Method to estimate the significance of coincident gravitational-wave observations from compact binary coalescence

Kipp Cannon,[1,*] Chad Hanna,[2,†] and Drew Keppel[3,4,‡]

[1]*Canadian Institute for Theoretical Astrophysics, 60 St. George Street, University of Toronto, Toronto, Ontario M5S 3H8, Canada*
[2]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[3]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany*
[4]*Leibniz Universität Hannover, D-30167 Hannover, Germany*

(Received 5 September 2012; published 16 July 2013)

Coalescing compact binary systems consisting of neutron stars and/or black holes should be detectable with upcoming advanced gravitational-wave detectors such as LIGO, Virgo, GEO and KAGRA. Gravitational-wave experiments to date have been riddled with non-Gaussian, nonstationary noise that makes it challenging to ascertain the significance of an event. A popular method to estimate significance is to time shift the events collected between detectors in order to establish a false coincidence rate. Here we propose a method for estimating the false alarm probability of events using variables commonly available to search candidates that does not rely on explicitly time shifting the events while still capturing the non-Gaussianity of the data. We present a method for establishing a statistical detection of events in the case where several silver-plated (3–5σ) events exist but not necessarily any gold-plated ($> 5\sigma$) events. We use LIGO data and a simulated, realistic, blind signal population to test our method.

DOI: 10.1103/PhysRevD.88.024025 PACS numbers: 04.30.−w, 02.60.−x, 97.80.−d, 07.05.Kf

I. INTRODUCTION

Detecting the gravitational-waves (GWs) from coalescing neutron stars and/or black holes should be possible with advanced GW detectors such as LIGO, Virgo, GEO and KAGRA [1–5]. If the performance of past detectors is any indicator of the performance of future GW detectors, they are likely to be affected by non-Gaussian noise [6]. Coincident observations are crucial in validating the detection of GWs but it is necessary to establish the probability that the coincident event could arise from noise alone.

If the detectors' data were Gaussian and stationary, it would be straightforward to compute the false alarm probability (FAP) of a coincident event based solely on its signal-to-noise ratio (SNR) and the number of independent trials. With nonstationary, non-Gaussian data the SNR is not sufficient to describe the significance of an event and, furthermore, the distribution of detector noise is not known *a priori*.

Estimating false-coincident backgrounds from time delay coincidence associated with searches for GWs was first proposed for targeted compact binary coalescence GW searches in [7]. This method has been the commonest used in subsequent searches [8–18]. We present a method to estimate the false alarm probability of a GW event from coalescing compact objects without time shifts by measuring the false alarm probability distributions for noncoincident events using a set of common variables available to the searches. This greatly simplifies analysis and lends itself nicely to an online analysis environment.

This paper is organized as follows. In Sec. II we describe a formalism for ranking GW events and establishing the probability distribution for a given event's rank in noise. In Sec. II C we present how to estimate the significance of a population of events associated with GW signals from compact binary mergers, which might include silver-plated (i.e. less than 5σ) events. In Sec. III, we test our method with a mock, advanced detector search that uses four days of LIGO fifth science run (S5) data that has been recolored to have an Advanced LIGO spectrum containing a plausible, simulated, blind population of double neutron star binary mergers. We demonstrate that we can detect GWs from neutron star binaries with very low false alarm probability.

II. METHOD

GW searches for compact binary coalescence begin by matched filtering data in the detectors [19]. If peaks in SNR times series for more than one detector are consistent with the light travel time between detectors and timing errors, these peaks are considered to be a coincident event.

GW data to date have not been stationary and Gaussian [6] thus making it difficult to model the noise in GW searches. Nonstationary noise degrades the effectiveness of standard matched filter searches. For that reason additional signal consistency tests are often employed, such as explicit χ^2 tests [20,21]. Nonstationarity occurs on several time scales. Here we are more concerned with short duration nonstationary bursts of noise called glitches for which χ^2 tests are very useful discriminators.

*kipp.cannon@ligo.org
†chad.hanna@ligo.org
‡drew.keppel@ligo.org

In this section we will present a method using common variables available to a compact binary search to estimate the FAP without relying on time shifting the detector data. Although many variables and measurements may be used, in this paper we consider two parameters: the matched filter SNR ρ_i and the χ^2 statistic χ_i^2, which depend on the detector i, as well as parameters intrinsic to the source that the template describes such as mass and spin, $\bar{\theta}$. In this section, we introduce the framework for evaluating the FAP of GW candidates.

A. Ranking events

Here is our concise definition of a coincident gravitational-wave search for compact binary sources. (i) The search consists of D detectors. (ii) We seek to find the significance of an event found in the D detectors localized in time. (iii) The intrinsic parameters of the event will be unknown *a priori*. Our detection pipeline will measure the significance as a function of the parameters of the template waveform $\bar{\theta}$.

For each detector i of a D detector network we use ρ_i and χ_i^2 to rank candidates with parameters $\bar{\theta}$ from least likely to be a gravitational wave to most likely. We use a standard likelihood ratio [22] defined as

$$\mathcal{L}(\rho_1, \chi_1^2, \ldots \rho_D, \chi_D^2, \bar{\theta}) = \frac{P(\rho_1, \chi_1^2, \ldots \rho_D, \chi_D^2, \bar{\theta}|s)}{P(\rho_1, \chi_1^2, \ldots \rho_D, \chi_D^2, \bar{\theta}|n)}, \quad (1)$$

where $P(\ldots|s)$ is the probability of observing $(\ldots)$ given a signal, and $P(\ldots|n)$ is the probability of observing $(\ldots)$ given noise. It is assumed that the signal distribution has been marginalized over all relevant parameters and the $\bar{\theta}$ refers only to the template waveform parameters that are measured by the pipeline. We make the simplifying assumption [23] that the likelihood ratio can be factored into products of likelihood ratios from individual detectors,

$$\mathcal{L}(\rho_1, \chi_1^2, \ldots \rho_D, \chi_D^2, \bar{\theta}) \approx \prod_i^D \mathcal{L}_i(\rho_i, \chi_i^2, \bar{\theta}). \quad (2)$$

The simplification that the likelihood ratio function can be factored implies statistical independence between detectors for both signals and noise. This results in a suboptimal ranking statistic. However, we can compute the FAP associated with this statistic, and in fact, it becomes much easier to do so.

B. Computing the FAP

The FAP is the probability of measuring a given $\mathcal{L}$ if the data contain only noise. N.B., this is not the same as assessing the probability that the data contain only noise, which requires knowing the prior probabilities of both signal and noise. In constructing the FAP, $P(\mathcal{L}|n)$, we start with

$$P(\mathcal{L}, \bar{\theta}|n) = \int_\Sigma P(\mathcal{L}_1, \ldots \mathcal{L}_D, \bar{\theta}|n) \mathrm{d}^{D-1}\Sigma, \quad (3)$$

where Σ is the surface of constant $\mathcal{L} = \prod_i^D \mathcal{L}_i$. From (2), we have, assuming that the likelihood ratio values in noise are independent between the detectors,

$$P(\mathcal{L}_i, \ldots \mathcal{L}_D, \bar{\theta}|n) = \prod_i^D P(\mathcal{L}_i, \bar{\theta}|n), \quad (4)$$

where $P(\mathcal{L}_i, \bar{\theta}|n)$ is obtained by marginalizing over ρ_i, and χ_i^2 in the single-detector terms,

$$P(\mathcal{L}_i, \bar{\theta}|n) = \int_\sigma P(\rho_i, \chi_i^2, \bar{\theta}|n) \mathrm{d}\sigma, \quad (5)$$

where σ is the contour of constant $\mathcal{L}_i$ in the $\{\rho_i, \chi_i^2\}$ surface at constant $\bar{\theta}$. Implicit in (4) and (5) is the assumption that the coincidence criteria do not depend on ρ_i, χ_i^2 or $\bar{\theta}$. Finally, $P(\mathcal{L}|n)$ is obtained by marginalizing over $\bar{\theta}$,

$$P(\mathcal{L}|n) = \int P(\mathcal{L}, \bar{\theta}|n) \mathrm{d}\bar{\theta}. \quad (6)$$

Given an event resulting from noise, the probability of observing it to have a likelihood ratio value at least as large as some threshold $\mathcal{L}^*$ is

$$P(\mathcal{L} \geq \mathcal{L}^*|n) = \int_{\mathcal{L}^*}^\infty P(\mathcal{L}|n) \mathrm{d}\mathcal{L}. \quad (7)$$

A GW search will typically produce multiple coincident events during a given experiment. That means that there will be multiple opportunities to produce an event with a certain likelihood value. We are ultimately interested in the probability of getting one or more events with $\mathcal{L} \geq \mathcal{L}^*$ after all the events are considered. The probability of getting at least one such event after forming M independent coincidences[1] can be adjusted by the complement of the binomial distribution

$$\begin{aligned} P(\mathcal{L} \geq \mathcal{L}^*|n_1, \ldots, n_M) &:= 1 - \binom{M}{0} P(\mathcal{L} \geq \mathcal{L}^*|n)^0 \\ &\quad \times (1 - P(\mathcal{L} \geq \mathcal{L}^*|n))^M \\ &= 1 - (1 - P(\mathcal{L} \geq \mathcal{L}^*|n))^M. \quad (8) \end{aligned}$$

This is the FAP at $\mathcal{L}^*$ in an experiment that yielded M coincident events.

[1]In practice it can be difficult to know if the coincidences formed are independent; however as long as they are related to the true number of independent trials by an overall scaling, one can adjust the number so that it agrees with the observed rate of coincidences for low significance events. This works because GWs are very rare and true signals will vastly underwhelm the false positives that a pipeline produces at high FAP. Thus the bias in calibrating M to high FAP events is very small.

C. FAP of populations of GW events

A population of events can collectively be more significant than the single most significant event alone. Indeed, population analyses have previously been employed in looking for GW signals associated with gamma-ray bursts (GRBs). For example, a Student's *t*-test was proposed in [24] to test for deviations in the cross correlation of detectors' output preceding a set of times associated with GRBs (i.e., *on-source* times) when compared to other *off-source* times not associated with GRBs, a binomial test was employed in [25,26] using the X% most significant events to test for excess numbers of events at their associated FAPs, a Kolmogorov test was used in [27] to look for deviations from isotropy in GRB direction based on the directional sensitivity of the bar detectors, and a Mann-Whitney U (or Mann-Whitney-Wilcoxon) test was performed in [28] to test if all the FAPs associated with the on-source events of the GRBs were on average smaller than the expected distribution given by the off-source events, as would be the case if the average significance were elevated due to the presence of GWs in the on-source events.

As noted in [25,26], seeking significance by considering different choices of populations diminishes the significance of each on account of the trials that have been conducted. We propose to control this by restricting ourselves to considering only populations consisting of events for whose ranking statistic values the expected number of background events was less than 1 (i.e., $MP(\mathcal{L} \geq \mathcal{L}^*|n) < 1$). Although it is conceivable that still less significant events could be sufficiently numerous to be statistically significant as a population, we consider it unlikely that such events will be interesting and so this is a natural stopping condition for considering events. The statistic we propose is

$$Q := \min_i \{P(\geq i|x_i)\}, \tag{9}$$

where x_i is related to the significance of the ith most significant event and $P(\geq i|x_i)$ is the probability of obtaining i or more events of that significance. In Appendix A we go into more detail defining x_i as well as providing an algorithm for computing the FAP of this statistic semi-analytically (i.e., without the use of a potentially computationally costly Monte Carlo simulation).

III. EXAMPLE

We have applied these techniques to a mock search for GWs from binary neutron stars in four days of S5 LIGO data that have been recolored to match the Advanced LIGO design spectrum[2] [29]. This provides a potentially realistic data set that contains glitches from the original LIGO instruments. A population of neutron star binaries was added at a rate of $4\mathrm{Mpc}^{-3}\,\mathrm{Myr}^{-1}$ (see [1] for the expected rates.) We self-blinded the signal parameters with a random number generator.

Our analysis targeted compact binary systems with component masses between 1.2 and $2M_\odot$. We used 3.5 post-Newtonian order stationary phase approximation templates to cover the parameter space with a 97% minimal match [30] by neglecting the effects of spin in the waveform models [31]. This required $\sim$15, 000 templates. We started the matched filter integrals at 15 Hz and extended the integral to the innermost stable circular orbit frequency. The analysis gathered the data, whitened it, filtered it, identified events in the single detectors, found coincidences and ranked the events by their likelihood ratios. The filtering algorithm is described in [32].

The previous section described our method for estimating the significance of events but did not describe many details of how the calculation is done in practice. We will point out a few of those details now.

The numerator of (1) is evaluated by assuming the signals follow their expected distribution in Gaussian noise. We note that this is a reasonable assumption because detections are likely to come from periods of relatively stationary and Gaussian data. Note that the expectation for ρ can be obtained by assuming that sources are distributed uniformly in space. The expectation for the χ^2 of a signal can be found in [20].

The denominator of (1) is found by explicitly histogramming the single detector events that are not found in coincidence. By excluding coincident events we lower the chance that a gravitational wave will bias the noise distribution of the likelihood ratios. In general the histogramming will suffer from finite statistics and "edge" effects. We generate the histograms at a finer resolution than required to track the likelihood ratio and then apply a Gaussian smoothing kernel with a width characteristic of the uncertainty in ρ.

We are unable to collect enough statistics to fully resolve the tail of the background ρ distribution. Thus, we add a prior distribution into the background statistics that models the ρ falloff as expected from a 2 degree of freedom matched filter in Gaussian noise, i.e. $p(\rho|n) \propto \exp[-\rho^2/2]$. This helps ensure that the likelihood ratio contours increase as a function of ρ at large ρ. At some point the probability of getting a given value of ρ, χ^2 becomes smaller than double-precision floating-point epsilon. We extend the background distribution above a given value of ρ with a polynomial in ρ that falls off faster than the signal distribution (which is $\propto \rho^{-4}$) but is shallow enough to prevent numerical problems. In both cases the point of the prior is not to influence the ranking of typical events but rather to make the calculations more numerically well behaved. The prior is added so that the total probability amounts only to a single event in each detector. Thus the background (as billions of events are collected)

[2]Specifically the zero-detuned, high-power noise curve was used.

quickly overwhelms the prior except for at the edges where there are no data. The point where the calculation is no longer based on having at least one actual event in background is important since it will effectively mark the limiting FAP. More discussion of that point follows.

In Fig. 1 we show some of the intermediate data used in estimating the significance of events in our example. Namely, we show the individual likelihood ratio contours for ρ and χ^2 described in (1) in the H1 and L1 instruments for signals with a chirp mass consistent with a neutron star binary ($1.2M_\odot$) in Figs. 1(a) and 1(b) respectively. The probability of getting an event with a likelihood ratio greater than $\mathcal{L}^*$ after M trials for the H1 and L1 instruments (8) is shown in Fig. 1(c). Our ability to measure $P(\mathcal{L} \geq \mathcal{L}^* | n_1, \ldots, n_M)$ is limited by the number of events that we collect in our background estimate. The shaded region shows the $\sqrt{N}$ error region found by assuming Poisson errors on the number of events that went into computing a given point on the curve. We have indicated the FAP at which there ceases to be more than 1 event collected in the background by a dashed line. The dashed line shows the $P(\mathcal{L} \geq \mathcal{L}^* | n_1, \ldots, n_M)$ has background events to $\mathcal{P} := 7 \times 10^{-5}$ which is nearly the FAP required for a 4σ detection. Below the dashed line the FAP estimate is dominated by the Gaussian smoothing kernel applied to the planes in Figs. 1(a) and 1(b). We believe that it is reasonable to trust the FAP estimate beyond the single background event limit but note that 5σ level confidence can still be reached without extrapolation with tighter coincidence criteria. Tighter coincidence criteria would reduce the trials factor and permit higher significances to be estimated. The best way to do this is to demand that three or more detectors see an event. In our example a third detector would lower the trials factor by ~ 100, which would shift the limiting FAP, $\mathcal{P}$ to $\sim 7 \times 10^{-7}$. It is worth mentioning that the background events and number of independent trials are accumulated at the same rate. Thus one cannot decrease the limiting FAP by collecting more data.

After assigning the FAP to events we also assign a false alarm rate (FAR), which is described in Appendix A and given by (A6). This allows us to produce the standard inverse FAR (IFAR) plot commonly produced in recent searches for compact binaries [14–18] without having relied on time shifting the detector events to estimate the background. This is shown in Fig. 2(a).

The IFARs of the most significant events that came out of this search in Fig. 2(a) can be identified as the long tail in the observed events distribution. The top event has a significance greater than 5σ, the level necessary for claiming the detection of GWs. The second loudest event has a significance greater than 4σ. Both events surpass the single background event limit $\mathcal{P}$. If restricted to this limit then both events are nearly 4σ.

Applying the population procedure we have put forth in Sec. II C, we produced a more significant statement about

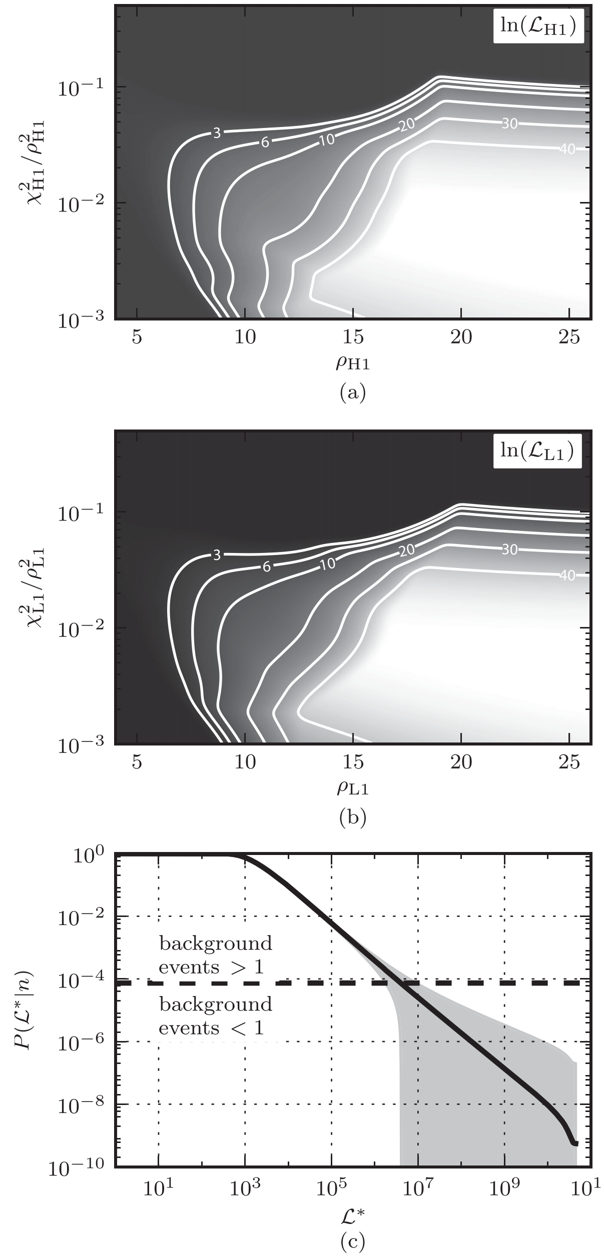

FIG. 1. (a), (b) Show the likelihood ratios $\mathcal{L}_{H1}$, $\mathcal{L}_{L1}$ as a function of ρ and χ^2 for H1 and L1 respectively for templates with masses consistent with neutron star binaries (1.2–$2M_\odot$). $\mathcal{L}_{H1}$, $\mathcal{L}_{L1}$ appear as the right-hand side of (2). White indicates high likelihood ratio values. (c) Shows the probability of having obtained a given value of likelihood ratio $\mathcal{L}^*$ or greater from noise as defined in (8) after M trials (where M is the number of independent coincidences formed). In this example $M = 6 \times 10^4$.

the presence of GWs beyond that of the loudest event. This effect is mostly attributed to the similar significance of the top two events. This could happen in a real analysis in two ways: (1) nature could just provide such a set of events as in this example; (2) both events exhaust our ability to measure significance and we must place an upper bound

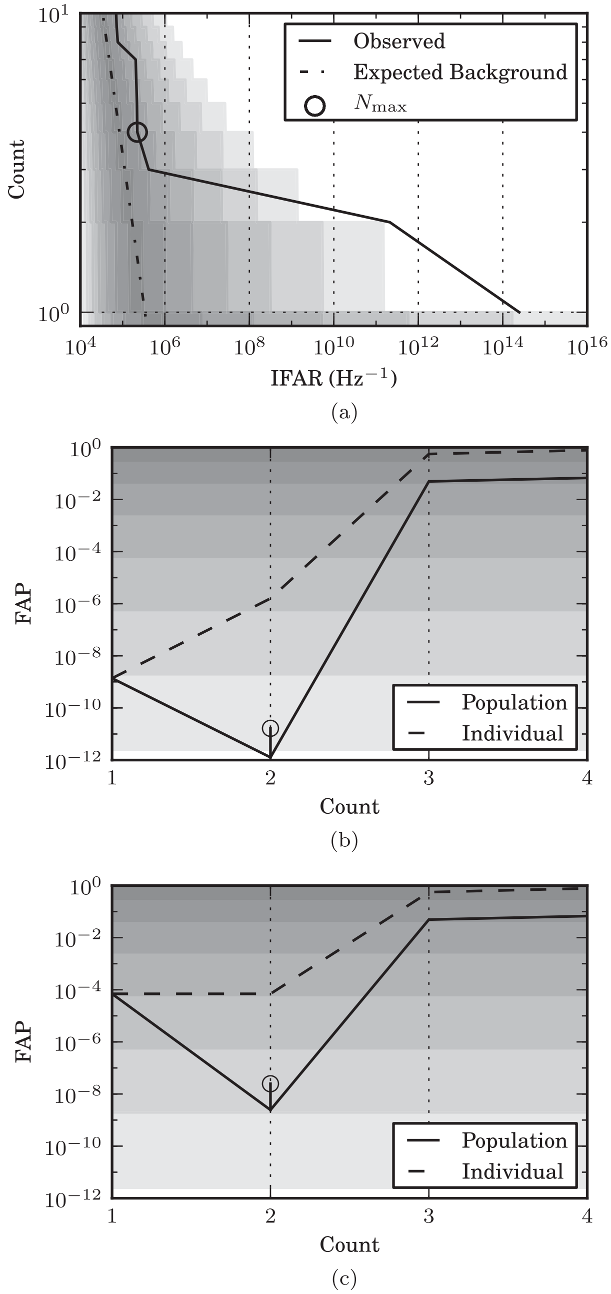

FIG. 2. (a) It is a standard IFAR plot where the shaded regions correspond to the "1σ" through "7σ" regions computed using the survival function and point percent function associated with the Poisson distribution. This is used to determine where to stop the accumulation of events for the population statement. (b) Shows the FAP associated with each of the individual events in the population we are considering, given by (8), as well as the FAP of obtaining the N loudest events, given by (A4). Also shown in (c) are the same traces obtained after restricting the individual FAPs to be greater than $\mathcal{P} = 7 \times 10^{-5}$. Note in both (b), (c) the circle marker indicates the FAP of having observed the minimum point on this curve, Q. It is important to note that even in the case where the limiting FAP for individual events cannot meet the "5σ" criterion, a population can. This could be a realistic scenario for early GW detections.

on the FAP. The latter case, although somewhat artificial, could still play an important role in analysis, especially if one is unable to confidently declare a single 5σ event but finds two or more events with 3 or 4σ. With our example analysis the combination of the two loudest events was a 5σ excursion even after restricting the FAP of both events to be $\mathcal{P}$. After examining the signal population we found that both candidates were separately associated with signal injections.

IV. CONCLUSION

We have provided a method for estimating the significance of GWs from compact binary coalescence using measurements of single instrument populations of ρ and χ^2 as a function of the template waveform intrinsic parameters. We demonstrated our method with mock Advanced LIGO data derived from initial LIGO data including a realistic population of compact binary merger signals and glitches. We found that between our two loudest events we were able to establish detection at greater than 5σ confidence. Both of the loudest two events exhausted the $\mathcal{P}$ ($\sim 4\sigma$) background estimate, but the extrapolated FAP of the loudest event exceeded 5σ on its own. Both of the loudest events were associated with the blind signal population introduced into the data and the remaining events were consistent with the expectation from background.

ACKNOWLEDGMENTS

Research at Perimeter Institute is supported through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation. K. C. was supported by the National Science and Engineering Research Council, Canada. D. K. was supported by the Max Planck Gesellschaft. This work has LIGO document number LIGO-P1200031. The authors wish to thank Rahul Biswas, Patrick Brady, Gabriela González Jordi Burguet-Castell, Sarah Caudill and Ruslan Vaulin for many fruitful discussions and John Whelan for a careful review of this work. The authors also wish to thank the referee of this work who exposed flaws in our original population analysis and pointed us in the direction of the final analysis shown here. We thank the UW-Milwaukee Center for Gravitation and Cosmology and the LIGO Laboratory for use of its computing facility to make this work possible through NSF Grants No. PHY-0923409 and No. PHY-0600953. We thank the LIGO Scientific Collaboration and the Virgo Scientific Collaboration for providing us with the data to test the methods described in this work.

APPENDIX A: FALSE ALARM PROBABILITY OF Q

In this section we describe how we can compute the FAP of our statistic (9) in two scenarios with different choices of x_i: the exact solution valid for all values of M, and the solution in the limit of $M \gg 1$.

The exact solution is given by computing the statistic using the binomial probability of observing k or more events for each of the events in the region $MP(\mathcal{L} \geq \mathcal{L}^*|n) < 1$, where $P(\mathcal{L} \geq \mathcal{L}^*|n)$ refers to the values calculated in (7). The probability of obtaining k events with $P(\mathcal{L} \geq \mathcal{L}^*|n) \in [0, x)$ is given by

$$P(k|x) = \binom{M}{k} x^k (1-x)^{M-k} \tag{A1}$$

and the probability of getting k or more events more significant than x is given by

$$P(\geq k|x) = 1 - \sum_{j=0}^{k-1} \binom{M}{j} x^j (1-x)^{M-j}. \tag{A2}$$

When $M \gg 1$, the factorials and exponential operations involved in (A1) and (A2) will become unwieldy. In the limit of $M \to \infty$, since the interval we are interested in is $[0, 1/M)$ the average number of events in the interval is unity and the distribution of events in this interval can be approximated by the Poisson distribution. In this case, the probability of observing k when λ are expected is

$$P(k|\lambda) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \tag{A3}$$

and the probability of getting k or more events when λ are expected is

$$P(\geq k|\lambda) = 1 - \mathrm{e}^{-\lambda} \sum_{j=0}^{k-1} \frac{\lambda^j}{j!}. \tag{A4}$$

To obtain the expected number of events due to background associated with the single event FAP $P(\mathcal{L} \geq \mathcal{L}^*|n)$, we can use (A4), setting $N = 1$, and solving for λ,

$$\lambda(\mathcal{L}^*) = -\ln[1 - P(\mathcal{L} \geq \mathcal{L}^*|n)]. \tag{A5}$$

This is the quantity that would be used in (A4) associated with our Q statistic (9) and only those events with $\lambda(\mathcal{L}^*) < 1$ are considered for this statistic. Historically, GW experiments have used false alarm rates to rank events [14–18]. The quantity, inverse false alarm rate, is given here by

$$\mathrm{IFAR}(\mathcal{L}^*) = T/\lambda(\mathcal{L}^*), \tag{A6}$$

where T is the observation time of the experiment.

For either case, the derivation of the FAP associated with the statistic from (9), $Q = \min_i\{P(\geq i|x_i)\}$, proceeds in the same manner, where x_i is $P(\mathcal{L} \geq \mathcal{L}_i^*|n)$ or λ_i for the binomial or Poisson cases, respectively. For ease of notation, we outline the derivation for the Poisson case in Appendix A 1 and give the differences in the final result for the binomial case in Appendix A 2.

1. The Poisson approximation

The computation of the FAP of the statistic Q from (9) for events that have $\lambda(\mathcal{L}_i^*) < 1$, where $P(\geq i|\lambda(\mathcal{L}_i^*))$ is given by (A4), proceeds as follows. Let us only consider events that are produced from the background alone. With each of these background events, let us associate with the ith most significant event a rate λ_i. The possible numbers of events that could have been obtained with $\lambda_i < 1$ and $\min_i\{P(i|\lambda_i)\} > Q$ are given by $\{k \in \mathbb{N}_N^*\colon P(k|1) \geq Q\}$, where $P(k|1)$ is given by (A3). Since there are only a finite number of these events, we find it easier to compute the probability of obtaining a statistic (9) less significant than Q, rather than more significant than Q, and then take the complement to compute the FAP of obtaining Q.

When k events are observed, the probability of getting $\min_i\{P(\geq i|\lambda_i)\} > Q$ is given as

$$P(\min_i\{P(\geq i|\lambda_i)\} > Q|k) = \frac{A_k(Q)}{B_k}, \tag{A7}$$

which can be computed as a series of integrals, one for each of the A's and B's.

The first set of k integrals compute the normalization B_k, which is the volume of the k-event parameter space. This is given by

$$B_k = \int_{\lambda_k=0}^{\lambda_k=1} \int_{\lambda_{k-1}=0}^{\lambda_{k-1}=\lambda_k} \cdots \int_{\lambda_1=0}^{\lambda_1=\lambda_2} P(\lambda_1)\ldots P(\lambda_{k-1}) \times P(\lambda_k)\mathrm{d}\lambda_1\ldots\mathrm{d}\lambda_{k-1}\mathrm{d}\lambda_k, \tag{A8}$$

where the distribution of each event in λ, $P(\lambda_i)$, can be approximated as uniform in the limit $M \gg 1$. The upper limits on these integrals impose the constraint that the events are ordered by their rates (i.e., $\lambda_1 < \cdots < \lambda_k < 1$). Performing these integrals, we find $B_k = 1/k!$.

The second set of k integrals computes the volume of the k-event parameter space that would have produced $\min_i\{P(\geq i|\lambda_i)\} > Q$. Let us identify the root z_i of the function $P(\geq i|\lambda) - Q$ as the solution of the transcendental equation

$$(1-Q)\mathrm{e}^{z_i} = \sum_{j=0}^{i-1} \frac{z_i^j}{j!}. \tag{A9}$$

For the ith event, if the event had $\lambda_i < z_i$, then this event would have produced $P(\geq i|\lambda_i) < Q$, an example of which is visualized in Fig. 3. Thus, in order to limit the integrals of (A8) to only the region where $\min_i\{P(\geq i|\lambda(\mathcal{L}_i^*))\} > Q$, we need to set the lower limit of the integral over the rate λ_i to be the root z_i. These integrals then take the form

$$A_k(Q) = \int_{\lambda_k=z_k(Q)}^{\lambda_k=1} \int_{\lambda_{k-1}=z_{k-1}(Q)}^{\lambda_{k-1}=\lambda_k} \cdots \int_{\lambda_1=z_1(Q)}^{\lambda_1=\lambda_2} P(\lambda_1)\ldots \times P(\lambda_{k-1})P(\lambda_k)\mathrm{d}\lambda_1\ldots\mathrm{d}\lambda_{k-1}\mathrm{d}\lambda_k. \tag{A10}$$

We find a recursion relation exists for the computation of $A_k(Q)$, where

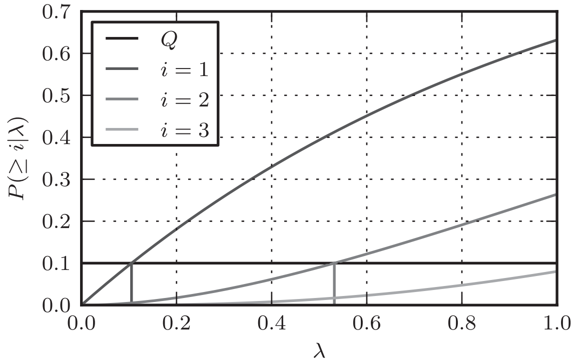

FIG. 3. The Poisson probability of obtaining i or more events with a rate λ for several values of i. The horizontal line is an example observed statistic of $Q = 0.1$. Looking at the right edge of the $i = 3$ curve shows us that we would have obtained a statistic value less that Q if there had been three events anywhere in the region $\lambda_3 < 1$. The vertical lines show the roots of $P(\geq i|\lambda) - Q$ for the cases $i = 1, 2$.

$$A_k(Q) = U(Q, k, 1), \tag{A11}$$

$$U(Q, k, l) = \frac{1}{k!} W(Q, k, k + l - 1) - \sum_{i=1}^{k-1} \frac{z_{l+i}^i}{i!} U(Q, k - i, l + i), \tag{A12}$$

$$W(Q, i, j) = 1 - z_j^i(Q). \tag{A13}$$

Combining the results from (A8) and (A10), we find (A7) to be

$$P(\min_i \{P(\geq i|\lambda_i)\} > Q|k) = k! U(Q, k, 1). \tag{A14}$$

Finally, combining the probability of observing k events (A3) with the conditional probability of k events producing a result less significant that Q (A14), we obtain the FAP of the statistic $\min_i \{P(\geq i|\lambda(\mathcal{L}_i^*))\}$,

$$P(\min_i \{P(\geq i|\lambda(\mathcal{L}_i^*))\} < Q|n) = 1 - \sum_{i=0}^{k(Q)} P(i|1) P(\min_j \{P(\geq j|\lambda_j))\} > Q|i). \tag{A15}$$

This result is displayed visually in Fig. 4 where we show the effective trials factor, $P(\min_i \{P(\geq i|\lambda_i)\} < Q|n)/Q$, as a function of the observed statistic Q.

2. The binomial solution

The equivalent statistic for which we wish to compute the FAP is $\min_i \{P(\geq i|P(\mathcal{L} \geq \mathcal{L}_i^*|n))\}$. This calculation proceeds as in Sec. A 1 with one minor difference. The polynomial (A13) of the recurrence relation is then given by

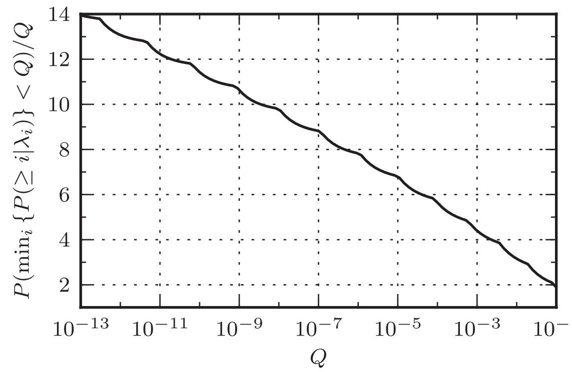

FIG. 4. The effective trials factor that is applied to the statistic $Q = \min_i \{P(\geq i|\lambda(\mathcal{L}_i^*))\}$ in order to obtain the FAP for Q. The effective trials factor grows with increasing significance of the observed statistic.

$$W(Q, i, j) = \left(\frac{1}{M}\right)^i - z_j^i(Q), \tag{A16}$$

where the roots $\{z_i\}$ are associated with $P(\geq i|z_i) = Q$ and $P(\geq i|z_i)$ is given by (A2). The final FAP of $\min_i \{P(\geq i|P(\mathcal{L} \geq \mathcal{L}_i^*|n))\}$ is then

$$P(\min_i \{P(\geq i|P(\mathcal{L} \geq \mathcal{L}_i^*|n))\} < Q|n) = 1 - \sum_{i=0}^{k(Q)} P(i|1/M) P(\min_j \{P(\geq j|P(\mathcal{L} \geq \mathcal{L}_j^*|n))\} > Q|i), \tag{A17}$$

where $P(i|1/M)$ is given by (A1) and the conditional probability $P(\min_i \{P(\geq i|P(\mathcal{L} \geq \mathcal{L}_i^*|n))\} > Q|i)$ is given by

$$P(\min_i \{P(\geq i|P(\mathcal{L} \geq \mathcal{L}_i^*|n))\} > Q|i) = i! M^i U(Q, i, 1), \tag{A18}$$

with $U(Q, i, 1)$ internally using (A16).

APPENDIX B: NUMERICAL CONSIDERATIONS

1. Equation (8)

As the duration of the experiment increases, the numerical evaluation of (8) using fixed-precision floating-point numbers becomes challenging. In this limit, the per-trial false alarm probability of interesting events is very small and the number of trials is very large. Using double-precision floating-point numbers, when the number of trials gets larger than about 10^{10}, FAPs of 10^{-6} and 0 become indistinguishable, and as the number of coincidences that are recorded increases further "4σ" and "5σ" events cannot be differentiated—it is no longer possible to make detection claims. The following procedure can be used to evaluate (8) for all $P(\mathcal{L} \geq \mathcal{L}^*|n)$ and M. If $MP(\mathcal{L} \geq \mathcal{L}^*|n) < 1$ the Taylor expansion of (8) about $P(\mathcal{L} \geq \mathcal{L}^*|n) = 0$ converges quickly:

$$1-(1-P)^M = MP-(M^2-M)\frac{P^2}{2} + (M^3-3M^2+2M)\frac{P^3}{6} - (M^4-6M^3+11M^2-6M)\frac{P^4}{24}+\cdots = \sum_{i=0}^{\infty} -1^i \frac{P^{(i+1)}}{(i+1)!}[(M-0)(M-1)\ldots(M-i)]. \tag{B1}$$

The last form yields a recursion relation allowing subsequent terms in the series to be computed without explicit evaluation of the numerator and denominator separately (which, otherwise, would quickly overflow): if the $(i-1)$th term is X, the ith term in the series is $X\frac{i-M}{i+1}P$.

If $MP(\mathcal{L} \geq \mathcal{L}^*|n) \geq 1$ the Taylor series still converges (in fact, as long as the number of trials M is an integer the series is exact in a finite number of terms) but the series is numerically unstable: the terms alternate sign and one must rely on careful cancellation of large numbers to obtain an accurate result. In this regime the expression's value is close to 1, so $(1-P)^M$ is small. If P is small, we can write

$$1-(1-P)^M = 1-\mathrm{e}^{M\ln(1-P)} \tag{B2a}$$

and then the Taylor expansion of $M\ln(1-P)$ about $P=0$ converges quickly,

$$M\ln(1-P) = -MP\left(1+\frac{P}{2}+\frac{P^2}{3}+\cdots\right). \tag{B2b}$$

Altogether, the algorithm for evaluating (8) is as follows: if $MP(\mathcal{L} \geq \mathcal{L}^*|n) < 1$ use (B1) computed via the recursion relation; otherwise if $P(\mathcal{L} \geq \mathcal{L}^*|n) < 0.125$ use (B2); otherwise evaluate (8) directly using normal floating-point operations. The threshold of $P(\mathcal{L} \geq \mathcal{L}^*|n) < 0.125$ for using (B2) is found empirically; the results are not sensitive to the choice of this number.

2. Equation (A5)

The evaluation of (A5) for events that are interesting as detection candidates after an experiment is concluded is straightforward using double-precision floating-point arithmetic. In this regime, $P(\mathcal{L} \geq \mathcal{L}^*|n) \sim 10^{-5}$, and there is plenty of numerical dynamic range available. However, the practical use of (A5) is in its ability to identify "once a day" or "once an hour" events for the purpose of providing alerts to the transient astronomy community. After just one day, 24 "once an hour" background events are expected, and their FAP—the probability of observing at least one such event from a Poisson process you expect to have produced 24—is 0.9999999999622486. After 37 events are expected, double-precision numbers can no longer be used to differentiate those events' FAPs from 1; that is, (A5) can only assign reliable false alarm rates to the 30 or so most significant background events in any experiment.

This problem is addressed by not computing the expected number of events, $\lambda(\mathcal{L}^*)$, from the false alarm probability, $P(\mathcal{L} \geq \mathcal{L}^*)$, as shown in (A5), but by first rewriting (7) and (8) as

$$1-P(\mathcal{L} \geq \mathcal{L}^*|n_1,\ldots,n_M) = \left(\int_0^{\mathcal{L}^*} P(\mathcal{L}|n)\mathrm{d}\mathcal{L}\right)^M, \tag{B3}$$

from which we can rewrite (A5) as

$$\lambda(\mathcal{L}^*) = -M\ln\int_0^{\mathcal{L}^*} P(\mathcal{L}|n)\mathrm{d}\mathcal{L}. \tag{B4}$$

This form of the expression presents no challenges to its evaluation using double-precision floating-point arithmetic.

[1] J. Abadie *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **27**, 173001 (2010).
[2] G. M. Harry (LIGO Scientific Collaboration), Classical Quantum Gravity **27**, 084006 (2010).
[3] Virgo Collaboration, Report No. VIR027A09, 2009, https://tds.ego-gw.it/itf/tds/file.php?callFile=VIR-0027A-09.pdf.
[4] B. Willke, P. Ajith, B. Allen, P. Aufmuth, C. Aulbert *et al.*, Classical Quantum Gravity **23**, S207 (2006).
[5] K. Somiya (KAGRA Collaboration), Classical Quantum Gravity **29**, 124007 (2012).
[6] J. Slutsky *et al.*, Classical Quantum Gravity **27**, 165023 (2010).
[7] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **69**, 122001 (2004).
[8] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **72**, 082001 (2005).
[9] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **72**, 082002 (2005).
[10] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **73**, 062001 (2006).
[11] B. P. Abbott *et al.* (LIGO Scientific Collaboration and TAMA Collaboration), Phys. Rev. D **73**, 102002 (2006).
[12] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **78**, 042002 (2008).
[13] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **77**, 062002 (2008).
[14] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **79**, 122001 (2009).

[15] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **80**, 047101 (2009).

[16] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **82**, 102001 (2010).

[17] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **83**, 122005 (2011).

[18] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **85**, 082002 (2012).

[19] B. A. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Phys. Rev. D **85**, 122006 (2012).

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

[21] C. Hanna, Ph.D. thesis, Louisiana State University, 2008.

[22] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, Cambridge, England, 2005).

[23] K. C. Cannon, Classical Quantum Gravity **25**, 105024 (2008).

[24] L. S. Finn, S. D. Mohanty, and J. D. Romano, Phys. Rev. D **60**, 121101 (1999).

[25] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **77**, 062004 (2008).

[26] B. P. Abbott *et al.*, Astrophys. J. **715**, 1438 (2010).

[27] P. Astone *et al.*, Phys. Rev. D **71**, 042001 (2005).

[28] J. Abadie *et al.*, Astrophys. J. **715**, 1453 (2010).

[29] LIGO Scientific Collaboration, "Advanced LIGO anticipated sensitivity curves," 2009, https://dcc.ligo.org/LIGO-T0900288-v3/public.

[30] B. J. Owen, Phys. Rev. D **53**, 6749 (1996).

[31] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. Sathyaprakash, Phys. Rev. D **80**, 084043 (2009).

[32] K. Cannon, R. Cariou, A. Chapman, M. Crispin-Ortuzar, N. Fotopoulos, M. Frei, C. Hanna, E. Kara, D. Keppel, L. Liao, S. Privitera, A. Searle, L. Singer, and A. Weinstein, Astrophys. J. **748**, 136 (2012).