.

background coincidences not flagged by the Q-occupancy test belong to this latter category of events. Figure 13 shows pixel-energy comparison of flagged and unflagged background coincidence. The Q scans of these unflagged background coincidences do not contain considerable noise above the pixel-energy threshold of 60 and thus were not picked up by the Q-occupancy test. However, since the background events have *ifar* > 50 years, these are considered highly significant events by the PyCBC search pipeline. With a Q-occupancy threshold of 2 and pixel-energy threshold of 60, the QoQ test flagged 43% of BBH background as 'Noise'. This is highly consistent with the amount of O3 retractions classified as 'Noise' by the QoQ test, as shown in table 1. Given a comparable rate of transient noise, this suggests that we can expect QoQ test to flag similar percentage of false positives in O4.

6. O3 offline analysis

In table 1 we compare the results of the QoQ test on O3 retracted events with non-retracted events found in low latency. Now we extend this comparison to include the events in O3 Catalog. As opposed to online analyses, offline analyses use cleaner and better-calibrated data over which the search is performed [8]. This leads to offline events not found in online analyses since more computationally extensive methods can be used on an improved dataset. Also, some of the events found in low latency may not pass the thresholds of FAR and/or p-astro and thus do not appear in the Catalog list. The catalog events list contains a total of 74 candidates, 30 of these were found only in the offline analysis [7, 8]. Table 3 shows the results on this full set of events from O3. Out of the 74 events in O3 catalog list, the QoQ test classifies 7 of these events as Signal with presence of transient noise, while none of the events are classified as just noise. In the Q scans plotted in figure A2 we can see the presence of transient noise near the candidate time for each of these 7 events. Astrophysical candidates polluted with nearby transient noise require a more careful event validation to rule out their instrumental origin [17].

Table 3. In this table, we show the results of QoQ classification on O3 events that include offline and online events and on O3 Catalog events, along with the retracted events. Appendix contains more details on the O3 events classified as Signal + Noise, and as Noise by the QoQ test.

	O3 online retractions	O3 online non retracted	O3 Catalog events
Signal	13/23	52/56	67/74
Signal + Noise	1/23	3/56	7/74
Noise	9/23	1/56	0/74

7. Discussion

The presence of short-duration transient noise adversely impacts the quality of GW strain data. The origin of these transients is in the complicated web of detector hardware, and it is often challenging to find their exact source. Apart from masking the GW signal morphology, the non-Gaussian transient noise may sometimes mimic the astrophysical signals and end up as search pipeline detection candidates. It is thus essential to build the ability to identify and characterize potential noise artifacts in real time so that false public alerts can be reduced. Our QoQ method approaches the classification problem from first principles on how excess power in a time-frequency decomposition is expected to appear for astrophysical signals vs noise and derives straightforward criteria to separate events into three (mutually exclusive) categories: signal, noise, and signal plus noise. Few analysis parameters allow the method to tune to specific astrophysical sources and/or noise artifacts. The application of the method we described here is within the context of BBH candidate events as they are identified by matched filtering (or otherwise) detection algorithms. Our analysis starts with the Q-transformation of the strain data and thresholds on the signal energy of the time-frequency pixels. Then it uses the pixel occupancy over several time-frequency windows in order to classify each event. We tested QoQ extensively with both simulated GW waveforms the LIGO-Virgo-KAGRA Collaborations introduced as part of testing the low-latency alert infrastructure [64] as well as background data from the search for BBHs with PyCBC in O3 [35]. This allowed us to explore various options for pixel energy thresholds and actual pixel occupancy ones, leading to ROC curves.

The application of our method in the above dataset was mostly geared in minimizing the false dismissal of GW-like events while maximizing the rejection of noise. While there is room for further optimization, our present choice of analysis thresholds leads to over 40% reduction of the background (noise) while misidentifying as noise about 2% of GW-like signals. These numbers are consistent with what we obtain when applying QoQ on actual public alerts that were issued by the LIGO-Virgo-KAGRA collaboration in real-time during O3 as well as with events in the O3 catalog as published by the LIGO-Virgo-KAGRA collaborations [8]. The ability to reject additional noise events with QoQ (and beyond the efficacy we currently have) faces challenges as about half of the noise and/or background events show little signal energy (SNRs below 10) in their time-frequency decomposition.

It is an efficient algorithm to run in real-time and integrate into workflows for astronomical alert generation, mostly in the form of preventing false alerts from dissemination. Additionally, its ability to identify the simultaneous presence of signal and noise within an analysis window

can facilitate the workflow of further denoising/deglitching as well as follow-up GW parameter estimation analyses. The occurrence of transient noise (glitches) may introduce biases in the parameters estimated for a given GW candidate [8, 23, 26, 27, 69]. The ability to quantify and identify automatically such types of events will further assist their analysis with parameter estimation techniques. The enhanced sensitivity of the GW detectors in O4 will result in an increased rate of both the GW candidates and likely new sources of noise transients. Tools such as QoQ will be helpful in improving the purity of resulting real-time alerts. In the first few months of O4, we plan to collect QoQ statistics on real time low latency alerts. This will allow us to assess the performance of QoQ and make any necessary changes. Following that we propose to deploy it as a low latency data quality tool.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

Appendix. QoQ analysis of O3 catalog and retracted events

Table A1. QoQ Non Overlapping method analysis of the retracted events. The table shows the pixel occupancy values of the flagged retracted events. The first nine events in this table were also flagged by the Overlapping method, as shown in figure 8 and thus are classified as 'Noise' by the QoQ test. The last event S191220af was only flagged by the Non-Overlapping method and is thus classified as 'Signal + Noise'.

IFO	Freq band [Hz]	t1	t2	t3	t4	t5	QoQ Flag	Superevent
L1	$10 - 100$	42.0	78.7	19.8	10.8	0.0	Noise	S200308e [70]
L1	$10 - 100$	0.0	0.0	17.2	5.1	0.0	Noise	S200108v [71]
L1	$10 - 100$	0.0	0.0	0.0	1.4	5.1	Noise	S190405ar [72]
H1	$10 - 100$	0.0	0.6	2.3	0.0	0.0	Noise	S191212q [60]
H1	$10 - 100$	24.0	34.8	24.6	0.0	0.0	Noise	S200106av [73]
L1	$10 - 100$	7.1	6.0	1.7	0.0	0.0	Noise	S200106au [73]
L1	$10 - 100$	0.0	23.3	22.0	0.0	0.0	Noise	S191117j [49]
L1	$10 - 100$	0.0	24.6	0.2	0.0	0.0	Noise	S191120at [74]
L1	$10 - 100$	0.0	0.0	2.5	0.0	0.0	Noise	S200116ah [75]
L1	$10 - 100$	0.0	0.3	1.1	0.0	0.0	Signal + Noise	S191220af [76]

Table A2. QoQ Non Overlapping method analysis of the O3 Catalog events. The table shows the pixel occupancy values of the 7 out of the 74 Catalog events classified as 'Signal + Noise' and 1 out of the 56 O3 online events classified as 'Noise' by the QoQ test. The candidate names in bold were not reported in low latency and were found in offline analyses. None of these 7 events were flagged by the Overlapping method, as shown in figure 8 and thus are classified as signal with transient noise. 4 of these events (S190513bm, S190514n, S191109d and S191127p) required noise mitigation [8, 62]. The last event in the table, S191213g was classified as 'Noise' as the L1 data is contaminated by scattering noise just before and after the event time.

IFO	Freq band [Hz]	t1	t2	t3	t4	t5	QoQ Flag	Superevent
L1	$10 - 100$	2.0	0.7	0.0	0.0	0.0	Signal + Noise	*S190513bm* [8, 77]
H1	$10 - 100$	0.1	0.1	0.0	0.0	0.0	Signal + Noise	*S191109d* [8, 78]
L1	$10 - 100$	0.0	0.0	0.0	0.3	0.0	Signal + Noise	*S200224ca* [8, 79]
H1	$10 - 100$	0.0	0.0	0.0	0.0	0.4	Signal + Noise	**S191127p** [8]
L1	$10 - 100$	0.0	0.0	0.0	0.0	0.5	Signal + Noise	**S200216br** [8]
H1	$10 - 100$	0.0	1.7	0.0	0.0	0.0	Signal + Noise	**S190803e** [8]
L1	$10 - 100$	1.9	0.4	0.0	0.0	0.0	Signal + Noise	**S190514n** [8]
L1	$10 - 100$	8.6	3.7	0.0	1.3	2.5	Noise	*S191213g* [8, 61]

Note: The candidate names in bold were not reported in low latency and were found in offline analyses.
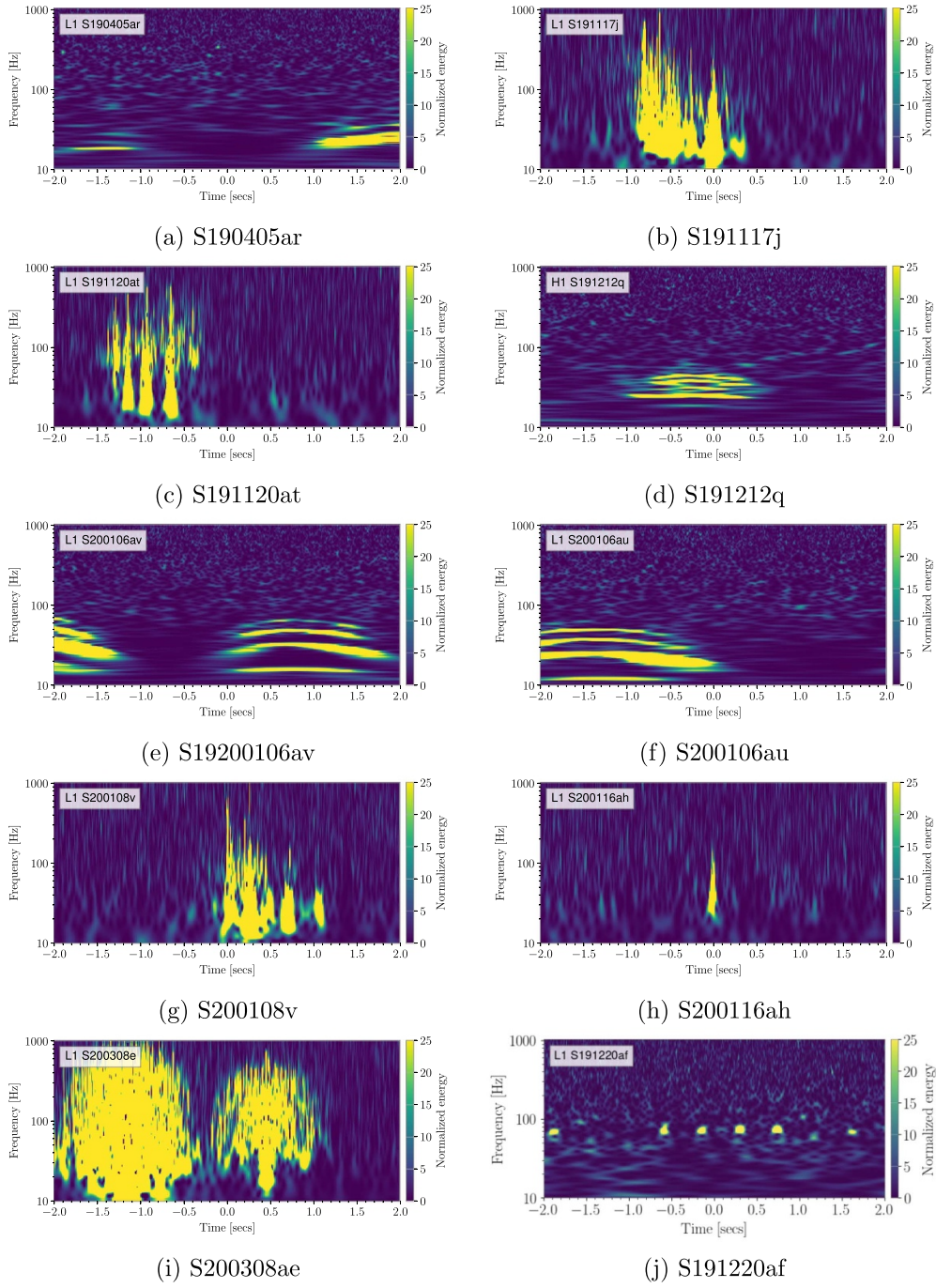
(a) S190405ar

(b) S191117j

(c) S191120at

(d) S191212q

(e) S19200106av

(f) S200106au

(g) S200108v

(h) S200116ah

(i) S200308ae

(j) S191220af

Figure A1. O3 retracted events flagged as 'Noise' by QoQ test. The last event S191220af is classified as 'Signal + Noise'.

(a) S190513bm   (b) S190514n

(c) S190803e   (d) S191109d

(e) S191127p   (f) S200216br

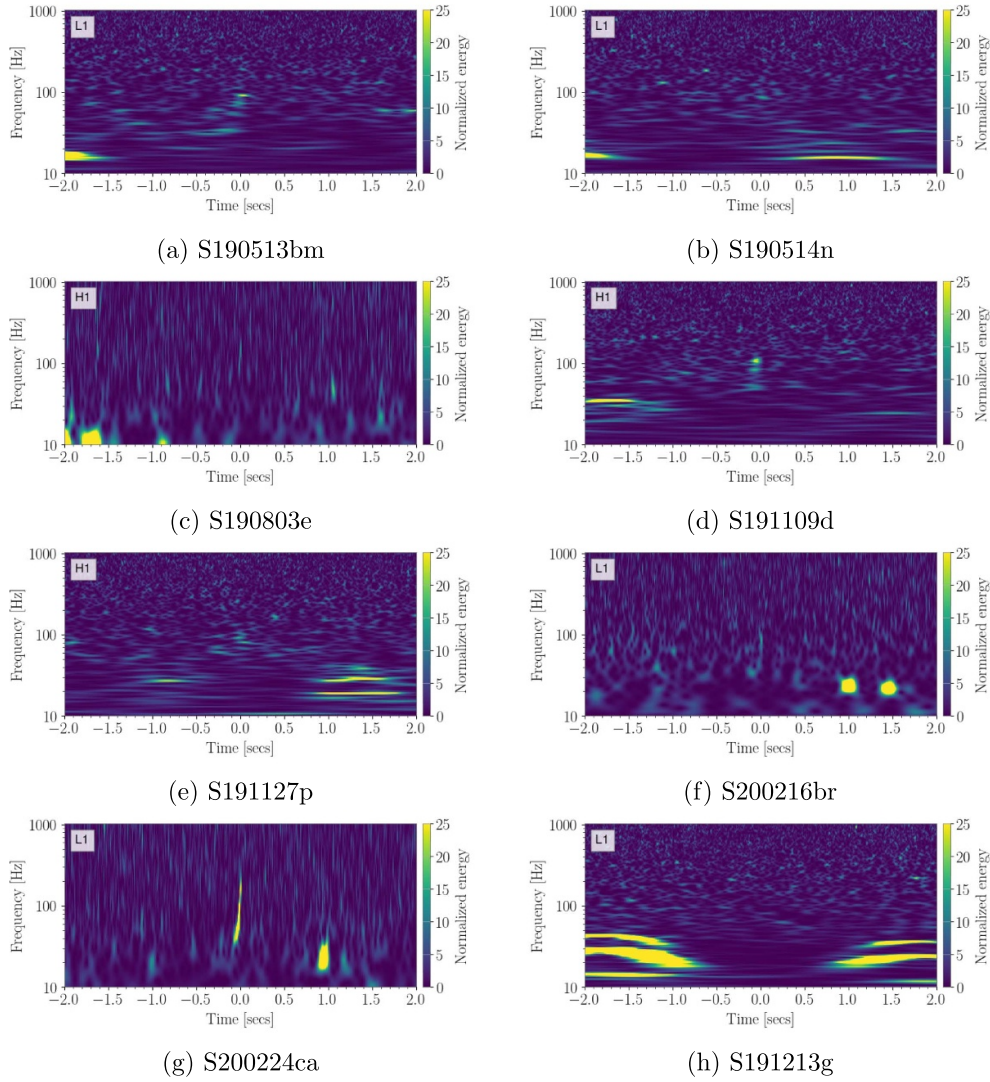(g) S200224ca   (h) S191213g

Figure A2. O3 Catalog events flagged as 'Signal + Noise' and 'Noise' by QoQ test. The first 7 events shown here are classified as 'Signal + Noise' while the event S191213g is classified as 'Noise' due the presence of high SNR scattering arches in the data.

ORCID iDs

Siddharth Soni   https://orcid.org/0000-0003-3856-8534
Reed Essick   https://orcid.org/0000-0001-8196-9267
G S Cabourn Davies   https://orcid.org/0000-0002-4289-3439
Michael W Coughlin   https://orcid.org/0000-0002-8262-2924

References

[1] Aasi J *et al* (LIGO Scientific Collaboration) 2015 *Class. Quantum Grav.* **32** 074001
[2] Acernese F *et al* (Virgo Collaboration) 2015 *Class. Quantum Grav.* **32** 024001

[3] Abbott B *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116** 061102

[4] Abbott B P *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2017 *Phys. Rev. Lett.* **119** 161101

[5] Abbott B P *et al* (LIGO Scientific Collaboration and Virgo Collaboration Collaboration) 2016 *Phys. Rev. X* **6** 041015 Abbott B P *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2018 *Phys. Rev. X* **8** 039903 (erratum)

[6] Abbott B P *et al* LIGO Scientific Collaboration and Virgo Collaboration 2019 *Phys. Rev. X* **9** 031040

[7] Abbott R *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2021 arXiv:2108.01045

[8] Abbott R *et al* (LIGO Scientific, Virgo and KAGRA Collaboration) 2021 arXiv:2111.03606

[9] Nitz A H, Capano C, Nielsen A B, Reyes S, White R, Brown D A and Krishnan B 2019 *Astrophys. J.* **872** 195

[10] Nitz A H, Dent T, Davies G S, Kumar S, Capano C D, Harry I, Mozzon S, Nuttall L, Lundgren A and Tápai M 2020 *Astrophys. J.* **891** 123

[11] Venumadhav T, Zackay B, Roulet J, Dai L and Zaldarriaga M 2019 *Phys. Rev. D* **100** 023011

[12] Venumadhav T, Zackay B, Roulet J, Dai L and Zaldarriaga M 2020 *Phys. Rev. D* **101** 083030

[13] Olsen S, Venumadhav T, Mushkin J, Roulet J, Zackay B and Zaldarriaga M 2022 *Phys. Rev. D* **106** 043009

[14] Abbott B P *et al* (LIGO Scientific, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observatory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VINROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chandra Team at McGill University, DFN, ATLAS Telescopes, High Time Resolution Universe Survey, RIMAS, RATIR and SKA South Africa/MeerKAT) 2017 *Astrophys. J. Lett.* **848** L12

[15] Magee R *et al* 2021 *Astrophys. J. Lett.* **910** L21

[16] Akutsu T *et al* (KAGRA) 2021 *Prog. Theor. Exp. Phys.* **2021** 05A101

[17] Davis D *et al* (LIGO Instrument Science) 2021 *Class. Quantum Grav.* **38** 135014

[18] Abbott B P *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2020 *Class. Quantum Grav.* **37** 055002

[19] Aasi J *et al* (VIRGO) 2012 *Class. Quantum Grav.* **29** 155002

[20] Akutsu T *et al* (KAGRA) 2021 *Prog. Theor. Exp. Phys.* **2021** 05A102

[21] Dal Canton T, Bhagwat S, Dhurandhar S V and Lundgren A 2014 *Class. Quantum Grav.* **31** 015016

[22] Abbott B P *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2018 *Class. Quantum Grav.* **35** 065010

[23] Pankow C *et al* 2018 *Phys. Rev. D* **98** 084016

[24] Powell J 2018 *Class. Quantum Grav.* **35** 155017

[25] Chatziioannou K, Cornish N, Wijngaarden M and Littenberg T B 2021 *Phys. Rev. D* **103** 044013

[26] Payne E, Hourihane S, Golomb J, Udall R, Davis D and Chatziioannou K 2022 arXiv:2206.11932

[27] Macas R, Pooley J, Nuttall L K, Davis D, Dyer M J, Lecoeuche Y, Lyman J D, McIver J and Rink K 2022 *Phys. Rev. D* **105** 103021

[28] Abbott B P *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Class. Quantum Grav.* **33** 134001

[29] Soni S *et al* (LIGO Instrument Science) 2020 *Class. Quantum Grav.* **38** 025016

[30] Dal Canton T, Nitz A H, Gadre B, Cabourn Davies G S, Villa-Ortega V, Dent T, Harry I and Xiao L 2021 *Astrophys. J.* **923** 254

[31] Sachdev S *et al* 2019 arXiv:1901.08580

[32] Aubin F *et al* 2021 *Class. Quantum Grav.* **38** 095004

[33] Hooper S, Chung S K, Luan J, Blair D, Chen Y and Wen L 2012 *Phys. Rev. D* **86** 024012

[34] Klimenko S, Mohanty S, Rakhmanov M and Mitselmakher G 2005 *Phys. Rev. D* **72** 122002

[35] Davies G S, Dent T, Tápai M, Harry I, McIsaac C and Nitz A H 2020 *Phys. Rev. D* **102** 022004

[36] Messick C *et al* 2017 *Phys. Rev. D* **95** 042001

[37] Roy S, Sengupta A S and Ajith P 2019 *Phys. Rev. D* **99** 024048

[38] Usman S A *et al* 2016 *Class. Quantum Grav.* **33** 215004

[39] Allen B 2005 *Phys. Rev.* D **71** 062001
[40] Nitz A H 2018 *Class. Quantum Grav.* **35** 035016
[41] Chatterji S, Blackburn L, Martin G and Katsavounidis E 2004 *Class. Quantum Grav.* **21** S1809–18
[42] LVC 2020 Compact binary mergers observed in the first half of the third LIGO-Virgo observing run *Technical Report* (G2001898 LSC)
[43] LVK 2020 GWTC-3 Webinar *Technical Report* (G2102416 LSC)
[44] Abbott R *et al* (LIGO Scientific, VIRGO and KAGRA) 2023 arXiv:2302.03676
[45] Cabero M *et al* 2019 *Class. Quantum Grav.* **36** 15
[46] Soni S *et al* 2021 *Class. Quantum Grav.* **38** 195016
[47] Glanzer J *et al* 2022 arXiv:2208.12849
[48] Glanzer J, Soni S, Spoon J, Effler A and González G 2023 arXiv:2304.07477
[49] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 26254 (available at: https://gcn.gsfc.nasa.gov/other/S191117j.gcn3)
[50] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 25301 (available at: https://gcn. gsfc.nasa.gov/other/S190808ae.gcn3)
[51] Vazsonyi L and Davis D 2022 arXiv:2208.12338
[52] Alvarez-Lopez S, Liyanage A, Ding J, Ng R and McIver J 2023 arXiv:2304.09977
[53] Kapadia S J, Dent T and Dal Canton T 2017 *Phys. Rev.* D **96** 104015
[54] Gabor D 1946 *J. Inst. Electr. Eng.* **93** 429–41
[55] Brown J C 1991 *J. Acoust. Soc. Am.* **89** 425–34
[56] Chatterji S K 2005 The search for gravitational wave bursts in data from the second LIGO science run *PhD Thesis* (Massachusetts Institute of Technology)
[57] Robinet F, Arnaud N, Leroy N, Lundgren A, Macleod D and McIver J 2020 *SoftwareX* **12** 100620
[58] Macleod D *et al* 2020 gwpy/gwpy: 2.0.1 (available at: https://doi.org/10.5281/zenodo.3973364)
[59] Robinet F 2015 Omicron: an algorithm to detect and characterize transient noise in gravitational-wave detectors *Technical Report* (VIR-0545C-14 Virgo)
[60] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 26394 (available at: https://gcn.gsfc.nasa.gov/other/S191212q.gcn3)
[61] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 26399 (available at: https://gcn.gsfc.nasa.gov/other/S191213g.gcn3)
[62] Abbott R *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2021 *Phys. Rev.* X **11** 021053
[63] Davis D, Littenberg T B, Romero-Shaw I M, Millhouse M, McIver J, Di Renzo F and Ashton G 2022 arXiv:2207.03429
[64] Ghosh S *et al* 2022 Mock data challenge update for 5th iteration *Technical Report* (G2201001 LVK)
[65] Chaudhary S *et al* 2023 Proceedings of the National Academy of Sciences submitted
[66] Abbott R *et al* (KAGRA, VIRGO and LIGO Scientific) 2023 *Phys. Rev.* X **13** 011048
[67] Tom K 2016 Importance sampling*Monte Carlo Methods - A Special Topics Course* pp 57–69 (available at: www.math.arizona.edu/~tgk/mc/book.pdf)
[68] Abbott R *et al* (LIGO Scientific Collaboration and Virgo Collaboration) 2021 *Astrophys. J. Lett.* **913** L7
[69] Kwok J Y L, Lo R K L, Weinstein A J and Li T G F 2022 *Phys. Rev.* D **105** 024066
[70] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2020 *GCN* 27347 (available at: https://gcn.gsfc.nasa.gov/other/S200308e.gcn3)
[71] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2020 *GCN* 26665 (available at: https://gcn.gsfc.nasa.gov/other/S200108v.gcn3)
[72] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 24109 (available at: https://gcn.gsfc.nasa.gov/gcn3/24109.gcn3)
[73] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2020 *GCN* 26641 (available at: https://gcn.gsfc.nasa.gov/other/S200106au.gcn3)
[74] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 26265 (available at: https://gcn.gsfc.nasa.gov/other/S191120at.gcn3)
[75] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2020 *GCN* 26784 (available at: https://gcn.gsfc.nasa.gov/other/S200116ah.gcn3)
[76] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 26512 (available at: https://gcn.gsfc.nasa.gov/other/S191220af.gcn3)

[77] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 24522 (available at: https://gcn.gsfc.nasa.gov/other/S190513bm.gcn3)
[78] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2019 *GCN* 26202 (available at: https://gcn.gsfc.nasa.gov/other/S191109d.gcn3)
[79] LIGO Scientific Collaboration and Virgo Collaboration Collaboration 2020 *GCN* 27184 (available at: https://gcn.gsfc.nasa.gov/other/S200224ca.gcn3)